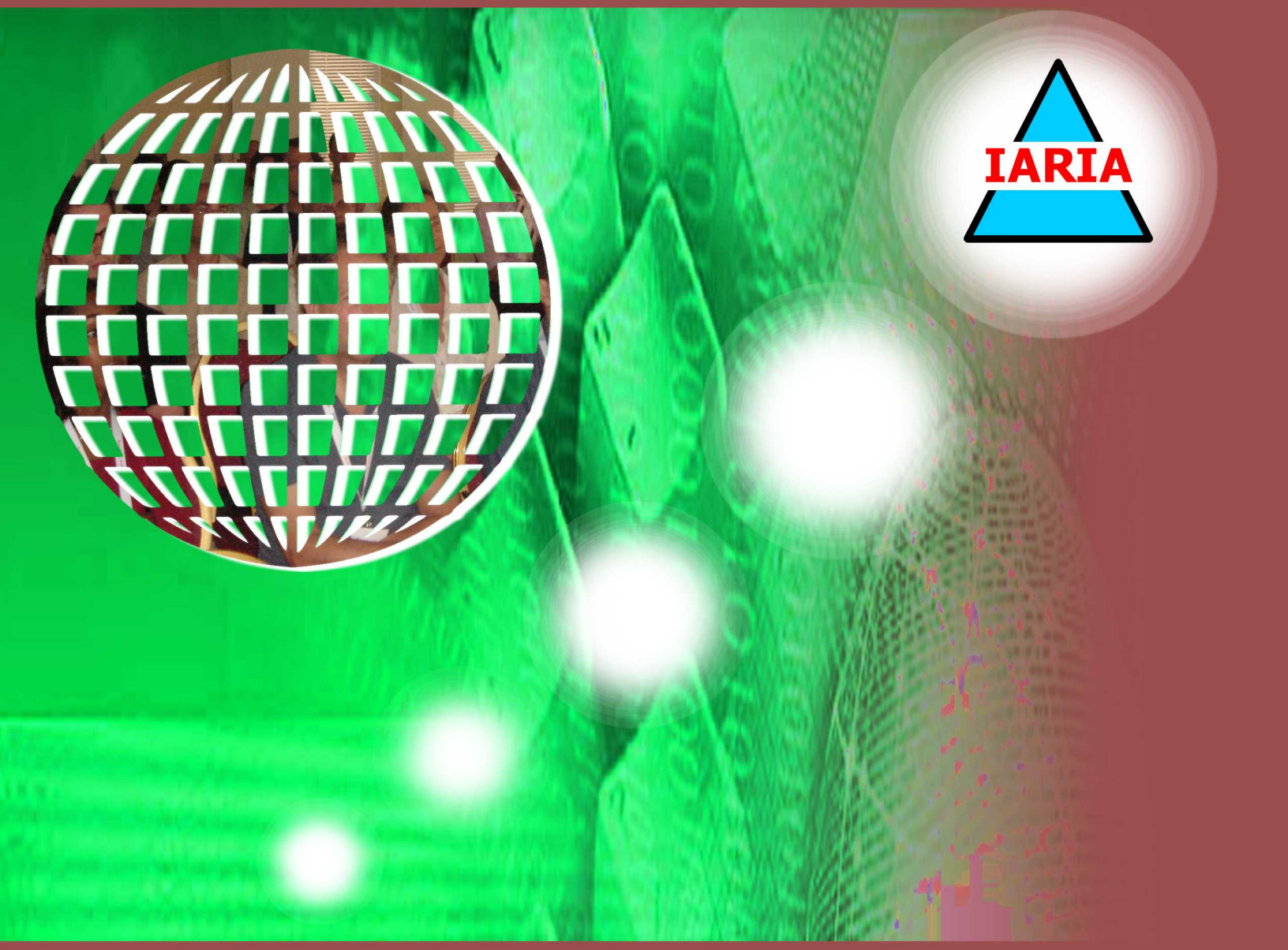
International Journal on Advances in Life Sciences



2021 vol. 13 nr. 1&2

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Life Sciences
Volume 13, Number 1 & 2, 2021

Bernd Krämer, FernUniversität in Hagen, Germany
Ramesh Krishnamurthy, Health Systems and Innovation Cluster, World Health Organization - Geneva, Switzerland
Roger Mailler, University of Tulsa, USA
Dirk Malzahn, OrgaTech GmbH / Hamburg Open University, Germany
Salah H. Mandil, eStrategies & eHealth for WHO and ITU - Geneva, Switzerland
Herwig Mannaert, University of Antwerp, Belgium
Agostino Marengo, University of Bari, Italy
Igor V. Maslov, EvoCo, Inc., Japan
Ali Masoudi-Nejad, University of Tehran , Iran
Cezary Mazurek, Poznan Supercomputing and Networking Center, Poland
Teresa Meneu, Univ. Politécnica de Valencia, Spain
Kalogiannakis Michail, University of Crete, Greece
José Manuel Molina López, Universidad Carlos III de Madrid, Spain
Karsten Morisse, University of Applied Sciences Osnabrück, Germany
Ali Mostafaeipour, Industrial engineering Department, Yazd University, Yazd, Iran
Katarzyna Musial, King's College London, UK
Hasan Ogul, Baskent University - Ankara, Turkey
José Luis Oliveira, University of Aveiro, Portugal
Hans C. Ossebaard, National Institute for Public Health and the Environment - Bilthoven, The Netherlands
Carlos-Andrés Peña, University of Applied Sciences of Western Switzerland, Switzerland
Tamara Powell, Kennesaw State University, USA
Cédric Pruski, CR SANTEC - Centre de Recherche Public Henri Tudor, Luxembourg
Andry Rakotonirainy, Queensland University of Technology, Australia
Robert Reynolds, Wayne State University, USA
Joel Rodrigues, Institute of Telecommunications / University of Beira Interior, Portugal
Alejandro Rodríguez González, University Carlos III of Madrid, Spain
Nicla Rossini, Université du Luxembourg / Università del Piemonte Orientale / Università di Pavia, Italy
Addisson Salazar, Universidad Politecnica de Valencia, Spain
Abdel-Badeeh Salem, Ain Shams University, Egypt
Matthieu-P. Schapranow, Hasso Plattner Institute, Germany
Åsa Smedberg, Stockholm University, Sweden
Chitsutha Soomlek, University of Regina, Canada
Monika Steinberg, University of Applied Sciences and Arts Hanover, Germany
Jacqui Taylor, Bournemouth University, UK
Andrea Valente, University of Southern Denmark, Denmark
Jan Martijn van der Werf, Utrecht University, The Netherlands
Liezl van Dyk, Stellenbosch University, South Africa
Lisette van Gemert-Pijnen, University of Twente, The Netherlands
Sofie Van Hoecke, Ghent University, Belgium
Iraklis Varlamis, Harokopio University of Athens, Greece
Genny Villa, Université de Montréal, Canada
Stephen White, University of Huddersfield, UK
Levent Yilmaz, Auburn University, USA
Eiko Yoneki, University of Cambridge, UK

International Journal on Advances in Life Sciences
Volume 13, Numbers 1 & 2, 2021

CONTENTS

New Way and Sensor Device for Investigation of Biosamples and Health Status of Humans and Animals

Tatiana Kuchmenko, Anastasiia Shuba, Ruslan Umarkhanov
Physical and Analytical Chemistry Department
Voronezh State University of Engineering Technologies
Voronezh, Russia
e-mail: tak1907@mail.ru, shuba1nastya@gmail.com, rus_270487@mail.ru

Darya Menzhulina
Endocrinology department
Voronezh Children's Clinical Hospital
named after N.N. Burdenko,
Voronezh, Russia
e-mail: darinochka_08@mail.ru

Anton Chernitskiy
Research Center for Clinical Pharmacology and Therapy,
Quality and Safety of Raw Materials and Products
All-Russian Scientific Research Veterinary Institute of
Pathology, Pharmacology and Therapy
Voronezh, Russia
e-mail: cherae@mail.ru

Abstract—**This paper discusses an application of a portable electronic nose based on an array consisting of 8 piezoelectric sensors with nanostructured solid-state coatings to detect volatile biomolecules secreted by nasal mucus and skin. A fundamentally new approach is proposed for a quick assessment of the status of the human body as a whole (normal, stress, inflammation) and the work of individual systems (reproductive, endocrine, digestive). This approach includes processing the output curves of sensors to assess of the qualitative and quantitative composition of the gas mixture of biomolecules secreted by the skin in the Zakharyin-Ged zone. Two algorithms for visualizing signals from an array of sensors are proposed. The first algorithm allows constructing the four "visual prints", reflecting the state of the human body (general, endocrine, energy, negative). The second algorithm constructs the health status sphere, which specifies the possible causes of deviation in health. Also, a portable electronic nose was applied in the veterinary field to assess the health status of calves' respiratory system. Unlike the traditional approach in diagnostics using a sensor array, one sample of nasal mucus was monitored for 5-9 hours with an interval of 2-3 hours. The new parameter is proposed to separate the diagnostic group of calves according to respiratory tract health. The changes in the composition of the gaseous phase over nasal mucus samples were considered. Using an electronic nose with nanostructured piezoelectric sensors, the speed and simplicity of measurement allow painlessly scanning the body for metabolic disturbances and estimating certain pathologies' presence and treatment effectiveness. The proposed algorithms and device for analysis of biosamples from humans and animals could be implemented to rapidly diagnose health status on farms, hospitals, and at home.**

Keywords- sensor; electronic nose; visualization; volatile compounds; metabolism; noninvasive diagnostic; skin; biofluid.

I. INTRODUCTION

Biosamples are complex objects to analyze. The problem of their analysis is connected with the absence of constant composition and in its almost instantaneous change when substances are excreted from the sample. Despite the emergence of new methods for analyzing and studying biostructures at the level of individual cells, the scheme remains traditional: a selection of biomaterial - sample preparation - detection of target components. The use of highly selective and effective methods of analysis (gas chromatography-mass spectrometry, high-pressure liquid chromatography.) suggests a special sample preparation, which can change the natural profile of the biosamples. The purpose of the analysis determines the method of sample preparation, but the integrity of the biological object is lost.

Furthermore, the results obtained by advanced analysis methods do not reflect a biological object's complex structure and behavior. Therefore, recently, in analyzing living objects (food, environmental objects, human and animal biosamples), complex methods with a multivariate analytical signal have been used more often. Such methods by methodology include artificial tongues, noses, eyes, and their portable variants [1][2]. The undoubted advantage of highly sensitive sensor systems with a rapid response is the ability to monitor the state of small volumes and masses of biological samples in a reasonably short time (from 2 to 9 hours). Given the lack of their contact with the environment (in vitro), primarily with oxygen, small volumes of biosamples, which means fast processes of changing their properties, open up a unique opportunity to obtain information about the status of the studied object, even if the specific methods for determining individual substances or

laboratory indicators (primarily microbiological) are unavailable.

A possible approach for assessing the body's status in the absence of biomaterial selection is to analyze the chemical composition of the gas, sweat of the skin in the zones of Zakharyin-Ged. Earlier, the presence of redness, peeling, rash, temperature changes in these areas was widely used as an additional parameter to confirm the malfunction of organs corresponding to these zones. The detection limits of modern methods of analysis, the complexity of the instrumentation of the most sensitive methods do not allow noninvasive scanning and determining the chemical composition of the gas phase of secretions from the skin. Therefore, creating an integrated system for scanning a volatile metabolome using a device with a sensitive detector to biomolecules of normal and disordered metabolism, inflammation, and microbial metabolites is actual for now.

The purpose of this paper is the development and application of a new mobile device based on piezoelectric sensors (portable electronic nose) for assessing the health status of organs and systems of humans and animals by analyzing the volatile metabolome.

We will demonstrate our approach in two ways: 1) analysis of nasal mucus samples of calves for the diagnosis of respiratory diseases and 2) characterizing the health status of humans by skin odor in the Zakharyin-Ged zones.

Further, in Section II, the related works and state-of-the-art technologies for health status assessment are described. In Section III, the features of the experiment, description of biosamples, and methods of analysis are presented. Section IV contains technical characteristics of the proposed device, characteristics of used sensors and their coatings, a description of the procedure for obtaining and recording output data of the sensor array. Section V shows the results of applying the proposed portable electronic nose (e-nose) for solving diagnostic problems according to the purpose of the work. Section VI is devoted to conclusions and perspectives of development.

II. Related work

There are some state-of-the-art approaches, including portable devices, for assessing the state of the body using electronic noses and various data processing methods: for exhaled breath air [3][4][5], for the analysis of biomaterials (blood, urine, secrets of the endocrine glands and others) [6][7][8]. The different approaches for assessing health status by skin based on electrochemical or optical methods are proposed [9][10][11].

At the same time, several devices have been developed to detect body odor with different types of sensors to discriminate volunteers and estimate changes in their body odor during physical activities [12][13]. There are examples of developing unique construction with sensors to diagnose tuberculosis [14], renal dysfunction [15], heart failure [16] by volatile compounds from the skin.

The mentioned researches are based on training the sensor array by volatile compounds, measuring the body odor, and calculating the relative of normalized signals used for processing by different multivariate data analysis methods for classification or discrimination. However, the investigation of healthy people in different conditions is missed. Also, the application of such complex multivariate algorithms is not understandable for users and difficult to use in a hospital.

We want to propose a sensor device with the simple measurement, calculation, and visualization of sensors signals with the verbal characteristics of state understandable by any user. The new technique of estimating respiratory organ state in animals and differentiating the disease on farms based on a simple parameter of the sensor array was one of the investigation tasks.

III. Materials and Methods

This section represents the additional methods used to diagnose the health state of animals and humans.

A. Diagnosis of Respiratory Diseases in Calves

The 17 samples of nasal mucus from calves (10-20 days of life), both with signs of respiratory system damage and conditionally healthy, were analyzed. A sampling of nasal mucus was carried out with sterile cotton swabs in individual sterile containers. The time from sampling to analyzing on e-nose was taken into account.

The calves were clinically studied in detail using a point system (WI score) developed at the University of Wisconsin, Madison (USA) [17]. Samples of nasal secretions were investigated in the laboratory at the All-Russian Scientific Research Veterinary Institute of Pathology, Pharmacology and Therapy using bacteriological and molecular genetic (PCR) analysis for infectious rhinotracheitis, parainfluenza-3, viral diarrhea-disease cattle mucous membranes, rotavirus, adenovirus, chlamydia, pathogenic mycoplasmas (*M. bovis, M. bovirhinis*). Also, the hematological indicators in the blood (leukogram, haptoglobin concentration) were determined to confirm inflammation.

For the isolation of cultures and typing of microorganisms from the nasal mucus samples, meat and peptone broth, milk salt, enterococcal agar, Endo medium, blood agar, glucose-serum broth and agar produced by Research Center for Pharmacotherapy (St. Petersburg, Russia) were used. The isolated Escherichia coli were typed in an agglutination reaction using O-serums.

B. Characterizing the Some Deviation from Normal Status by Human Skin Odor

The forearm area of human skin was chosen to analyze the volatile metabolome by a portable electronic nose. Over 100 volunteers aged 3 to 80 years took part in the investigation for two years (72 female, 35 men). The volunteers periodically were clinically (visits to physicians) and laboratory (general analysis of blood, urine, biochemical analysis of blood (glucose, cholesterol, some hormones: thyroid-stimulating hormone, free thyroxine, progesterone, follicle-stimulating hormone, estradiol, lutein hormone, adrenaline, cortisol)) tested to control health status.

Clinical monitoring of health status and measurement of volatile metabolome were performed on average once a month in the absence of pathology. When volunteers had

diseases (acute respiratory viral infections, bronchitis) or exacerbation of chronic pathologies (diabetes, heart failure), monitoring of the volatile metabolome of the skin and the main clinical and laboratory parameters (glucose level, heart rate, blood pressure) was carried out 2-3 times a day until the health state stabilized due to the therapy. The state "inflammation" corresponds to both an increase in laboratory parameters (leukocyte count, erythrocyte sedimentation rate) and some cases, a change in biochemical parameters indicating the inflammation in organs - lactate dehydrogenase, aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, C-reactive protein. During the experiment, we monitored inflammation with different diagnoses: arthritis (n=2), gastritis (n=5), sinusitis, pyelonephritis, pneumonia (n=2), adnexitis (uni- and bilateral, n=3), endometritis (n=3), osteochondrosis (n=2), diabetes (n=2), inflammation of soft tissue and tendon, systemic lupus erythematosus.

For conditionally healthy volunteers, the results of laboratory tests corresponded to the typical reference values, and the symptoms did not match with clinically significant ones for illness. Clinically not diagnosed conditions, so-called descriptive states (tiredness, excitement, agitation, stress, lack of sleep, spasm, pain), were recorded from the volunteers' words.

In total, by March 2020, we performed 964 measurements.

IV. Description of Portable E-nose and Measurement Technique

The characteristic of used piezoelectric sensors and coatings, hardware and software of a portable e-nose, as well as technique of measurement, are presented below.

A. Making of Piezoelectric Sensors

We used piezoelectric quartz resonators (PQR) with a natural frequency of 14 MHz with an established linear response with a film mass on its electrodes up to 20 μg/cm^2. The array contained eight piezoelectric sensors with electrodes covered by films of carbon nanomaterial, hydroxyapatite, zirconium salts of different mass (1-5 μg) (NANO-BIO array).

1) *Characteristics of the Used Sorbents*

Hydroxyapatite (HA) $Ca_5(PO_4)_3OH$ was obtained by the sol-gel method developed at Nizhny Novgorod State University, named after N. I. Lobachevsky [18] and optimized by us to obtain nanostructured coatings with good sorption properties.

The reaction was carried out according to the following equation:

$5Ca(NO_3)_2 \cdot 4H_2O + 3H_3PO_4 + 10NaOH \rightarrow Ca_5(PO_4)_3OH + 10NaNO_3 + 29H_2O$

To a solution of calcium nitrate (2 mol/dm^3) prepared from $Ca(NO_3)_2 \cdot 4H_2O$ in bidistilled water has added a solution of H_3PO_4 in the amount necessary to maintain the ratio Ca/P = 5/3. The resulting solution was thermostated for one hour at 37 °C; then its pH was adjusted to 7-8 using a NaOH solution with a concentration of 2 mol/dm^3. A hydroxyapatite sol began to form at pH = 4. The reaction mixture was kept at 37 °C for 1 hour. Then, the resulting gel was centrifuged and dried in air. The obtained sorbent can be stored for at least 0.5 years in airtight conditions. Multi-walled carbon nanotubes (CNT) were obtained in the Institute for Extra Pure Materials of the Russian Academy of Sciences (Chernogolovka, Russia) by gas-phase chemical deposition during ethanol pyrolysis. Nickel was used as a catalyst; deposition temperature was 450-500 °C. Then nanotubes were washed with HNO_3 concentrated (Reachem, Russia). The solvent for the suspension of HA and CNTs was chloroform. Zirconium nitrate (ZrO) ($ZrO(NO_3)_2 \cdot 2H_2O$) (chemically pure) was obtained from aqueous solutions containing zirconium and nitrate ions (Reachem, Russia). The solvent for the suspension of ZrO was acetone. According to the preliminary experiment results, these sorbents are selective and sensitive to volatile metabolites of bacteria and inflammation [19]. Also, the variation of sorbent mass affects the selectivity and sensitivity of volatile organic compounds (VOCs) micro weighting [20].

Therefore, the main array of sensors consists of:

Sensors 1, 8 – CNT phases of different masses.

Sensors 2, 7 – phases of ZrO of different masses.

Sensor 3 – dicyclohexane-18-Crown-6, DCH-18C6.

Sensors 4, 5 – HA phases of different masses.

Sensor 6 – polyethylene glycol succinate, PEGS.

Additionally, for the analysis of the gas phase over the nasal mucus samples, we used an array with polymer polar films (additional sensor array), which we previously used in our work with other biological samples [21][22], namely polyethylene glycol 2000 (PEG-2000), polyoxyethylene sorbitan monopalmitate (Tween), t-octylphenoxypolyethoxyethanol (TX-100), dicyclohexane-18-crown-6 (DCH-18C6), polyethylene glycol sebacinate (PEGSb), bromocresol blue (BCB). These polymeric compounds were dissolved in organic solvents according to their polarity: PEG-2000, Tween, TX-100 – in acetone, DCH-18C6, PEGSb - in toluene, BCB – in ethanol. The solutions of polymeric sorbents were immediately used to form films on the piezoelectric resonator electrodes without their pretreatment with ultrasound (step 3).

2) *The Method of Forming Films on the Surface of the Piezoelectric Quartz Resonator*

The films were uniformly deposited to the electrodes of PQRs, fat-free with acetone or chloroform, by immersion in solutions of sorbents.

The forming of films on the electrodes of resonators was performed using the following procedure:

Step 1 – the measurement of the initial oscillation frequency of the piezoelectric resonator (10 or 14 MHz) F_0, Hz with an accurate record, for example, 9999280 Hz;

Step 2 – suspension was prepared in the beaker as dissolution of sorbent (0.5 g) in 10 ml of solvent;

Step 3 – processing in an ultrasonic bath for 15 minutes at a power of 90 W;

Step 4 – exposure of the quartz piezoelectric resonator in suspension for 15 s;

Step 5 – drying the coating in an oven (40 minutes at a temperature from 50 ° C) in the holder vertically;

Step 6 – the measurement of the oscillation frequency of the sensor, calculation of the coating mass (Δm) according to the Sauerbrey equation [23]:

$$\Delta m = \frac{\Delta F \cdot 0.2}{2.27 \cdot 10^{-6} \cdot F_0^2} \quad (1)$$

where ΔF is the change in the oscillation frequency of the quartz plate of the resonator after film deposition and removal of an unbound solvent, MHz;
$2.27 \cdot 10^{-6}$ – calibration constant of PQR at normal condition, cm^2/g;
F_0 — base oscillation frequency of the PQR, MHz;
0.2 – the area of electrodes of PQR, cm^2.

B. Characteristic of Portable E-nose

The portable device for diagnosing the status of humans and animals is a miniature case, consisting of two functional parts (Fig. 1): head 1 and the protective part of the body 6. A microprocessor 2 with terminals for sensor mount sockets, a block for fixing and transmitting information 3 to the recording device (laptop, tablet, personal computer). The sockets are located in cover 4, into which removable sensors 5 are mounted on the outside, separated from the environment by the protective part of the body 6, which is tightly attached to the head 1. Optionally e-nose is supplemented by an internal gas-permeable gasket 7, which separates the body's sensor and free air region of the body 6.

The protective nozzles of various types 8 from inert materials (fluoroplastic) are used depending on the nature of the analyzed sample to reduce interfering factors (external fluctuation in airflow, temperature, air composition in the near-sensor space).

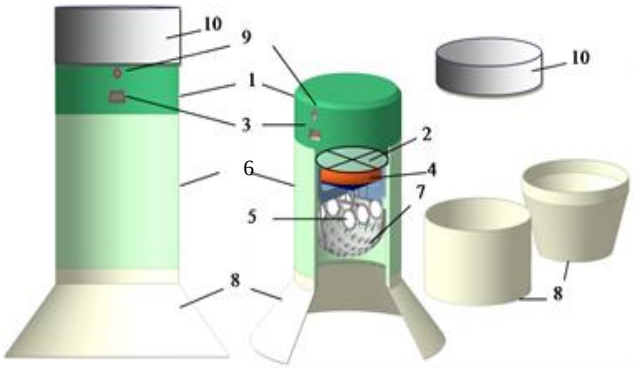


Figure 1. General view of the 3D model of the portable e-nose for diagnostics: 1 – head; 2 – microprocessor with terminals for sensor mount sockets; 3 – block recording and transmitting information to a recording device of any type; 4 – cover; 5 – removable sensors; 6 – protective part of the body; 7 – internal gas-permeable gasket; 8 – nozzles; 9 – power supply from the electricity; 10 – removable battery.

The e-nose is powered by either an electronic device via a USB cable, either from electricity 9 or a removable battery 10. The developed portable device is an electronically counting frequency meter with 8 channels to measure the oscillation frequency of BAW-type PQRs with a base oscillation frequency of 5 to 20 MHz with a resolution of 1 Hz a time interval (step) of 1 second. The electronic counting frequency meter is switched on in the network (220 V); it warms up for 10-15 minutes. In this case, the sensors should be in the device to reduce measurement errors. Nevertheless, their subsequent inclusion is also possible. It takes about 5-10 minutes to stabilize the baseline of the oscillation frequency of the quartz plate.

To simultaneously record (read) the oscillation frequency of each sensor every second for a specific time interval (from 1 s to a maximum of 6000 s), the device is connected to a computer via USB cable, and other connection options are possible (via Wi-Fi, Bluetooth).

Operating conditions and technical specifications of E-nose are presented below:

- Ambient temperature from +15 to +35 ° C.
- Humidity is up to 98% at temperature +35 °C.
- The device is powered by an alternating current with a voltage of 220 ± 22 V and a frequency of 50 ± 0.5 Hz.
- Frequency range of using PQR 4 MHz – 20 MHz.
- The reference frequency oscillator is 4 MHz.
- Overall dimensions – 38x120x170 mm.
- Weight with a cover – 0.40 kg.

C. Specification of Software

The responses are recorded in the instrument software, which saves the measurement and converts it into analytical information – a change in the oscillation frequency of each resonator individually at each measurement moment relative to the starting point of measurement (-ΔF, Hz). The total output curve is displayed in a set of chronograms for all resonators installed in the e-nose (Fig. 2).

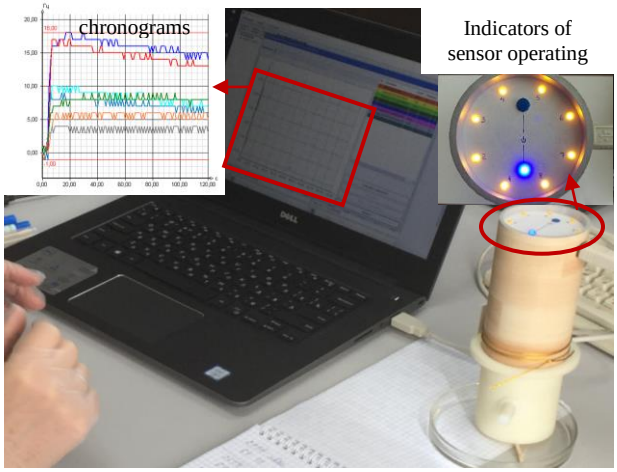


Figure 2. E-nose connected with the laptop when measuring the sample of nasal mucus.

During the interaction of vapors with the surface of the piezoelectric sensors, sorption occurs on the films or electrodes, resulting in frequency changes. Individual colors reflect a change of the base oscillation frequency in time for each of the eight piezoelectric sensors (Fig. 2).

In the developed software based on chronograms, the "visual print" is constructed using different algorithms depending on the purpose of analysis. The quantitative characteristic of "visual prints", therefore, the total amount of volatile substances excreted by samples and sorbed by piezoelectric sensors, is the area of "visual print" ($S_{v.p.}$, Hz·s). The "visual print" area is calculated in software as a sum of definite integrals of time dependence the signals of sensors during measurement (chronograms).

Additionally, in software, the parameters of sorption (*A(i/j)*) are calculated by equation (2), which can be used for the identification of volatile substances in the gas phase over samples [24][25] or to describe additional analytical information about sample characteristics.

$$A(i/j) = \Delta F_i / \Delta F_j , \quad (2)$$

where $\Delta F_{i(j)}$ are the maximum responses of piezoelectric sensors during the sorption of the gaseous phase over the biosamples or skin.

D. Technique of Measurement

The gas phases over nasal mucus samples and skin were studied with the front input method into the detection cell. The open detection cell of the device contacted the forearm area (20 cm^2) of the skin or Petri plate with mucus sample to analyze the volatile substances excreted by the biosamples. The registration time of the sorption of volatile substances excreted by biosamples was 80 s, the registration of desorption was 120 s. Thus, the total time of one measurement was 200 s. The frontal analyte input method in detail is described in [26][27].

Preliminary, the electronic nose was trained in the same technique of measurement by 45 individual substances of normal and pathogenic metabolism: C_1-C_5 alcohols, ketones, C_5-C_7 cyclic ketones, aldehydes, N-, S-containing aldehydes, C_1-C_5 carboxylic acids; primary, tertiary, cyclic amines, O-containing amines.

Analysis of the gas phase above the nasal mucus samples using an additional array of sensors with films of polymer sorbents was carried out immediately after measurements with the main array of sensors with solid-state nanostructured films by the same technique. A detailed description of the measurement technique of nasal secretion samples is provided elsewhere [27].

V. Results and Discussion

Examples of using developed sensor device for medical applications are provided in this section, including sensor data processing algorithms to extract information concerning health status.

A. Diagnosis of Respiratory Diseases in Calves

We selected three diagnostic groups of calves based on the results of clinical studies, the determination of hematological and biochemical markers of inflammation (leukocytosis, an increase in the concentration of haptoglobin in the blood serum), pathogens of viral and bacterial infections accompanied by damage to the respiratory system: 1 - "healthy respiratory system" (n=4), 2 - "with signs of respiratory disease" (n=7), 3 - "with the subclinical course of respiratory diseases" (n=6).

A natural change in the mucus composition taken at a weekly interval can reflect only significant changes in the condition – for example, a vivid manifestation of the inflammatory process. Unlike state-of-the-art approaches to diagnosing respiratory diseases by one measurement of biosamples [6-8], for the first time, it has been proposed to monitor one sample for 5-9 hours with an interval of 2-3 hours. It allows recording changes in the state at the micro-level associated with microbiological contamination of the sample or its absence. The areas of "visual prints" were calculated for all samples of nasal mucus. Early it was shown that the values of the "visual prints" area correlate with biochemical indicators of inflammation characterizing the disease of respiratory organs in calves [28].

The results of one-day monitoring of nasal mucus samples can be divided into three groups (Fig. 3).

1) Positive (increasing) dynamics of changes in the value of the integral analytical signal of the sensor array (area of "visual print") - indicates the destruction of nasal mucus and production of a large number of volatile compounds, including microorganisms metabolites.

2) The negative (decreasing) dynamics of the change in the value of the analytical signal of the sensor array indicates the decreasing of volatile substances excreted from nasal mucus due to increase of its viscosity by the high level of proteins, mucin, which is observed in the acute phase of respiratory disease [29].

3) The almost constant value of the integral signal from the array of sensors, which is observed at the first sign of respiratory disease (subclinical course), indicates that the excretion of substances at the destruction of nasal mucus and the production of metabolites by microorganisms are not so active.

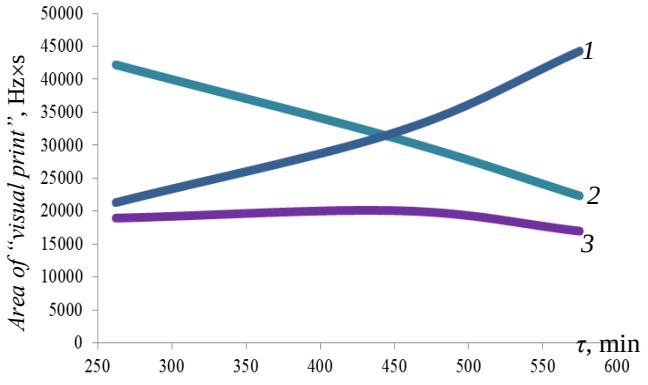


Figure 3. Total "visual prints" area of signals of the sensor array in vapors of nasal mucus of calves with different diagnoses: 1 –healthy respiratory system, 2- with signs of respiratory disease, 3 - with the subclinical course of respiratory disease.

When a sample of nasal mucus from an animal with the subclinical course of the respiratory disease is taken within 5 minutes after the first selection, then the time dependence of area of "visual print" for this sample will be like for conditionally healthy (curve 1, Fig. 3). These facts are in good agreement with the data on the formation of local protection of the respiratory tract of calves [29].

The maximum exposure time of nasal mucus samples from the moment of sampling to analysis by the array of sensors should not exceed 7 hours; with a longer exposure time, the differences in the concentration of volatile substances between diagnostic groups are disappeared.

The study of the sorption of the gas phase over nasal mucus samples on an additional array of sensors with polymer hydrophilic films coatings showed the same trend with more significant differences between groups, since sensors with polymer films are more sensitive than solid-state sensors [30], although with a shorter operation time [31]. Consequently, during the delayed analysis of nasal mucus samples, it was proposed to calculate the ratio of the sum of sensor signals ($\Sigma F_{ref.}$) with hydrophilic polymer coatings of the first measurement (1mtesurement) to the one carried out after 2 hours for each sample using the equation (3):

$$\sum F_{ref.} = \frac{\sum_{i=1}^{6} \Delta F_{i(1measurement)}}{\sum_{i=1}^{6} \Delta F_{i(after\,2hous)}} \quad (3)$$

The deviation in sensor array responses due to internal factors (microelectronic scheme, reproducibility of sensors surface) is no more than 5 % [16]. Therefore, if indicator ΣF_{ref} deviates by more than 15%, the differences in gas phases over biosamples are statistically significant. Consequently, based on the established trend by one-day monitoring of nasal mucus samples, the group "healthy respiratory system" – $\Sigma F_{ref.} \leq 0.85$, and for the group "early signs of respiratory disease" – $\Sigma F_{ref.} \geq 1.15$, the values $F_{ref.}$ 1.00 ± 0.15 is assigned to the "with the subclinical course of respiratory diseases" group.

Table I shows the proposed classification based on the value of indicator $\Sigma F_{ref.}$ and actual diagnostic group for the 17 studied samples. The proposed indicator was used to rank additional 4 samples of nasal mucus of calves to check the implementation of such classification. These calves were also clinically and laboratory tested to determine the diagnostic group (in italics at the bottom of Table I). The proposed ranking for 4 samples also correlates with the established diagnostic group. Hence, samples of nasal mucus from calves with or without infections of the upper respiratory tract (nasal cavity, trachea, bronchi) can be differentiated using the $\Sigma F_{ref.}$.

Thus, several measurements of one biosample during the day and simple indicator $\Sigma F_{ref.}$ allow clarifying the degree of damage respiratory tract in calves associated with the contamination sample.

TABLE I. RELATIVE CHANGE $\Sigma F_{REF.}$. IN THE TOTAL AMOUNT OF VOLATILE SUBSTANCES IN THE GAS PHASE OVER NASAL MUCUS SAMPLES FROM CALVES.

Sample	**ΣF_{ref}**	**WI**	**Results of bacteriological, mycological, PCR investigations of nasal mucus samples**	**CFU /ml[a]**	**Diagnostic group**
2675	0.78	3	*E. coli, Ent. faecium, yeast-like fungi*	800	1 – healthy respiratory system
2664	0.85	2		200	
2444	0.82	2		350	
2442	0.85	3		150	
2	1.26	8	*E. coli O142, Ent. faecium, Ent. faecalis, Pseudom. aeruginosa, yeast-like fungi, Adenovirus, Mycoplasma bovis*	3500	2 – with signs of respiratory disease
3	1.29	8		2400	
2679	1.25	5		600	
2667	1.31	5		1500	
2669	1.40	7		4500	
2670	1.47	8		6000	
2671	1.44	8		5600	
2677	0.96	4	*E. coli O142, Ent. faecium, Ent. faecalis, yeast-like fungi, Adenovirus*	700	3 – with the subclinical course of respiratory disease
1	0.98	5		540	
4	1.09	4		200	
5	1.13	4		450	
2666	0.99	4		850	
2668	0.92	4		750	
2672	0.99	4	*E. coli, Ent. faecalis*	1000	3
2665	1.37	6	*E. coli, Ent. faecium, Adenovirus*	3400	2
2663	1.15	4		2300	3
2669	1.44	8	*E. coli, adenovirus, yeast-like fungi*	4200	2

a - colony forming units (CFU) of mesophilic aerobic and facultative anaerobic microorganisms.

Trends in changes in the composition of the gas phase of biosamples during a day were established based on assessing changes in proportions of VOC classes in the gas phase over samples $w_{i,ref.}$ calculated by the formula (4):

$$w_{i,ref.} = \frac{w_{i,(followingmeasurement)}}{w_{i,(previousmeasurement)}}, \quad (4)$$

where w is a proportion of VOCs calculated via equation (5) [32]:

$$w_i = \frac{\Delta F_i}{\sum_{i=1}^{8} \Delta F_i} \quad (5)$$

where $\sum_{i=1}^{8} \Delta F_i$ is sum of sensors signals in an array (main or additional).

For samples of nasal mucus from calves from the group "healthy respiratory system" in the first 4 hours, changes in the proportions of different classes of VOCs in the gas phase were observed within 25%. In the next 3-4 hours, the proportion of volatile aromatic, cyclic amines increases several times and the proportion of aldehydes, ketones, organic acids decreases by 30-60% (Table II).

TABLE II. RELATIVE CHANGE IN THE PROPORTION ($W_{i,REF.} \pm 0.08$) OF INDIVIDUAL CLASSES OF VOCS IN THE GAS PHASE OVER SAMPLES OF NASAL MUCUS DURING ONE DAY MONITORING

τ, min	Diagnostic group (samples)	$W_{i,ref}$; *i* – sensor coating					
		HA[a]	*CNT*	*ZrO*	*BCB*	*DCH-18C6*	*PEGS*
124	1 (2675, 2444, 2442, 2664)	1.30	1.27	1.37	0.74	1.41	1.27
246		1.19	1.14	1.22	0.79	1.22	1.16
377		1.01	0.88	1.00	1.14	0.99	1.05
453		0.53	0.76	2.86	1.39	0.75	1.79
517		0.69	0.67	0.63	0.91	0.56	0.64
126	2 - (2679, 2, 3, 2667, 2669, 2670, 2671)	1.19	1.03	1.06	0.69	1.08	1.10
136		1.19	1.09	1.22	0.80	1.22	1.25
258		1.37	1.39	1.45	0.89	1.45	1.37
380		1.59	1.49	1.52	0.91	1.45	1.61
453		1.14	1.08	1.11	2.44	1.09	1.19
557		0.93	0.98	0.88	0.92	1.02	0.88
126	3 - (2677, 1, 4, 5, 2666, 2668	0.63	0.57	0.53	0.60	0.63	0.60
256		1.35	1.15	1.33	0.80	1.41	1.33
377		1.39	1.25	1.32	1.15	1.25	1.32
498		1.27	1.15	1.19	1.69	1.23	1.25
503		0.91	1.03	1.01	0.91	0.85	1.01

a - Predominantly adsorbed VOC classes on sensor coatings: DCH-18C6 – hydroxyl, carboxylic acids; BCB – aromatic, cyclic amines, nitro compounds; PEGS – aromatic compounds, oxy-amines; CNT – C_1-C_5 primary, secondary alkylamines, ammonia; HA – C_2-C_5 alcohols, acetals, branched and cyclic ketones, amines; ZrO – alkylamines, cyclic amines, oxy-amines.

For samples of nasal mucus from calves from the group "with signs of respiratory disease", conversely, in the first 2-3 hours after sampling in its gas phase, the proportion of volatile aromatic, cyclic amines, nitro compounds is significantly decreased and the proportion of oxy- and hydroxyacids is increased, with an increase in the proportion of ketones, oxy-compounds in the next 4 hours. For samples of nasal mucus from calves from the group "with the subclinical course of respiratory diseases", a decrease in the proportions of all VOC classes in the gas phase over samples is observed in the first 2 hours after sampling. In the next 6 hours in the gas phase, the proportion of VOCs increases up to 35% with the predominance of ketones, acids, amines (Table II). Thereby, parameters $\Sigma F_{ref.}$ and $w_{i,ref.}$ can be applied for estimating respiratory organ state in animals and differentiating the disease on farms.

B. *Characterizing the Some Deviation from Normal Health Status by Human Skin Odor*

A forearm zone of skin was chosen to scan the whole organism's health status according to information about the diagnostic significance of the Zakharyin-Ged zone. The primary measurement database contained the responses of 8 sensors in 200 seconds (1590-1600 signals).

Based on the analysis of the skin odor of 100 volunteers in various states according to their words and the results of laboratory tests, a primary algorithm of “visual print” construction has been developed for linking the features of the forms of “visual prints” with the human condition and possible causes of deviation (Table III, Fig. 4).

"Visual print" is a circular diagram of the sensor signals at specific points in the process of sorption and desorption of volatile substances from the gas phase over biological samples.

TABLE III. ALGORITHM FOR CONSTRUCTING “VISUAL PRINTS” OF SIGNALS FROM AN ARRAY OF 8 SENSORS FOR ASSESSING THE HEALTH STATUS OF PEOPLE

Name of an algorithm of “visual print” construction (characteristic)	Time of recording the sensor responses, s / number of sensors used to build a “visual print”
"General state" (the most complete information about reproductive, digestive systems)	30, 45, 60, 80, 100, 120, 180 / 8
"Energy" (reflects the strength and intensity of the metabolome part, which shows the ability of the body to act)	110, 120, 130, 140, 150 / 3
"Endocrine system" (reflects malfunctions of the endocrine glands, primarily the pancreas)	10 20 30 60 / 4 Additionally, the parameter is calculated: $\gamma = \Delta F_4(60\ s)/\Delta F_4(20\ s)$. At $\gamma \leq 2$ pathology occurs
"Negative" (the severity of destructive processes in the body)	20, 30, 170, 180 / 8 Additionally, parameters are calclulated: $\beta_1 = \Delta F_4(20\ s)/\Delta F_4(170\ s)$. $= \Delta F_4(30\ s)/\Delta F_4(180\ s)$. At $\beta_1, \beta_2 \leq 2.5$ pathology occurs

Integral analytical signals of the sensor array when measuring the volatile metabolome of the forearm skin are visualized in the form of 4 “visual prints” (general state, energy, endocrine, negative, Table III). These multidimensional responses represent a particular, differentiated part of the chronograms of the most informative sensors, reflecting the sorption features of volatile organic substances established in the preliminary training experiment. "Visual prints" differ in the set of time points of recording the sensor responses (mask) and the number of sensors.

The largest number of points on the "visual print" corresponds to the "General state" algorithm - all 8 sensors and 7 time points on the chronograms of complete measurement (Table III). It reflects the complete information about the nature and concentration of volatile substances in the gas phase above the skin.

The "Energy" algorithm includes the desorption time on the most sensitive sensors. It estimates the rate of desorption of the volatile metabolome of the skin, presumably associated with the production of adenosine triphosphoric acids and, as a result, production of the energy by the cell.

The "Endocrine system" algorithm uses only the beginning of the sorption of volatile substances, where the difference between volatile products of hormone metabolism and other substances is maximally manifested.

The "Negative" algorithm takes into account the initial sorption time and the final desorption time of substances, allowing the maximum differentiation of heavy compounds (cyclic and aromatic amines, ketones), which are associated with the development of negative processes (oncology, inflammation, tissue and organ damage).

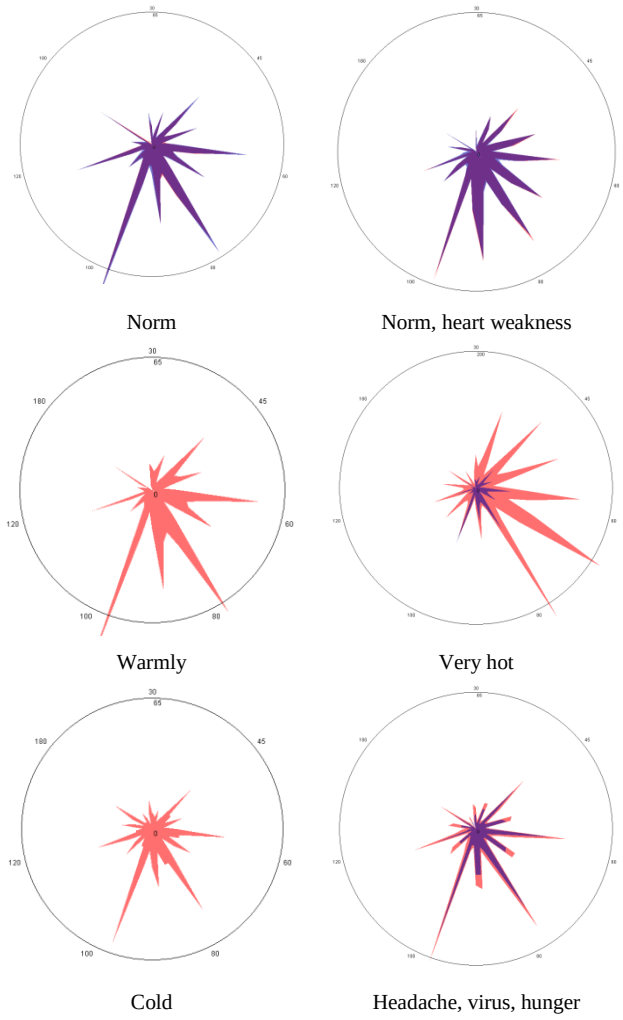


Figure 4. Statistically significant typical changes in the shape of the integral signals of the sensor array of the portable e-nose ("General state" algorithm) for different conditions of one person (for left forearm – the blue color and for right –red color).

Masks and a set of sensors for constructing multidimensional visual signals were selected based on the results of preliminary training of the array of sensors for individual substances – biomarkers of metabolic disorders and the disfunction of individual organs, the presence of inflammatory processes in the gaseous phase of biosamples [33][34][35][36][37][38][39][40][41][42][43][44]. The proposed algorithms are different from the traditional approach to visualize the sensor signals used in the state-of-the-art methods [4]. Moreover, such visualization is more straightforward and more evident for the user.

Based on the data obtained for 400 measurements of volatile compounds secreted by the forearm zone of skin, the geometric characteristics of "visual prints" and their typical shape were determined, corresponding to the reference boundaries of the norm for different categories of people (men, women, adolescents, dependence on the type of central nervous system). The geometric shape of the "visual print" is characterized by its rays, area, and the ratio of parts relative to the line of symmetry (passes through 80s). Changes in the geometric shape of typical "visual prints" at different time intervals (for example, the first three points) are signs of a change in the composition of the volatile metabolome. The features of these changes were compared with a verbal description of the condition (headache, stress, fatigue, hunger, and others), clinical examination and the presence of disease symptoms (runny nose, cough), parameters *A(i/j)* for substances. Fig. 4 shows the comparison of "visual prints" forms for left and right hands, used for visual assessment of descriptive states. As a result, statistically reliable responses and "visual prints" forms were determined, which correspond to the physically normal functioning of the body (norm), stress, tiredness, inflammation, weakness, in total no less than 17 states. The deviation of the geometric shape of the "visual print" from the typical one is the first sign of a change in health state.

In a normal state, the qualitative and quantitative composition of the volatile metabolome on the left and right hand practically does not change - the shape and area of the "visual print" practically do not differ (Table IV, Fig. 4). During an extended stay in a warm or cold room, when a person can describe his condition as "warm" or "cold", a regular increase or decrease in the amount of excreted substances occurs, while the shape of the "visual print" does not change (Fig. 4). When the body overheats, changes in the volatile metabolome are more significant, the shape of the "visual print" changes, associated with a more intense excretion of VOCs to maintain homeostasis.

TABLE IV. THE MOST TYPICAL EXAMPLES OF CHANGES IN THE AREA OF THE "VISUAL PRINTS" WITH SOME CHANGES IN HEALTH STATUS

Person's state	**The trend of changing and the relative difference in the parameter $S_{v.p.}$,%**
Norm	Differences between the left and right hands - 5-10%, for the left more than for the right
Norm, after eat	11-30% more for the right hand than for the left one
Norm, easy hunger	For the right hand less than for the left one by 11-25%
Norm, severe hunger	If the gallbladder is malfunctioning, a response increase of 15-20% for the right hand is observed due to the excreted of propandial
Cold, temperature for a long time 15-20°C	20% (before meals) – 10% (after meals), the shape of "visual prints" changes
Headache, toothache, other pain, spasm	10-35%, the shape of "visual prints" changes
Increase the air temperature up to 26-30°C	10-30%, at perspiration till 500%
Virus, malaise without fever	Decrease on the right hand to 38-40%
Menstruation	An increase in the side of the working ovary by 15-25%.
Fatigue, heart failure	The left side is smaller than the right in normal to 12-20%, the shape of "visual prints" changes
Bronchitis, inflammation	The difference between left and right Zakharyin-Ged zone of bronchi for acute bronchitis - 40 %, recovery – up to 15 %, healthy lung and bronchi – up to 5 %

In the presence of pathology and severe symptoms of the disease (viral infections, headache, heart weakness, Fig. 4), the shape of the "visual print" changes in different ways compared to the normal state, which is associated with the VOCs of intoxication of the organism. If we normalize for one person its average quantitative parameter of the smell trace (area of "visual print") by 500 measurements in different periods during two years, then the general nature of the bias of this indicator will obey the laws presented in Table IV. The geometric shape of the "visual print" is strictly individual for each person and the calculated parameters (*A(i/j)*) – for health status (Fig. 4, Table V).

The number of changed parameters and indicators depends on the degree of deviation of the health state from the norm and correlates with the severity of destructive processes during the disease. The shape of "visual print" is influenced to a greater extent by the body health, a psycho-emotional state during measurement, and gender, to a lesser extent by age. The reference limits of parameters *A(i/j)* for the degree of deviation from the norm have been established for each descriptive state, for instance: the low level of tiredness, the middle level of tiredness, the high level of tiredness, and the critical level of tiredness. Besides, the appearance of individual substances in the descriptive states can be evaluated, comparing values of parameters *A(i/j)* for 45 substances of normal and pathogenic metabolism with those for descriptive states. The ranges of parameters *A(i/j)* for identifying VOC classes and individual compounds are described in the works [25][26]. For an array of eight sensors, the maximum number of parameters *A(i/j)* is 28. Of these, the values of 21 parameters are statistically significant and reliably (according to Pearson's and Student's criteria) correlate with changes in clinical and laboratory test results or the verbal description of the state by volunteers. Namely, four parameters *A(i/j)* are associated with inflammatory processes, 4 – with malfunctions of the gastrointestinal tract (spasms of bile streams, slowed metabolism in the liver, pancreatitis, ketoacidosis), 3 – with changes in hormones or their metabolites (growth hormone, adrenaline, cortisol, sex hormones, hypothyroidism). Several parameters reflect descriptive conditions, for example, fatigue and its severity up to critical exhaustion (Table V).

For the convenience of deciding the health status of the organism, the state sphere is constructed using parameters *A(i/j)* in the software. This program is written in a high-level Java language as an Android application, designed to interact with the electronic nose device (Fig. 5) and based on previous our development [45][46]. Two algorithms have been developed to analyze received signals from an array of 8 sensors. The first allows getting the measurement results understandable to any user. Initial processing of the measurement results is carried out according to the maximum sensor responses. The body's health state decoding is displayed on the screen together with a complete set of calculated parameters in the form of a sphere (health status sphere), where each parameter corresponds to its sector (Fig. 6).

TABLE V. THE VALUE OF SOME PARAMETERS OF SORPTION *A(I/J)* IN VARIOUS STATES

Parameter A(i/j)	**The numerical range of parameter values**			
	Norm	***Description of the deviation state***		
Sector color	green	yellow	red	burgundy
A(1/5)	< 0.75	0.75 – 0.94 Stress, body weakness	> 0.90 Hormone imbalance	> 0.81 Stress, weakness, severe inflammation
A(1/7)	≤ 1.90	> 1.9	> 2.3 Adrenaline, cortisol, severe stress	-
A(1/2)	> 1.15	0.90 – 1.14 Inflammation, very hot	< 0.9 Alcohol, ketones	-
A(1/4)	-	-	<0.30 Sharp pain, inflammation in the organ (depends on the projection of which organ by Zakharyin-Ged zone was measured), weakness, exhaustion	0.38 – 0.42 Risk of diabetes, weakness, exhaustion of the body
A(2/4)	≤ 0.1	-	0.25 – 0.30 Ketones, sugar is above normal, hormones are very unbalanced	0.16 – 0.24 Weakness, Exhaustion
A(2/5)	> 0.52	0.48 – 0.52	<0.48 Ketones, high level of glucose, endocrine system disorders	-
A(2/6)	0.62 – 0.73	0.75 – 1.2 Weakness	1.3 – 1.5 Spasm in the gastrointestinal tract, pain	-
A(3/5)	5.3 – 3.2	3.0 – 1.6	1.5 – 1.7 Norm after meal, inflammation, hormones	-
A(4/5)	1.6 – 3.2	-	3.2 – 3.8 Ketones, problems with the pancreas	-
A(4/6)	< 4.5	-	4.8 – 5.2 Severe inflammation, stress, exhaustion, ketones	> 6.0 Stress, weakness, severe inflammation
A(4/8)	-	-	0.62 – 0.80 Problems with pancreas or stress and inflammation	-
A(5/6)	-	1.3-1.5 Hunger, fatigue, lack of sleep, stress	1.8-2.2 Hot, stress	-

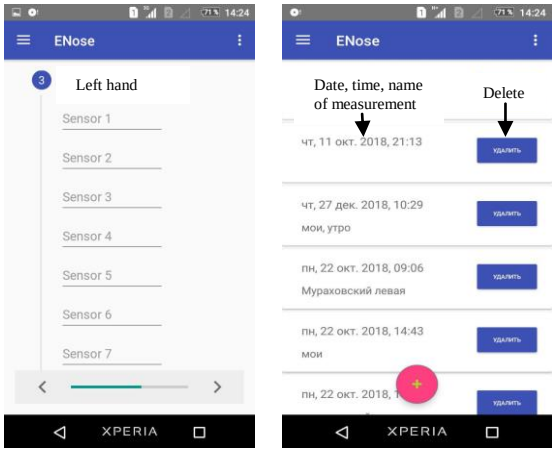
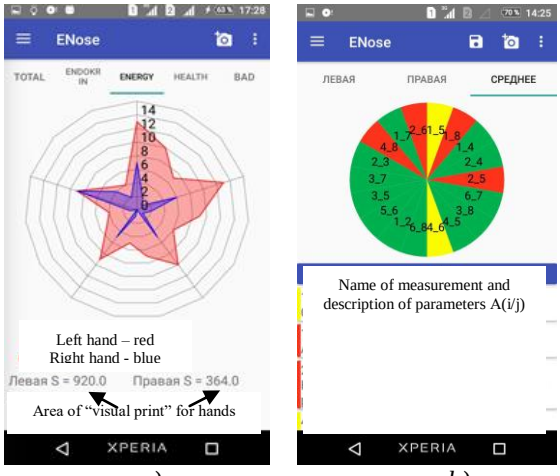


Figure 5. The dialoge windows of program to input of sensor responses and save the mesurements.

The sector's colors are determined by the numerical value of parameters in Table V. The comments about the body's health state or the volatile substances in the analyzed sample are displayed depending on the interval in which the calculated parameters fall. The complete data about measurements (entered and calculated) are saved to the database (on a personal device) (Fig. 5). The more green sectors in the health status sphere, the closer it is to norm. For instance, in Fig. 7 on the sphere of the state, 4 red and 2 yellow sectors are noted, which corresponds to many deviations from the norm: inflammation, disturbances in the functioning of the endocrine glands, a violation in the work of the excretory system. The value of parameters *A(2/6), A(4/8)* is in the red zone, which corresponds to the spasm and problems with the gastrointestinal tract or inflammation and the results of clinical and laboratory tests confirm inflammatory process in the digestive and respiratory tracts.



a) *b)*

Figure 6. The dialoge windows of program with text and graphical information and results of comparison for two measurent for left and right forearm (*a*) and for average measurement with norm (*b*) –green sector means that values of parameter included in diapazon for norm state, yellow, red color – deviation from norm.

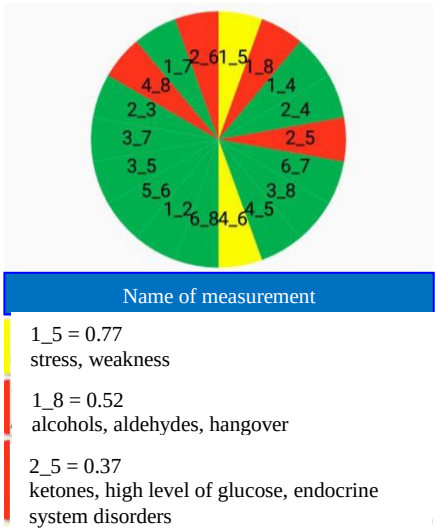


Figure 7. Example of the sphere for human health status on parameter *A(i/j)* in the dialogue window of the software.

In the example in Fig. 8, the presence of inflammation is noted, presumably of the gastrointestinal tract by the parameters *A(2/6), A(4/5), A(4/6)*, severe exhaustion and stress according to parameters *A(5/6), A(1/7)*. Parameter *A(3/5)* reflects the residual processes of food digestion. The doctor's conclusion by the examination of the volunteer with the special test was gastritis. According to the survey of this volunteer, the study was carried out after a snack due to a strong feeling of hunger. Although in the cases of monitored diseases, correlations were established between deviations in laboratory indicators and parameters of the sensor array, the study of the mechanisms of changes in the volatile metabolome in the process of biological adaptation in the case of specific diseases requires clarification.

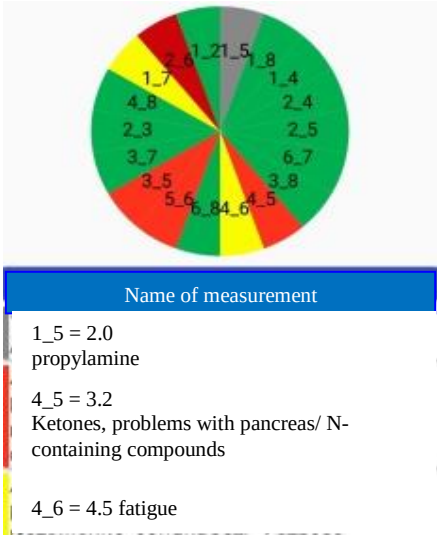


Figure 8. Example of the sphere for human health status on parameter *A(i/j)* in the dialogue window of the software.

Thus, according to the parameters of sensor signals, one can easily and quickly obtain comprehensive information about the deviation of the health status from the average norm.

The general algorithm for assessing the condition using a portable "electronic nose" consists of:

1) Measurement of the volatile metabolome of the forearm skin with an array of 8 sensors.

2) Processing of sensor chronograms – constructing of the "visual prints" and calculation of its area, parameters *A(i/j)*.

3) Comparison of the shape and areas of "visual prints" for the left and right hands according to the "General condition" algorithm

4) If there are deviations, we analyze the "visual prints" using other algorithms to determine the presumptive reason for the deviation.

5) Deviations identified by other prints are confirmed and refined by the parameters *A(i/j)* in the health status sphere.

6) If serious deviations are found (the presence of a disease, hormonal imbalance, inflammation), it is necessary to consult a specialist and take specific laboratory tests.

At the moment, according to the collected base of measurements, the sensitivity and specificity of determining descriptive states (fatigue, anxiety, stress, excitement, and others) is 98 %, clinical cases of the disease, including exacerbation of chronic illness, – up to 80%. The sensitivity of determining deviations in some blood test indicators (changes in hormonal balance, including during a monthly cycle in women, an increase in glucose levels) reaches 100%.

Thereby, using portable e-nose for scanning volatilome secreted by the skin allows determining the diverse variation in health status, including the descriptive states, the presence of individual volatile substances at ppm-level.

VI. Conclusion and Future Work

According to the responses of the electronic nose, when monitoring biosamples of nasal mucus from 17 calves during 5-9 hours, with an interval of 2-3 hours, the changes in the qualitative and quantitative composition of biosamples gas phase possible to assess atraumatic, on a place. The proposed parameter ΣF_{ref}. can be used to differentiate the diagnostic group of calves, identifying the animals that needed medical treatment.

For the first time, a fundamentally new approach is proposed for a quick assessment of the health status of the human body as a whole (normal, stress, inflammation) and the work of individual systems (reproductive, endocrine, digestive), based on the results of an assessment of the qualitative and quantitative composition of the gas mixture of biomolecules secreted by the skin in the forearm and the Zakharyin-Ged zone.

The correct interpretation and prediction of the status of biosamples and a person's state have been proved by different analyses (leukogram, biochemical, microbiological, molecular genetic analysis). The algorithms are proposed for reading and visualizing signals from an array of sensors understandable to any user.

The time of one analysis of volatile substances excreted by the biosamples or skin using portable e-nose, including visualization and processing, is up to 5 min that is faster than described in works [3][4][5][6], and without any sample preparation, unlike other modern research [3][8][9].

We believe that proposed in this work approach in the analysis by sensor array is appropriate for other biosamples, such as blood, cervical mucus, exhaled breath condensate, and urine.

In future works, we plan to expand the set of parameters when processing sensor chronograms to better specify the state of a person and differentiate states corresponding to one parameter value, such as stress, inflammation, and gastrointestinal problems. Investigation of idiosyncratic changes in humans and animals is one of the future work tasks. Also, we plan to improve the algorithm for deciding on the dominant reason for the deviation from the norm state to include it in the sensor data processing program.

Acknowledgment

The authors thank the committee of the Fifth International Conference on Advances in Sensors, Actuators, Metering and Sensing for selecting our paper to provide an extended version. The part of the work is financially supported by the Russian Scientific Fund (grant number 18-76-10015).

References

[1] A. Shuba, T. Kuchmenko, R. Umarkhanov, and A. Chernitskiy, "Portable E-nose for Diagnostic of Inflammation and Diverse Variation in Health Status of Humans and Animals", In Proceedings of the Fifth International Conference on Advances in Sensors, Actuators, Metering and Sensing (ALLSENSORS 2020), IARIA, Nov. 2020, pp. 56-62, ISSN: 2519-836X, ISBN: 978-1-61208-766-5.

[2] A. D. Wilson and M. Baietto, "Advances in Electronic-Nose Technologies Developed for Biomedical Applications", Sensors, vol. 11, pp. 1105-1176, Jan. 2011, doi: 10.3390/s110101105.

[3] A. T. Güntner, S. Abegg, K. Königstein, P. A. Gerber, A. Schmidt-Trucksäss, and S. E. Pratsinis, "Breath Sensors for Health Monitoring", ACS Sens., pp. 268-280, Jan. 2019, doi: 10.1021/acssensors.8b00937.

[4] M. Yakob, D. Mustika, R. I. Nila, and R. A. Putra, "Design of E-nose as an Instrument Identification of Diseases Through the Respiratory Tract", Journal of Physics: Conference Series (The 4th International Conf. on Applied Physics and Materials Application), IOP Publ., Feb. 2020, vol. 1428, 012062, doi:10.1088/1742-6596/1428/1/012062.

[5] E. Kabir et al., "Recent Advances in Nanomaterial-Based Human Breath Analytical Technology for Clinical Diagnosis and the Way Forward", Chem, vol. 12, pp. 3020-3057, Dec. 2019, doi:10.1016/j.chempr.2019.08.004

[6] Y. Y. Broza et al., "Disease Detection with Molecular Biomarkers: From Chemistry of Body Fluids to Nature-Inspired Chemical Sensors", Chem. Rev., vol. 119, pp. 11761-11817, Nov. 2019, doi:10.1021/acs.chemrev.9b00437.

[7] T. A. Kuchmenko, A. A. Shuba, V. V. Bityukova, and N. A. Matveeva, "Application of a Chemical Sensor Array to Assessing the Presence of Neoplasms by Blood Smell", Journal of Analytical Chemistry, vol. 73, pp. 91-101, Jan. 2018, doi: 10.1134/S1061934817110053.

[8] Q. Gao and W.-Y. Lee, "Urinary Metabolites for Urological Cancer Detection: a Review on the Application of Volatile Organic Compounds for Cancers", Am J Clin Exp Urol, vol.

7, pp. 232–248, Aug. 2019; ISSN: 2330-1910/AJCEU0101247.

[9] G. E. Bergdahl, T. Andersson, M. Allhorn, S. Yngman, R. Timm, and R. Lood, "In Vivo Detection and Absolute Quantification of a Secreted Bacterial Factor from Skin Using Molecularly Imprinted Polymers in a Surface Plasmon Resonance Biosensor for Improved Diagnostic Abilities", ACS Sens., vol. 4, pp. 717-725, Mar. 2019, doi:10.1021/acssensors.8b01642.

[10] S. Zhang, S. Li, Z. Xia, and K. Cai, "A Review of Electronic Skin: Soft Electronics and Sensors for Human Health", J. Mater. Chem. B, vol. 8, pp. 852-862, Dec. 2019, doi:10.1039/C9TB02531F.

[11] A. Amann and D. Smith "Volatile Biomarkers. Noninvasive Diagnosis in Physiology and Medicine", Amsterdam: Elsevier, 2013. p. 568.

[12] P. Lorwongtragool, E. Sowade, N. Watthanawisuth, R. R. Baumann, and T. A. Kerdcharoen, "A Novel Wearable Electronic Nose for Healthcare Based on Flexible Printed Chemical Sensor Array", Sensors, vol. 4, pp. 19700-19712, Sept. 2014; doi:10.3390/s141019700.

[13] T. Seesaard, S. Seaon, P. Lorwongtragool, and T. Kerdcharoen, "On-cloth Wearable E-nose for Monitoring and Discrimination of Body Odor Signature", 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), IEEE Press, June 2014, pp. 1-5, doi: 10.1109/ISSNIP.2014.6827634.

[14] R. Vishinkin et al., "Profiles of Volatile Biomarkers Detect Tuberculosis from Skin", Adv. Sci, vol. 8, 2100235, June 2021, doi: 10.1002/advs.202100235.

[15] A. Voss et al., "Smelling Renal Dysfunction via Electronic Nose". Ann Biomed Eng, vol. 33, pp. 656-660, May 2005, doi: 10.1007/s10439-005-1438-2.

[16] A. Voss et al., "Smelling heart failure from human skin odor with an electronic nose", 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE press, Nov. 2012, pp. 4034-4037, doi: 10.1109/EMBC.2012.6346852.

[17] S. M. McGuirk, "Disease Management of Dairy Calves and Heifers", Veterinary Clinics of North America: Food Animal Practice, vol. 24, pp. 139-153, April 2008, doi: 10.1016/j.cvfa.2007.10.003.

[18] E. N. Bulanov, "Obtaining and investigation of nanostructured boicompatible materials based on hydroxyappatite", Nizhnii Novgorod: Nizhnii Novgorod State University, 2012. pp. 64-100, (in Russian). http://www.lib.unn.ru/students/src/bulanov.pdf

[19] T. A. Kuchmenko and L. B. Lvova, "A Perspective on Recent Advances in Piezoelectric Chemical Sensors for Environmental Monitoring and Foodstuffs Analysis", Chemosensors, vol. 7, pp. 39-45, Aug. 2019, doi: 10.3390/chemosensors7030039.

[20] T. A. Kuchmenko, R. U. Umarkhanov, A. A. Shuba, and D. A. Menzhulina, "The Application of Gas Sensor with Biohydroxyapatite to Study the Volatile Profile of Nasal Secretion", The SEIA Conference Series Proceedings (The 7th International Conference on Sensors and Electronic Instrumentation Advances), IFSA Publ., Sep. 2021, pp. 56-61. ISBN: 978-84-09-33525-1.

[21] T. A. Kuchmenko, A. A. Shuba, R. U. Umarkhanov, E. V. Drozdova, and A. E. Chernitskii, "Application of a Piezoelectric Nose to Assessing the Respiratory System in Calves by Volatile Compounds", Journal of Analytical Chemistry, vol. 75, pp. 645-652, May 2020, doi: 10.1134/S1061934820050093.

[22] T. A. Kuchmenko, A. A. Shuba, and I. V. Cheremushkina, "Application of Chemical Sensors to the Rapid Assessment of the Digestive Tract of Birds", Journal of Analytical Chemistry, vol. 71, pp. 1096-1103, Nov. 2016, doi: 10.1134/S1061934816110071.

[23] G. Sayerbrey, "Messung Von Plattenschwingungen Sehr Kleiner Amplitude Durch Lichtstrommodulation", Zeitschrift Fuer Physik, vol. 178, pp. 457-471, 1964.

[24] T. A. Kuchmenko, A. A. Shuba, and N. V. Belskikh, "The Identification Parameters of Organic Substances in Multisensors Piezoquartz Microbalance". Analitika i kontrol' [Analytics and Control], vol. 21, pp. 151-161, June 2012, (in Russian).

[25] T. A. Kuchmenko and A. A. Shuba, "Informative Nature of the Electronic Nose Output Signals Based on the Piezoelectric Sensors", Analitika i kontrol' [Analytics and Control], vol. 21, pp. 72-84, June 2017, doi: 10.15826/analitika.2017.21.2.001 (in Russian).

[26] T. A. Kuchmenko, A. A. Shuba, R. U. Umarkhanov, and L. B. Lvova, "The New Approach to a Pattern Recognition of Volatile Compounds: The Inflammation Markers in Nasal Mucus Swabs from Calves Using the Gas Sensor Array", Chemosensors, vol. 9, 116, June 2021, doi: 10.3390/chemosensors9060116.

[27] T. A. Kuchmenko, A. A. Shuba, R. U. Umarkhanov, and A. E. Chernitskiy, "Portable Electronic Nose for Analyzing the Smell of Nasal Secretions in Calves: Toward Noninvasive Diagnosis of Infectious Bronchopneumonia", Veterinary Sciences, vol. 8, 74, May 2021, doi:10.3390/vetsci8050074.

[28] T. A. Kuchmenko, A. A. Shuba, R. U. Umarkhanov, and A. E. Chernitskiy, "Electronic Nose Signals Correlation Evaluation for Nasal Mucus and Exhaled Breath Condensate of Calves with the Clinical and Laboratory Indicators", Analitika i kontrol' [Analytics and Control], vol. 23, pp. 557-562, Dec. 2019, doi: 10.15826/analitika.2019.23.4.014 (in Russian).

[29] Y. Alekhin, M. Zhukov, V. Morgunova, G. Chusova, and I. Klement'eva, "Formation of Local Protection of the Respiratory Tract in Holstein Calves", Turk. J. Vet. Anim. Sci., pp. 656-661, Oct. 2019, doi:10.3906/vet-1903-73.

[30] T. A. Kuchmenko, "Electronic Nose Based on Nanoweights, Expectation and Reality", Pure and Applied Chemistry, vol. 89, pp. 1587-601, Oct. 2017, doi: 10.1515/pac-2016-1108.

[31] T. A. Kuchmenko, A. A. Shuba, and E. V. Drozdova, "Substantiation of the Operating Life of Gas Piezosensors in Detection of Vapors of Organic Compounds", Russian Journal of Applied Chemistry, vol. 88, pp. 1997-2008, Dec. 2015, doi: 10.1134/S10704272150120150.

[32] V. N. Skorikov, T. A. Kuchmenko, V. I. Mikhalev, and R. U. Umarkhanov, "The Use of the Electronic Nose Device for the Diagnosis of Postpartum Metritis in Cows", IOP Conference Series: Earth and Environmental Science, vol. 640, pp. 072032, Feb. 2021, doi: 10.1088/1755-1315/640/7/072032.

[33] A. D. Wilson, "Biomarker Metabolite Signatures Pave the Way for Electronic-nose Applications in Early Clinical Disease Diagnoses", Current Metabolomics, vol. 5, pp. 90-101, Aug. 2017, doi: 10.2174/2213235X04666160728161251.

[34] H. Shin et al., "Acetaldehyde and Hexanaldehyde from Cultured White Cells", J. Transl. Med., vol. 7, pp. 31, Apr. 2009. doi:10.1186/1479-5876-7-31

[35] R. Xue et al., "Investigation of Volatile Biomarkers in Liver Cancer Blood Using Solid-phase Microextraction and Gas Chromatography/Mass spectrometry", Rapid Commun Mass Spectrom., vol. 8, pp. 1181-1186, Apr. 2008, doi: 10.1002/rcm.3466.

[36] C. Brunner et al., "Discrimination of Cancerous and Non-cancerous Cell Lines by Headspace-analysis with PTR-MS", Anal Bioanal Chem., vol. 397, pp. 2315-2324, Jul. 2010, doi: 10.1007/s00216-010-3838-x.

[37] W. Filipiak et al., "TD-GC-MS Analysis of Volatile Metabolites of Human Lung Cancer and Normal Cells in

vitro", Cancer Epidemiol Biomarkers Prev., vol 19, pp. 182-195, Jan 2010, doi: 10.1158/1055-9965.EPI-09-0162.

[38] J. N. Labows, K. J. McGinley, G. F. Webster, and J. J. Leyden, "Headspace Analysis of Volatile Metabolites of Pseudomonas aeruginosa and Related Species by Gas Chromatography-mass spectrometry", J Clin Microbiol., vol 12, pp. 521-526, Oct. 1980, doi: 10.1128/JCM.12.4.521-526.1980.

[39] B. De Lacy Costello et al., "A Review of the Volatiles from the Healthy Human Body," J Breath Res., vol. 8, pp. 014001, Jan. 2014, doi: 10.1088/1752-7155/8/1/014001.

[40] O. Lawal et al.,"Headspace Volatile Organic Compounds from Bacteria Implicated in Ventilator-associated Pneumonia Analysed by TD-GC/MS", J Breath Res., vol. 12, pp. 026002, Jan. 2018, doi: 10.1088/1752-7163/aa8efc.

[41] Y. Zhang, L. Guo, Z. Qiu, Y. Lv, G. Chen, and E. Li, "Early Diagnosis of Breast Cancer from Exhaled Breath by Gas Chromatography-mass spectrometry (GC/MS) analysis: A prospective Cohort Study", J Clin Lab Anal., vol 34, pp. e23526, Dec. 2020, doi: 10.1002/jcla.23526.

[42] P. Mochalski, J. King, M. Haas, K. Unterkofler, A. Amann, and G. Mayer, "Blood and Breath Profiles of Volatile Organic Compounds in Patients with End-stage Renal Disease", BMC Nephrol., vol. 15, pp. 43, Mar. 2014, doi: 10.1186/1471-2369-15-43.

[43] P. Neyer, L. Bernasconi, J. A. Fuchs, M. D. Allenspach, and C. Steuer, "Derivatization-free Determination of Short-chain Volatile Amines in Human Plasma and Urine by Headspace Gas Chromatography-mass spectrometry", J Clin Lab Anal., vol. 34, pp. e23062, Feb. 2020, doi: 10.1002/jcla.23062.

[44] P. M. Santos, M. Del Nogal Sánchez, A. P. C. Pozas, J. L. P. Pavón, and B. M. Cordero, "Determination of Ketones and Ethyl acetate - a Preliminary Study for the Discrimination of Patients with Lung Cancer", Anal Bioanal Chem., vol. 409, pp. 5689-5696, Sep. 2017, doi: 10.1007/s00216-017-0508-2.

[45] T. A. Kuchmenko, R. U. Umarkhanov, and T. Chikake, "Server for Reading, Calculation and Storage of Geometric Parameters of the Elastic e-nose Server Sensor Response Matrix," Certificate of computer program registration RU 2019616314, 05.22.2019. Application No. 2019615346 dated 05.13.2019. (Т.А. Кучменко, Р.У. Умарханов, Т. Чикаке, "Сервер для чтения, расчета и хранения геом. парам. матрицы откликов сенс. elastic e-nose server" Св-во RU 2019616314, 22.05.2019. Заявка № 2019615346 от 13.05.19).

[46] T. Chikake, D. A. Kuchmenko, T. A. Kuchmenko, and R. U. Umarkhanov, "Cloud Service for the Collection and Processing of Information Systems Piezoelectron Nose," Proc. Of the Conf. "Problems and Innovative Solutions in Chemical Technology", Oct. 2019, pp. 23-24. (Т. Чикаке, Д.А. Кучменко, Т.А. Кучменко, Р.У. Умарханов, "Облачный сервис для сбора и обработки информации систем пьезоэлектронный нос," В сборнике: Проблемы и инновационные решения в химической технологии, Окт. 2019, ст. 23-24.)

Exploring Caregivers' Perceived Stress with Their Need-Factors

Ah-Choo Koo
Multimedia University, Cyberjaya, Malaysia
ackoo@mmu.edu.my

Kok-Yew Ang
Universiti Tunku Abdul Rahman, Sungai Long, Malaysia
angky@utar.edu.my

Sian-Hoon Teoh
Universiti Teknologi MARA, Puncak Alam, Malaysia
teohsian@salam.uitm.edu.my

Wei-Fern Siew
International Medical University, Bukit Jalil, Malaysia
weifern_siew@imu.edu.my

Tenku Putri Norishah, Peter Charles Woods
Multimedia University, Cyberjaya, Malaysia
tengku.norishah@mmu.edu.my, p.woods@mmu.edu.my

Weng-Ping Chin
Puchong, Malaysia.
lydiadigitalclassroom@gmail.com

Abstract **- Caregivers face numerous challenges in providing care for a wide range of illnesses and health conditions associated with communicable diseases or non-communicable diseases. Pandemics, ageing populations, environmental health concerns, lifestyle changes, and other factors contribute to health issues, which lead to a greater demand for caregiving. This warrants further research on caregivers' wellbeing. The study reported six need-factor of caregivers, which are the needs for regular communications, personal well-being, basic healthcare, access to information, coping with change and learning about caregiving. This paper explores these objectives: 1) To determine perceived stress level of caregiving among the caregivers; 2) To determine perceived stress level of caregiving among caregivers based on gender, age, duration of caregiving, and frequency of using internet for information seeking; 3) To determine the relationship of the need-factors of caregivers with their perceived stress level. A total of 84 caregivers responded in a survey by using questionnaire. The t-test, analysis of variance, and correlation were used in data analysis. Findings reveal a normal distributed perceived stress level with a weak relationship with the need for information on caregiving. Caregivers should be able to access and acquire reliable information about the care guides which would aid in the provision and management of care, especially in this unprecedented time due to COVID-19.**

Keywords- Caregiver; perceived stress; information need; factor analysis; care model.

I. INTRODUCTION

The rise of noncommunicable diseases over the years has resulted in an increasing disease burden in our society, with people suffering from life-limiting chronic illnesses. Their debilitating illness necessitates the assistance of caregivers. As a result, there is a greater demand for caregivers to provide care informally or professionally. Caregivers frequently faced with a range of challenging circumstances that they must handle while doing their jobs, and stress becomes unavoidable for them. In a study published in Global Health 2020 Contribution 70025, Koo et al. [1] identified six need-factors that facilitate caregivers in carrying out their responsibilities and making care decisions. The six needs-factors are regular communications, personal well-being, basic healthcare, access to information, coping with change and learning about caregiving. In this study, the authors proceeded to determine if caregivers' perceived stress level of giving care influenced by these factors, particularly in a case study of cancer caregivers.

Caregiver is defined as "someone who performs hands-on care and/or provides emotional support to patients, such as a partner, relative or friend" [6, p. 388]. "Formal caregivers typically undergo training and certification and may inherently have greater health literacy capacities compared to informal caregivers" [26, p. 12]. The term "caregivers" used in this study is to encompass both partners and family caregivers, who are inherently informal caregivers.

Caregivers to elderly recipients have a significant influence on their treatment [17]. If the caregiver is depressed or lacks resilience, the care recipient may not be able to get the most of care, thus affecting the quality of life for both parties. Support is needed for caregivers, such as physical, psychological, social, and spiritual [4][20] and encompass many decades of care for patients with chronic neurological conditions, such as multiple sclerosis, Alzheimer's, Huntington's disease, stroke, and Parkinson's disease [22]. However, the capacity of informal caregivers to source and utilize information for them to perform their tasks is not well understood [24]; the roles of family caregivers are multifaceted and challenging. Health and social well-being have been an issue for the current society, hence care services (informal or formal) play an increasing role, and therefore the motivation of conducting this study.

Yuen et al. [26. p. 12] mentioned that "accessing information from the internet may entail additional demands and capacities compared to traditional health literacy due to the factor of competing sources, identifying accurate and trustworthy resources, technological and internet literacy, as well as access to technology and the internet". They propose future research could examine various strategies in providing information to caregivers through eHealth modalities. Ribeiro et al. [18] suggested that innovative health education, such as mobile learning applications will further expand the context of a smart learning ecosystem for cancer education.

The objectives of this paper are as follows: 1) to determine perceived stress level of caregiving among the caregivers; 2) to determine perceived stress level of caregiving among caregivers based on gender, age, duration of caregiving, and frequency of using internet for information seeking; and 3) to determine the relationship of the need-factors of caregivers with their perceived stress level.

Section II reports relevant literature based on various aspect of caregiving, especially informal ones, and the challenges faced by caregivers. Section III reports the methods of the study which include the process of data reduction analysis using factor analysis. Section IV provides analysis and findings for the research questions. Section V is discussions based on findings and finally, Section VI is conclusion, and future works.

II. LITERATURE REVIEW

A. Types of caregiver

The need of caregiving in society is ever increasing. According to WHO's Global Health Observatory [3] on a global scale, the life expectancy of a person is 73.3 years in 2019. For South-East Asia is 71.4 year; Europe being the highest, 78.2. Persons could expect a healthy life of 63.7 years. According to [3], the gap between healthy and unhealthy life is about 9 years, which the populations may need care services for that duration.

Generally, the relationship of informal caregiver and care recipient is family members relationships. Reference [7] reported that, 5.1% (95% CI = 4.45, 5.87) of adults (from the data of National Health and Morbidity Survey 2019 with 11,160 adults involved, estimated to represent 21.3 million adults aged 18 years and above) in Malaysia reported to be informal caregivers. [7] provided two levels of caregiving - high and low intensity. Those high ones were likely to be actively employed and provide longer duration of care compared to the low intensity ones. For low intensity caregiving, females, those aged 35–59 years, and those with long-term condition were more likely to have negative effects on health. For high intensity caregiving, caregivers aged 60 and over, those received training and those without assistance were more likely to have negative effects on health [7]; the research team further suggested that "caregiving, regardless of intensity, has a significant impact on caregivers" (p. 1). To reduce the negative effects of caregiving duties, all caregivers need some assistance from the supporting environments such as, the community and government to support their needs. [8] revealed that the issue of social care, particularly the role of informal caregivers and their wellbeing were often neglected or missing in the discussions. In the UK today, 6.5 million people are caregivers, to care for different type of care recipients including seriously ill persons. Providing care, especially long-term care, can impact on a person's health, finance, and relationships [8]. Many caregivers provide care in long hours or many years which affect their wellness. The survey conducted by [8] reported that 25% of caregivers experienced bad or very bad physical health and 29% of them reporting bad or very bad mental health; 81% of all caregivers reported to have felt lonely or isolation. Since many countries are having a development of an ageing population, shrinking of family sizes, and increased of women participated in workforce has posted many challenges in social care.

B. Framing caregiving based on triadic model

There are some models related to caregiving. [10] has come up with a triadic model, namely Figure 1, which shows the factors influencing the three groups of stakeholders in caregiving. This paper highlighted the importance of caregiving has only recently been acknowledged by the nation, such as, the US, as an important topic to the nation.

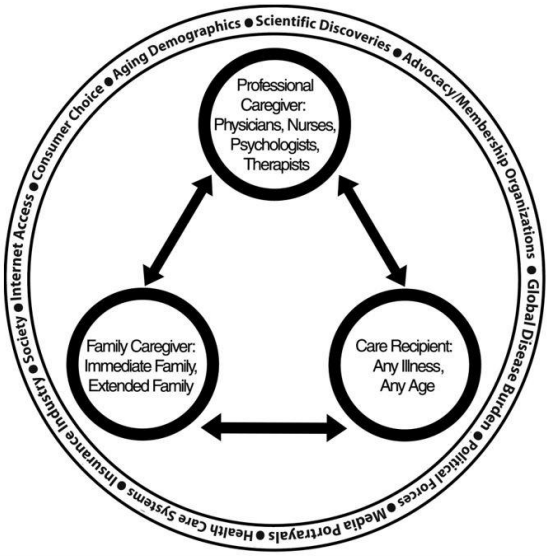


Figure 1. "A triadic model of caregiving: factors influencing the care recipient, family caregiver, and professional caregiver team" (Source: [10]).

Figure 1 shows that caregiver's type/group is identified to be either a family caregiver or a professional caregiver. A family caregiver could be an immediate or extended family member. Care recipients can be inflicted with any illness, and at any age or age that requires care from others. Caregiving resides in the domain of public health. Figure 1 shows a care triad within a complex system of variables that influence the system of caregiving from a larger environment. The triadic is influenced by the complex force from societal, political, and scientific issues that shape the context of care, such as global disease burden, demographic changes, health insurance coverage, and scientific discoveries. Care triad deals with a variety of internal as well as external variables that facilitate or inhibit the care situation, enhancing the chances for success or hindering them. More studies should be conducted in understanding various factors influencing the wellbeing of caregivers, particularly the family caregivers.

C. Care model during COVID-19

Digital technology has transformed the care model in this unprecedented challenge time for healthcare systems internationally (the whole world) [11]. Caregiving during COVID-19 have affected the caregiver's burden and mental health [13]. Subsequently, this affected the parents and children relationship. Digital health transformation has been the rapid development and implementation of new models of care which incorporate digital technology health [11]. The forces to shield people from COVID-19 have resulted in an increased in information seeking using telemedicine consultation approaches, as well as the rapid rollout of digital apps, digital education / training, etc. Reference [11] concluded that the human costs of COVID-19 will be high and long remembered; the change of the use of new ways of remote and digital health and sustained these care models (extended even for caregiving) will be the future developments (p. 2).

D. Information technology for seamless learning and the infodemic phenomenon: A mixed blessing

The blessing for caregivers is that they are privileged in this era to receive healthcare information at their fingertips. The idea of a Seamless Learning Model in context of caregivers' informal learning and information seeking pursuits, implemented seamlessly via the internet and social media, which impacts their public and private learning spaces. Many activities of learning and information seeking, especially for caregivers are inherently informal, self-directed, independent, and critical as they frequently influenced by online technologies and social media.

There are many internet-based information platforms for supporting and developing skills in caregiving and social care, for example [27][28][29]. It is also recognized as an authentic and just-in-time learning (or training), especially for caregivers or informal caregivers who need some help and guidance. For catering the needs of receiving guidance, there are formal caregiving courses or websites created caregivers (such as, MOOC courses that promote lifelong learning). [14] proposed a Seamless Learning Framework as shown in Figure 2. It explains the learning environments or dimensions experienced by most of the learners. The learning space is no longer defined by a "physical / formal class" but by "learning unconstrained by scheduled class hours or specific locations" [14, p.156], thus promoting seamlessness, with informal learning and information access at ones' fingertips.

The seamless environment is labeled as "community" which comprises different categories of people such as, teachers, experts, and learners. The community has access to any relevant sources of knowledge through cognitive tools, within the dimension of time (anytime), space (anywhere), and artefacts (any learning artifacts / contents).

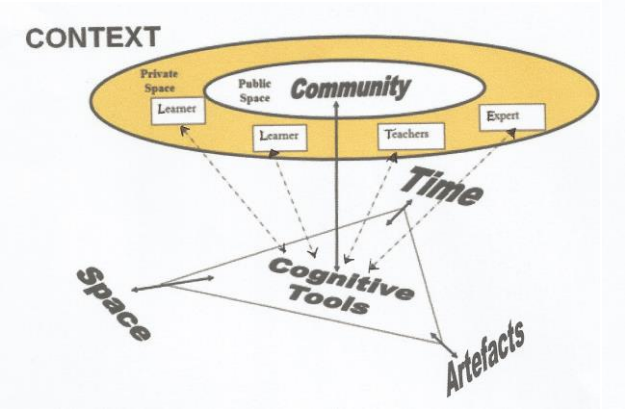


Figure 2. A Seamless Learning Framework [14].

This concept can be applied to fulfil the informal learning needs of cancer caregivers since caregivers interact with all kinds of information and materials seamlessly. They interact with internet support groups or cancer survivors through social media, and conduct discovery learning about the disease from the cancer journeys shared by others. There are also professional cancer learning and sharing materials, which can be accessed with ease. [25] suggested that public health information campaigns could be conducted using social media. This new means of communication, especially for prevention purposes, will complement other methods of communication.

However, according to [2, p. 627], the importance of these online media and technologies has not been clearly revealed in previous studies, especially on the information-seeking behavior of family caregivers. Not a blessing part is the current infodemic. The Internet and social media are a 'mixed blessing' for the healthcare sector. However, the challenge will be to adhere to the legal framework that preserves the quality of the healthcare information provided on the internet and consume by the caregivers / patients. Infodemic challenges have even become a 'disastrous' than the pandemic itself. According to [15] that "…a global epidemic of misinformation—spreading rapidly through social media platforms and other outlets—poses a serious problem for public health". Fighting infodemic and misinformation is a joint effort. This is because fighting every outbreak will be accompanied with tsunami of information and misinformation, rumors, etc., and social media amplifying it [15]. Users, such as caregivers must critically analyze and select information posted on YouTube to make effective healthcare decisions. Moreover, diagnostic information online is sometimes used to confront doctors [12][15].

Consumption of online healthcare information and services is increasing [15] and has empowered caregivers (and patients) to enhance their health and digital literacy for improved decision-making [9]. Hence, getting more insights on the proper way of obtaining useful and informative inputs for caregivers is essential.

E. Identifying challenges and needs of an informal caregiver

Among the bigger challenges faced in long term caregiving are the need to give up their personal life or career and becoming a full-time caregiver, this is normally a case for Asian culture [16]. Most caregivers will be coming from family members who are informal mostly, and conflict may happen between them with care recipients or the other family members who are not able to empathize the roles and feeling of caregivers. Another challenge that they faced, is the 'shrink' of social life to be just in the circle of family and a gradual distant from their career life [16].

Seattle Cancer Care Alliance (SCCA) [19] listed the challenges of caregiving such as, caring for themselves (self-care), supporting and caring patients emotionally and physically, maintaining the home environment for patients, gathering information, helping with decision-making on cancer care or treatment, arranging patients' hospital visits and patient's financial support. According to Wingate and Lackey cited in [2], family caregivers need knowledge, information, or understanding that can be gained through education, experience, study, or through explanations by qualified specialists. Chen [2] reported a qualitative study that discloses the information needs of cancer family caregivers are varied along the cancer journey, and they used diverse information sources, including healthcare professionals, hospital booklets, interpersonal networks, besides the internet, mass media, and books - to satisfy their needs. Her study found that demographic variables of caregivers (such as, gender, age, level of education, socioeconomic status, and culture) affected their information-seeking behaviors. Girgis et al. [6] measured the psychometric properties to capture the multidimensional supportive care needs of cancer caregivers. The instrument used was given to 547 cancer caregivers. Psychometric analyses found four dimensions of need: healthcare service, psychological and emotional, work, and social, and information. Caregivers with anxiety and depression were more likely to report, "at least one unmet moderate or high need in comparison to non-anxious participants"; younger caregivers faced at least one unmet moderate or high need around "psychological and emotional", and "work and social", as compared to the older participants. Girgis et al. [6] suggested the findings can be used to prioritize healthcare resources and tailor supportive cancer care service accordingly. The model and framework reviewed above facilitated the conception of need-factors as reported by [1].

This study focuses on investigating significant factors that correlate to caregivers' perceived stress levels. The contributing constructs provide guidelines on further plans or actions to reduce stress among caregivers. Cancer caregivers were chosen as the study's setting due to the rising number of cancer patients in Malaysia [21] and the need for improving cancer management, which certainly requires greater focus in research, such as, on the caregiving aspect.

III. METHODS

A survey was conducted in a hospital with two cancer specialist clinics run by National Cancer Society Malaysia (NCSM), an NGO for cancer awareness and cancer care. Permission was granted from NCSM to conduct this study at the waiting lounge of the two cancer clinics. The population of the study involved all cancer clinics run by NCSM across the country. However, only two cancer clinics were purposively selected to participate in this study due to their strategic location in the center of Kuala Lumpur city. A total of 84 participants were involved in this study. Majority of them were Malaysian Chinese. This is because the hospital is traditionally or historically relevant and popular among Malaysian Chinese community.

A. Instrument

The instrument of the survey was adapted from [6] to study needs of accessing information in the context of informal learning environments. Items related to this context were added to the instrument. The instrument was then reviewed by three experts in wellness and preventive medicine (Expert 1), management of a cancer wellness center (Expert 2), and healthcare informatics (Expert 3) respectively. Items were reviewed; some were dropped and merged because of overlapping meaning; some were rephrased for the suitability of local caregivers in Malay and English. A comprehensive factor analysis was conducted to establish the factors. The items were analyzed in a comprehensive factor analysis. It involves two stages as below.

1) Stage one: Extracting factor

In PCA, the suitability and adequacy of data in terms of variability of data were tested based on Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy. In this study the value of KMO is 0.842 which is greater than 0.7 which indicates a very good condition to proceed with the factor analysis. In normal practice, value of KMO should be larger than 0.5 for achieving condition of satisfactory.

On the other hand, the Barlett's Test of Sphericity takes consideration of testing of correlations among the variables. This value is referred to ensure that there are sufficient correlations among the variables. The sufficiency of correlations is indicated in the associated probability in the chi-square. If the p value of the associated chi-square statistic is less than 0.5, it shows the items are sufficiently correlated for further analysis in PCA. In this study, the results of Barlett's Test of Sphericity show that value of chi-square is 2470 with df=703 and p-value < 0.05, indicating that the variables were sufficiently correlated to form the specific components and factors.

TABLE 1. KMO AND BARLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.842
Bartlett's Test of Sphericity	Approx. Chi-Square	2470.094
	Df	703
	Sig.	.000

An examination of the scree plot of the Eigenvalue versus Component has shown a clear "knee point / elbow", that is at the point of component number of 6 (refer to Figure 3). Other components have lower value of eigenvalue, which will not be considered for the next step.

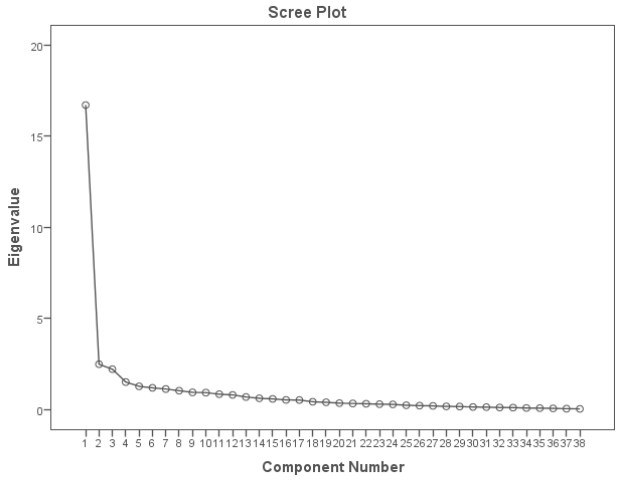


Figure 3. Scree plot showing the elbow at the point of the sixth component at the value of Eigenvalue >= 1.

The six components which have eigenvalues more than one (namely 16.714, 2.483, 2.211, 1.504, 1.270, 1.187 as observed in the analysis of the total variance explained by components) indicate there are six factors or constructs in the extraction process.

2) Stage two: Factor rotation

The final step of PCA is to examine the factor loadings of each item in relation to the first six components. The factor loading presents the results of component rotations and interpretation of components.

The factor rotation method used in the analysis is Viramax rotation, a method used frequently in social science and psychological study. A check on oblique based rotation has also produced the similar set of items according to these components. In determining the factors from the factor loading, the loading of absolutes 0.4 is used as a cut-off value. For practical significance, loadings of absolutes0.4 and above, but less than 0.5 are considered just enough to be significant. Loadings above 0.5 indicates highly significant. On the other hand, if the factor loading difference between two factors across a particular item is less than 0.2, the item should be dropped.

Table 2 is a matrix table showing the rotated components and their related items shows the loading value of each item which are greater than 0.4 in relation to the six identified components in the factor extraction, and other components (i.e., Components 7 and 8). The bold loading values are the selected items according to components; items which loading values are *italicized* due to the difference of loading values less than 0.2 were discarded (i.e. Item 18, 35, 26, 33 and 32). Components 7 and 8 were also not considered for interpretation.

TABLE 2. A MATRIX TABLE SHOWING THE ROTATED COMPONENTS AND THEIR RELATED ITEMS WITH FACTOR LOADINGS.

Item #	1	2	3	4	5	6	7	8	Difference of two factor loadings with nearest values	**Decision on item#**
14	0.819									
13	0.809									
19	0.658									
29	0.641									
20	0.622									
18	*0.575*				*0.433*				0.142 (<0.2)	Item 18 is dropped
15	0.469									
30		0.769								
27		0.766								
36		0.701								
28		**0.679**			**0.456**				**0.223 (>0.2)**	**Item 28 is maintained**
25		**0.638**			**0.424**				**0.214 (>0.2)**	**Item 25 is maintained**
10		0.611								
35		*0.593*	*0.467*						0.126 (<0.2)	Item 35 is dropped
26		*0.546*			*0.448*				0.098 (<0.2)	Item 26 is dropped
9			0.656							
6			0.605							
7			0.54							
33	*0.411*		*0.518*			*0.426*			0.015 (<0.2)	Item 33 is dropped
32			*0.515*		*0.486*		*0.432*		0.054 (<0.2)	Item 32 is dropped
8			0.476							
11			0.462							
5			0.45							
12			0.44							
3				0.783						
1				0.781						
2				0.706						
4			**0.425**	**0.64**					**0.215 (>0.2)**	**Item 4 is maintained**
24					0.787					
23					0.732					
17					0.506					
37						0.814				
38						0.809				
34						0.594				
21							0.757			
31			*0.429*				*0.444*		0.015 (<0.2)	Item 31 is dropped
22								0.624		

16	*0.418*	*0.5*	0.082 (<0.2)	Item 16 is dropped

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 10 iterations.

Note: The minimum loading score accepted is 0.4, loadings below 0.4 are not shown; *items with loading difference <0.2 are discarded from further analysis / interpretation*. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. A. Rotation converged in 10 iterations.

The bold items were retained and used for interpreting the need factor. In total, there were 29 items to explain six (6) components of need factors. Table 2 is a matrix table showing the rotated components and their related items. A total of 9 items were discarded, and not included for further interpretation of the construct or need factors.

The internal reliability analyses with the Cronbach's Alpha values for the items emerged for six components / factors are stated in column 3 of Table 4: Factors with items loaded to the six-factor of needs (Refer to the last page of this paper). All factors have the alpha values greater than 0.7 ($\alpha > 0.7$), indicating an acceptable internal reliability measure for the factors.

Collectively, none of the factors scored 'high need' and there is no factor indicating "no need". The need for communication is at the level of 'moderate'. Other need factors such as, personal well-being, basic healthcare, access to information, coping with change are between 'low' to 'moderate need'. The need for learning through online information and connection is 'low'.

B. Results of the Factor Analysis

The needs-factors are identified as F1-F6 and are detailed as follows:

- Factor 1 (F1): Regular communication for better understanding and balance of needs between caregivers and person with cancer.
- Factor 2 (F2): Personal well-being especially on the control of emotion, communication, and spiritual beliefs mainly on the quest of meaning of life and the faith in the healing process.
- Factor 3 (F3): Basic healthcare, counselling, and service.
- Factor 4 (F4): Access to information related to cancer or patient care information and services.
- Factor 5 (F5): Coping with change especially the change of life routine and perspective on life.
- Factor 6 (F6): Learning through online information and connection with others on cancer care.

The items were analyzed using factor analysis. Factors on the needs of cancer caregivers were then identified in the analysis (This part of analysis was reported in [1]). The instrument also collected data of profiles, caregiving experiences and online activities behavior, caregivers' perceived stress level. The instrument also measures the perceived stress level of caregivers according to a scale of 1 - 10. Respondents were freely to mark their level of stress guided by the simple semantic such as, minimal stress, some stress and higher stress (in which data is gathered from a scale bar from 1 to 10, where 1 represents minimal stress, the middle scale represents some stress and 10 represents the higher stress (Refer to Figure 4)

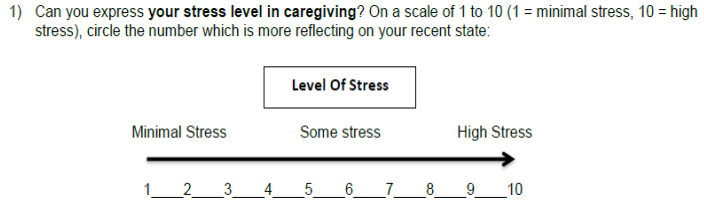


Figure 4. Scale of perceived stress level.

C. Procedures

The caregivers in the waiting lounge were invited to participate in the survey. The researcher assistant took turns to be in the waiting area for two or three days in a week for two months. The data collection process stopped when there were very few new caregivers in the waiting area. Caregivers who agreed to participate in this study were asked to sign a participation consent form. They were briefed that at any time, they could freely withdraw from the study. Some caregivers were assisted by the researchers to clarify the meaning of items. Overall, the total number of usable responses was 84 out of 91; seven responses were not included due to incompleteness.

IV. ANALYSIS AND FINDINGS

Two parts of analysis were conducted, firstly is Part A - demographic information and perceived stressed level, and Part B - probing caregivers' profile such as, age, gender, duration of caregiving, frequency to be online for accessing information, and the six need-factors with perceived stress level among the group of caregivers.

D. Description of the demographic information

The total participants were 84. More than half were females, most of them were Chinese (83%) with their religion Buddhism or Taoism, middle income or lower (77%), holding diploma and above (59%), aged more than 30 years old (73%). More than half of them (62%) have been caregiving for 6 months and above, mostly the care recipients were affected with female related cancer (54%). The caregivers were quite active online to look for health information and connection with others via online. About 60% of them were in this category.

Profile of caregivers, especially their age, gender, duration, and frequency of online access to information are further examined, whether these variables influence caregivers' stress levels.

RQ1 – What are the caregivers' stress levels?

Figure 5 presents the distribution caregivers' stress levels. The bell shape of the distribution depicts the data is normally distributed. Further testing of normality is conducted by calculating the Z-score. The Z-score value (as shown below) is calculated from the skewness statistic (skewness statistic= -0.359) and standard error (standard error = -0.269) of it (refer to Table 3).

TABLE 3. STATISTIC AND STANDARD ERROR FOR THE CALCULATION OF Z-SCORE

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
C1_stress level scale 1 - 10	80	1	10	5.68	.204	1.826	3.336	-.359	.269	.350	.532
	80										

Z-score = skewness statistic/ standard error= -0.359/-0.269 = -1.386

The Z-score = -1.386 is located between –3.29 and +3.29 shows a fulfilment of requirement of normality. It is indicated the medium sample size (50< n < 300), a threshold value of Z score is 3.29 to determine the distribution of data is normal. In this case, the null hypothesis (H0: the data is not distributed normally) is rejected at alpha level of 0.05. Hence, the data is distributed normally. In addition, the data satisfies normality since the ratio of standard error (-0.269) to its standard error is in between -2 and +2.

Refer to Figure 5, overall, the caregivers were averagely stressed on their caregiving task, with most of them having middle stressed level ranging at the mean value of 5.68 (Std. Dev. = 1.826). Majority of them have some stress (within the range of scale of 4 – 8), i.e., some stress in caregiving. The distribution of stress level is showing a normal distribution (Figure 5).

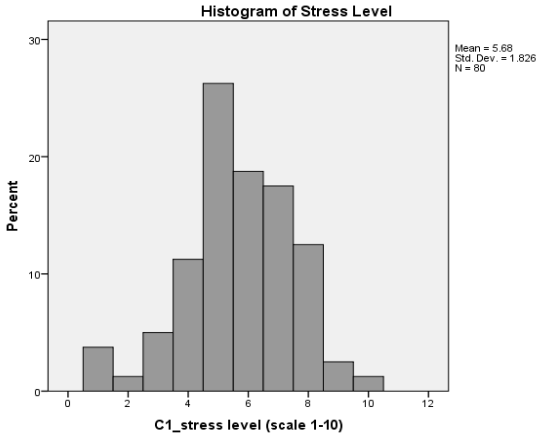


Figure 5. Histogram of perceived stressed level.
*Scale for perceived level is 1 – 10. * Note: 4 missing values reported.*

RQ2: Is there any difference in perceived stress level between/among the grouping variables (namely gender, age, period of caregiving, frequency of using internet for information seeking)?

Table 4 shows descriptive data for four grouping variables. The t-statistic and F-statistic are referred to examine whether the grouping variables has an influence on perceived stress level. The findings show that the perceived stress level does not shows difference between/ among each of the grouping variables (in Table 4) since all the p-values of the statistical test (t-test and ANOVA) exceed 0.5. Thus, it has sufficient evidence that there are no differences in perceived stress level among different groups (gender, age, period of caregiving, frequency of using internet for information seeking) of caregivers.

TABLE 4. COMPARING DEMOGRAPHIC AND EXPERIENCE VARIABLES WITH PERCEIVED STRESS LEVELS

Grouping Variables	**Mean (Std. Dev.)**	**Statistical Test**	**P value**
Main measure	Perceived Stress Level		
Gender		t = 1.122	0.265 (not
Male	5.94 (1.722)	(df = 78)	significant)
Female	5.48(1.894)		
Age			0.184
< 30	5.21(1.865)	F = 1.734	
30-49	6.09(1.505)		
50-69	5.50 (2.259)		
Duration/Period of caregiving			
<0.5 year	6.04 (1.503)	F = 1.302	0.281
0.5 year-1 year	5.43 (2.507)		
1– 2 year	5.71 (1.978)		
>2 year	5.00 (2.000)		
Frequency of using internet for information seeking:-			
Very frequently (everyday and almost everyday)	5.54 (1.584)	F = 1.6695	.176
Weekly, fortnightly or seldom	5.88 (2.748)		
Never	6.69 (1.653)		

Note: The dependent variable is caregivers' perceived stress level.

RQ3: Is there any relationship between caregivers' need factors with their perceived stressed on caregiving?

Table 5 shows the correlations of Perceived Stressed Scale with the six need-factor identified in [1], where the bivariate correlation coefficient values are displayed with the indication of p-value. The findings show that the need for 'access to information', 'basic healthcare' and 'personal wellbeing' are among the significant factor that correlate positively (have positive relationships) with perceived stress level since the p-values are less than 0.05. The correlation coefficient for the one variable which is the need for accessing information is the highest (r=0.341, p-value < 0.05) among these variables. Another three need-factor (regular communication need, need to cope with change, and the need for learning through online information and connection) were not correlated with perceived stress.

TABLE 5. CORRELATIONS BETWEEN CAREGIVERS' NEED FACTORS WITH THEIR PERCEIVED STRESS ON CAREGIVING

Need-factor (with reference to [1])	Correlation Coefficient, r	p value Sig. (2-tailed)
F1_Communication	0.193	.087
F2_Personal wellbeing	0.221	.049*

F3_Basic healthcare	0.317	.004*
F4_Access to info	0.341	.004**
F5_Cope with change	0.179	.112
F6_Learning through online	0.173	.126

Correlation of six need-factor with perceived stress; n = 84

V. DISCUSSION

The need for basic healthcare, personal wellbeing and information access are the three crucial factors which have some correlations with caregivers' perceived stress level. The personal wellbeing (in the form of positive feeling, counseling, spiritual) has often been neglected in modern care model. Personal wellbeing and the access for the basic healthcare for caregivers is vital for stress relief and to build a more robust and holistic model for patient care system. The number of caregivers is increasing due to the unprecedented pandemic situations.

The information seeking behaviors (push and pull for information) is now seamlessly a part of the life for caregivers. The information areas can be referred to online or offline channel. The infodemic viral or fake news are also a key challenge for everyone, especially more critical for those with low literacy in digital skills and health knowledge. The findings have shown that, information access to caregivers on appropriate information on healthcare, caregiving, etc. has a stronger relationship to perceived stress. This was also evidenced in the descriptive findings (mean and standard deviation) of the group of caregivers who never access information via internet to be higher stress compared to other group of users.

The seamless learning concepts by [14] are adapted to suit the needs for caregivers and keeping them informed and empowered while at the same time learning to acquire skills and knowledge for their caregiving tasks from various resources and people. Keeping caregivers well-informed and information-literate is vital in this age. In recent years, the rapid shift of digital technologies has put a toll on caregivers, particularly those who are health-illiterate and would not know how and where to acquire information as technology advanced. Most information is sought online these days. Those who never online due to many reasons (illiterate, no access to mobile data or devices, etc.) had reported higher stress level. Caregivers should be provided training or education on information seeking skills, digital and health literacy skills. They will be more empower when dealing with the new world of information era.

The current study and analysis finding is also aligned with [6] which reported that, "more caregivers experienced unmet needs varied across cancer types for the Health Care Service Needs and Information Needs domain." The unmet needs may result uneasiness in caregiving across cancer types, and these two factors contributing to some stress. [6] further stressed that by experiencing any one of the factors, "even if just one unmet need, can be quite distressing". [2] explained that caregiving needs for information is not as simple as we thought. There are these information needs related to these areas: treatment, dietary, disease specific information, homecare, psychological support, health, insurance and social welfare and funeral arrangement. These categories of information are complex, demanding, and unfamiliar to many caregivers, could be unmanaged by just one or solely one caregiver.

Information to guide new family caregivers can be provided early by healthcare bodies which can reduce the anxiety and uncertainty exhibited by caregivers [2].

A. Limitations

The study has sampling limitations, the caregivers are mostly from one ethnicity and the number of respondents is less than 85. The convenience sample for this study comes from one hospital. The sample of the study are caregivers who were selected purposively from cancer clinics in a hospital, they were affected by different kinds of cancer. All these limitations are acknowledged. Future research is proposed to incorporate more participating hospitals and caregivers.

VI. CONCLUSION AND FUTURE WORK

Overall, the caregivers involved in this study perceived that they have some stress with a few said they have minimal stress or higher stress.

Future directions of this research will consider studying on one or two aspects of need factors, especially on the needs for information during caregiving by informal caregivers. Interaction and communication of family members and decision-making processes are all quite relevant or maybe different at different stages of caregiving.

Future research will consider the different duties of caregivers and the level of quality of treatment received by patients. These variables may influence the requirements of needs by caregivers and subsequently affecting the quality of care provided by them, especially to their loved ones. [23] proposed an urgent need for research directions on the impact of COVID-19. Among the highlighted directions are the general issues with special considerations, i.e., role of technology as the 'oxygen' (as vital medium), and the importance of contextualization of research. [2] stressed about the culture aspect in caregiving; the influence of culture in health information studies requires further research. This understanding will move toward better ways of living and coping with COVID-19. Although caregiving is a routine task, it is however, a very challenging task with stress and emotion. Ethnographic method for this research area can be used to enhance understanding of the information-seeking behavior of family members.

ACKNOWLEDGMENT

The authors would like to express gratitude for the support given to the research team by Malaysia Ministry of Higher Education (MOHE)'s Fundamental Research Grant Scheme (FRGS) support to this research work. Our appreciation also goes to National Cancer Society Malaysia (NCSM) for their support to data collection process. Thanks to all the research respondents who were going through the

experiences of giving care to their loved ones. Finally, we also would like to thank the anonymous reviewers for providing constructive feedback to this paper.

REFERENCES

[1] A. C. Koo et al., "Cancer caregivers' needs for their well-being in information era," The Ninth International Conference on Global Health 2020 Contribution 70025, 25-29 October, 2020, pp. 9-12.

[2] S. Chen, "Information needs and information sources of family caregivers of cancer patients," Aslib Journal of Information Management, vol. 66(6), 2014, pp. 623–639.

[3] WHO, "Global health observatory data repository: Life expectancy and healthy life expectancy," World Health Organization,2020. https://apps.who.int/gho/data/view.main-searo.SDG2016LEXREGv?lang=en%0Ahttp://apps.who.int/gho/data/view.main.SDG2016LEXREGv?lang=en [retrieved: August, 2021].

[4] M. K. Barton, "Earlier adjuvant therapy is beneficial in patients with breast and colon cancer," CA: A Cancer Journal for Clinicians 0(0): n/a-n/a. 2015. [Online]. Available from: http://doi.wiley.com/10.3322/caac.21296 [retrieved: September, 2021].

[5] J. Li et al., "Communication needs of cancer patients and/or caregivers: A critical literature review," Journal of Oncology, vol. 2020, Article ID 7432849, pp. 1-12.

[6] A. Girgis, S. Lambert, and C. Lecathelinais, "The supportive care needs survey for partners and caregivers of cancer survivors: Development and psychometric evaluation," Psycho-Oncology, vol. 20(4), 2011, pp. 387–393.

[7] S. Jawahir, E. H. Tan, Y. R. Tan, S. N. Mohd Noh, and I. Ab Rahim, "The impacts of caregiving intensity on informal caregivers in Malaysia: findings from a national survey," BMC Health Serv. Res., vol. 21, no. 1, 2021, pp. 1-16.

[8] Carers, U. K. "State of caring: A snapshot of unpaid care in the UK." State of Caring Report, London, 2019.

[9] M. Househ, E. Borycki, and A. Kushniruk, "Empowering patients through social media: The benefits and challenges," Health Informatics Journal, vol. 20(1), 2014, pp. 50–58.

[10] R. C. Talley and J. E. Crews, "Framing the public health of caregiving," American Journal of Public Health, vol. 97, no. 2. 2007, pp. 224–228, doi: 10.2105/AJPH.2004.059337.

[11] T. Robbins et al., "COVID-19: A new digital dawn?," Digit. Heal., vol. 6, 2020, pp. 1-3.

[12] K. Krishna, "A heavy responsibility for Dr H. Krishna Kumar," Fit for Life "People" Sunday Star, 3 August, 2014, p.10.

[13] B. S. Russell, M. Hutchison, R. Tambling, A. J. Tomkunas, and A. L. Horton, "Initial challenges of caregiving during COVID-19: Caregiver burden, mental health, and the parent–child relationship," Child Psychiatry & Human Development, 51(5), 2020, pp. 671-682.

[14] C. K. Looi, P. Seow, and B. Zhang, "Leveraging mobile technology for sustainable seamless learning: A research agenda," British Journal of Educational Technology, vol. 41(2), 2010, pp. 154–169.

[15] J. Zarocostas, "How to fight an infodemic," Lancet, vol. 395, no. 10225, 2020, p. 676.

[16] S. H. Tey, "Long term caring journey: How to comfortably be yourself?" [translated from Chinese] AiFM, July 12, 2021.https://web.facebook.com/watch/live/?v=1162201980960056&ref=watch_permalink [retrieved: August, 2021].

[17] R. M. Razali, P. C. Bee, and G. G. Gan, "Survey of willingness to accept chemotherapy among elderly Malaysian patients," Asian Pacific Journal of Cancer Prevention, vol. 14(3), 2013, pp. 2029–2032.

[18] N. Ribeiro, L. Moreira, A. M. P. Almeida and F. Santos-Silva, "Mobile seamless learning tool for cancer education," In Ó. Mealha, M. Divitini, M. Rehm (eds) Citizen, Territory and Technologies: Smart Learning Contexts and Practices. SLERD 2017. Smart Innovation, Systems and Technologies, vol. 80, 2018, Cham: Springer.

[19] Seattle Cancer Care Alliance, "Caregiver companion guide," 2018. https://www.seattlecca.org/sites/default/files/inline-files/Caregiver%20Guide_1.pdf [retrieved: November, 2021].

[20] P. Sercekus, D. B. Besen, and N. P. Gunusen,, "Experiences of family caregivers of cancer patients receiving chemotherapy," Asian Pacific journal of cancer prevention : APJCP, vol. 15, no. 12, 2014, pp. 5063–5069.

[21] WHO, "Cancer country profile 2020: Malaysia burden of cancer," 2020. https://www.who.int/cancer/country-profiles/MYS_2020.pdf [retrieved: October, 2021].

[22] A. B. Sullivan and D. Miller, "Who is taking care of the caregiver?", Journal of Patient Experience, vol. 2(1), 2015, pp. 7–12.

[23] V. Venkatesh, "Impacts of COVID-19: A research agenda to support people in their fight," Int. J. Inf. Manage., vol. 2020, no. 55, 2020, pp. 1–6.

[24 E. Y. N. Yuen, S. Dodson, and R.W. Batterham, "Development of a conceptual model of cancer caregiver health literacy," European Journal of Cancer Care, vol. 25, no. 2, 2016, pp. 294-306.

[25] I. Veuillotte, G. Morel, and S. Pitois, "General practice and the Internet revolution. Use of an Internet social network to communicate information on prevention in France," Health Informatics Journal, vol. 21, no. 1, 2015, pp. 3–9.

[26] E. Y. N. Yuen, T. Knight, L.A. Ricciardelli, and S. Burney, "Health literacy of caregivers of adult care recipients: A systematic scoping review," Health and Social Care in the Community, vol. 26, no. 2, 2018, pp. e191- e206.

[27] Cancer.Net Website https://www.cancer.net/ [retrieved: August, 2020].

[28] Skill For Care Website https://www.skillsforcare.org.uk/About/About-us.aspx [retrieved: August, 2021].

[29] Dignity In Care Website. https://www.dignityincare.org.uk/About/ [retrieved: August, 2021].

Data Merging for the Study of Eye Diseases in Relation to Demographic and Weather Factors in Telangana State, India

Amna Alalawi[1], Les Sztandera[2], Dr. Richard Derman[3], Parth Lalakia[4], Anthony Vipin Das[5], Gumpili Sai Prashanthi[6]

[1]DMgmt in Strategic Leadership Program, Thomas Jefferson University, Philadelphia, PA, USA
Email: Amna.Alalawi@Jefferson.edu

[2]Kanbar College of Design, Engineering, and Commerce, Thomas Jefferson University, Philadelphia, PA, USA
Email: Les.Sztandera@Jefferson.edu

[3]Global Health Research, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA
Email: Richard.Derman@jefferson.edu

[4]Global Affairs, Thomas Jefferson University, Philadelphia, PA, USA
Email: Parth.Lalakia@jefferson.edu

[5] Department of EyeSmart EMR & AEye, Indian Health Outcomes Public Health and Economics Research Centre (IHOPE), LV Prasad Eye Institute, Hyderabad, Telangana, India
Email: vipin@lvpei.org

[6] Department of EyeSmart EMR & AEye, Indian Health Outcomes Public Health and Economics Research Centre (IHOPE), LV Prasad Eye Institute, Hyderabad, Telangana, India
Email: saiprashanthi.g@lvpei.org

***Abstract* – Big data is the new gold in today's VUCA world. VUCA is a systems-derived acronym used to refer to a volatile, uncertain, complex, and ambiguous system. Taking a holistic approach in studying and examining big data can yield new insights that can benefit healthcare practitioners and decision-makers in a VUCA system. This study explores advancing the method to collect and process Electronic Medical Records (EMR), to expand the interest in the use of big data in health care. The objective of this study was to determine whether the EMR record in LV Prasad Eye Institute (LVPEI) in India can contribute to the management of patient care, through studying how climatic and socio-demographic factors relate to eye disorders and visual impairment in the State of Telangana, using data merging as the main technique to study the data. Our findings highlighted the presence of cataract in rural areas and throughout the different weather seasons. Men tend to be the most, while home makers make the most visit to the hospital.**

Keywords –big data; ophthalmology; ocular diseases; artificial intelligence.

I. INTRODUCTION

This extended research is based on the critical global healthcare issue of examining the demographic and weather factors that correlate to the development of eye diseases in Telangana, India. A brief presentation of the topic was addressed at the Ninth Data Analytics 2020 Conference in Nice, France [1]. Responding to the major interests from the discussions in Nice, this research includes more statistical analysis of the relationship, as well as a detailed explanation of the systems-approach method used to study the data.

India is home to over 8.3 million people with Vision Impairment (VI), the highest in the world [4]. Even though, in 1976, India became the first country in the world to start a national program for control of blindness for the goal to reduce blindness prevalence to 0.3 percent by 2020, the prevalence of blindness still stands at 1.99 percent, according to the National Blindness and Visual Impairment Survey, released in October 2019 by the Union Ministry of Health and Family [2]. The prevalence of blindness and visual impairment is one of the highest in Telangana, a state in Southern India, as inferred from survey [2]. The significant reasons indicated in the survey were due to cataract and refractive error [3].

All surveys in the country have shown that cataract is the most common cause of blindness and all prevention of blindness programs have been "cataract-oriented."

However, it has recently been recognized that the visual outcome of the cataract surgeries, as well as the training of ophthalmologists has been less than ideal.

This study uses Artificial Intelligence (AI) and machine learning techniques to explore a dataset containing information on 873,448 patients who visited LV Prasad Eye Institute (LVPEI), a multi-tier ophthalmology hospital network, based in Hyderabad. The data was extracted from Eye Smart, the name given to the hospital's Electronic Medical Record (EMR) and health management system. The EMR was then merged with climatic factors to test the correlation between climatic variables and ocular diseases presented by the patients [4]. Studying risk factors, primarily associated with climate and the environment can lead to a better understanding of the causes, diagnosis, and treatment of several eye diseases [6].

In healthcare, ophthalmology deals with the diagnosis and treatment of eye disorders. Some known diseases in ophthalmology are cataracts, retinal disorders, macular degeneration, and others. The relatively rapid and recent adoption of EMRs in ophthalmology has been associated with the promise that the accumulation of large volumes of clinical data would facilitate quality improvement and help answer a variety of research questions. Given that EMRs are relatively new in most practices and that clinical data are inherently more complex than other fields that have been altered by the digital revolution, these proposed benefits have yet to be realized [5].

With the rise of big data, it has now become easier to study how culture, race, climate, and other socio-demographic factors correlate to the spread of ocular diseases. This has shed light on recent research in medicine and ophthalmology. An investigation has been conducted with an aim for the application of AI-based hierarchical clustering as a tool to optimize the in total excellence, values, and the security for the Adult Spinal Deformity Surgery (ADS) [17]. It has been observed that prior to this the ADS classification was based on certain radiographic parameters which have been correlated with the patient related outcomes. But the problems faced immensely by the researchers is to separate out the patients and the patterns manually that is in turn was based on hundreds of data parameters and the process was considered to be practically not feasible.

Therefore, as a methodological approach for every probable cluster of patient (N) done on the basis of (M) surgery were normalized for two year enhancement and the massive rate of complication were computed. Thus, this particular study has therefore, highlighted that unsubstantiated hierarchical clustering of the patterns of findings that helps to initiate the prior operative judgment making by formulating a two-year risk benefit grid. In this way the novel AI-ASD pattern of classification and identification have helped to diminish the risk and overall improvement [17]. It should be mentioned that smaller cities face a lot of difficulties to maintain the sustainable welfare of countries in amalgamation with notable standards of living [18].

In another investigation, the emergency bases of hospitalization along with the causes of mortality were utilized as the replacements of frailty. The researchers used to two different models to assign the deteriorating risk score to every subject of the elderly population residing within the Municipality of Bologna, Italy [18]. The study design was a cohort study with of 58 789 subjects as sample size for overall six years with a four-year monitoring period. The study findings reported excellent power of discrimination along with calibration that demonstrated an excellent anticipating ability of the models utilized [18].

In this respect, another study could also be illustrated that had utilized the application of health administrative databases along with authenticated algorithms to show a correlation in between the residential proximity towards foremost roadways along with the prevalence of three major neurological diseases like dementia, multiple sclerosis, and Parkinson's disease [19]. This particular study design was also a cohort study based on two different populations having the age group in the range of 20–50 years with sample size of 4.4 million suffering from multiple sclerosis and the matured adult groups having the age range of 55–85 years with sample population size of 2·2 million suffering from dementia or other Parkinson's disease. The researchers of the study estimated the associations among the following parameters such as the traffic proximity, incidence of dementia, Parkinson's disease, and the multiple sclerosis using the model of Cox proportional hazards which would also take into consideration of certain individualistic contextual parameters such as any injury to brain, diabetes, and the local income [19].

The study demonstrated the successful application of the health record databases along with the specialized analytical tool for the categorization of patterns of large or big data [19]. In today's world, the relationship based on function in between the non-coding RNA (ncRNA) and the varied types of human diseases, is considered to be a central task within the field of modern research for the formulation of effective therapeutic approaches.

In Section 1, we discuss the methodology used in the study, breaking the sections down to explaining the systematic approach used as a framework for studying the date, and an explanation into the computer software used, Microsoft Power BI, to explain the data merging technique. In Section 2, we discuss the analysis and highlight the specific statistical tools and methods used to investigate the study of several demographic and climatic factors that impact upon the individuals in Telangana. Section 3 focuses on providing a thorough analysis of the data findings. Section 4 highlights key findings and trends, and Section 5 includes the conclusion and recommendations pertaining to the use of big data and data merging in EMR to reveal new insights in the study of visual impairment and eye disorders in Telangana.

II. METHODOLOGY

A. Systematic-Approach Using the DIKW Pyramid

Data science, a discipline that has been emerging over the past few years, centers on analyzing data. Since there is no specific definition of data science, we relied on the explicit meaning of the term, and therefore decided to work on complex data sets and determine a way through which it can be evaluated. To make the process simpler, we decided to study the data from a holistic approach. Putting a framework in place, helped us in creating a plan for studying the data to achieve results that can be of use and that can identify and analyze patterns in the data sets, which can be used for future clinical practices.

The following Data, Information, Knowledge, and Wisdom (DIKW) pyramid framework was chosen as the main framework that explains the reasons for transforming the data. By transforming the data into information, we gained certain knowledge about the topic which then was transformed into wisdom that helped us in not only conducting the investigation effectively, but also to gain effective understanding about the research topic. The readers can also rely on this wisdom to gain conceptual clarity and understanding of the subject matter.

The DIKW Pyramid

Figure 1. DIKW Pyramid

Figure 1 depicts the DIKW Pyramid. Dr. Russel Ackoff, wrote that "wisdom is located at the top of a hierarchy of types," suggesting that as the highest level of the hierarchy of types, wisdom is somehow superior to the types below it. Subsequent depictions of the knowledge hierarchy typically exclude the understanding level [22]. Ackoff's hierarchy does not require that data transform into information, information into knowledge, or knowledge into wisdom. Instead, he stated that each category is included in the next. For example, there can be no wisdom without understanding and no understanding without knowledge [22]. Big data analytical tools have a strong potential that allows healthcare professionals to gander upon the clinical data stored within repositories and assist in the process of informed making of decisions. Today, especially with the COVID-19 situation, the healthcare sector is making use of AI for a wide range of findings.

Social researchers are now expected to possess various skills and abilities, in particular, the ability to apply qualitative methods. Such researchers are also expected that they acknowledge as well as explain their personal positioning. This is mainly due to ‘subjectivity’. Every person has their own thoughts and opinions; due to this reason, another researcher exploring the same topic might have a different perspective which will be seen through their investigation. Such differences in perceptions and ‘frames of mind’ influence the way the study is carried out and understood by the readers [24]. As Morison [24] has written: “Sociological research is a complex enterprise involving a dynamic interplay between personal values, theories and practical data gathering skills. Different sociologists, looking at the same community but not starting from the same theoretical viewpoint, may direct their attention to different aspects of the place they are studying and come up with extremely contrasting results.”

B. Data-Merging Using Microsoft Power BI

To gain insight into the climatic and socio-demographic factors that correlate to the risk of ocular diseases in the State of Telangana, we used multiple approaches utilizing AI and statistical software and programming languages, including Microsoft Power BI and Python to explore the dataset, which contained information on 873,448 patients complaining of eye disorder symptoms across multiple categories of ocular diseases. Publicly available climatic variables were obtained and aligned to the dataset through a process called column mutation, and then examined by Microsoft Power BI, which heavily relies on visual illustrations and statistical storytelling to present findings and new insights. It should be noted that Microsoft Power BI is considered an assortment of software or apps which all together works in amalgamation to transform the unrelated sources of data into a visually pattern oriented, continuous and dynamic insights. Column mutation, which is the merging of datasets, was done through Python, an interpreted, object-oriented programming, that codes the columns in a language called Syntax. It works out on the principle of logical and arithmetic computation. This tool has an advantage to handle large and complex datasets.

The process was however timely given the large volume of patients. The process was repeated more than once to ensure minimal error in the merging process. Microsoft Power BI was then used to model the data in order to obtain the visualizations and insights.

III. Analysis

It has proven valuable to first observe which diseases are the most prevalent in the different areas of Telangana, and what age and gender are most affected to get a full understanding of the criticality of the eye disorder epidemic and to provide a baseline against which to compare the climate and patient demographic variables examined.

The use of EMRs in generating new insights has been an increasing trend in the area of ophthalmology. Research in ophthalmology has benefited greatly from the use of EMRs in expanding the breadth of knowledge in areas such as disease surveillance, health services utilizations and outcomes. In addition, the quantity of data available has increased, that it is now highly recommended to work on data linkage systems in eye research, as such data can offer insights into advantages and limitations for future direction in eye research [6].

The timespan of this dataset is between 2011-2019, a total of 8 years. There has been consistency to already known information through the analysis, specifically on gender and age-related eye conditions. Creative featuring techniques have been used to shed the light on the most critical variables through trial and error of testing for relationship in accordance to eye disease.

The master dataset or the big data, which was explored and analyzed, covered clinical visits are from the year 2011-2019, and included demographic information of the patient, including age, gender, profession, data of visit, district of resident, and symptoms and diagnosis of the patient in relation to eye disease. To look further into this issue, we merged climate variables to the dataset to explore the relationship with eye disorders. The AI approach can be of varied types namely conventional symbolic AI, Computational intelligence, and statistical tools or the combination of all of the above. Here in the present assignment the Computational intelligence approach has been adopted for the analytical purpose [23].

The climate variables we examined were average temperature, minimum temperature, maximum temperature, humidity, rainfall, and solar radiation. This data was retrieved from the Telangana State Development Planning Society in the state of Telangana. The findings that relate to temperature and its effect on Cataract in older age was consistent in high and low temperatures.

IV. Findings and Trends

This section highlights key findings of the study, as well as trends in relation to the subject matter as per the demographic and climatic variables tested.

A. Gender and Eye Disorders

First, we broke down the data to understand which gender in Telangana has a higher rate of risk in developing eye diseases. This break- down helped us to understand the cultural background of Indian culture among men verses women.

Gender and Eye Disorders

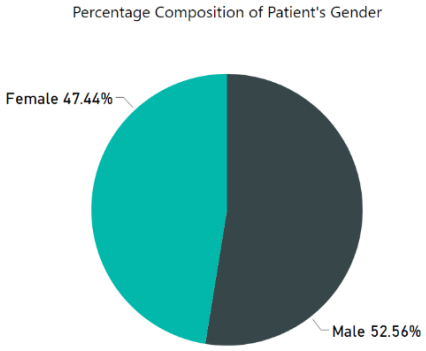


Figure 2. Clinical Visits by Gender (2011-2019)

Figure 2 indicates that between the years 2011-2019, 53% of the patients were male patients who were seen for eye disorders, and 47% were female patients. This finding is in line with the gender study that was conducted on 2.3 million patients of all those who presented to LV Prasad Institute from the years 2011-2019 [4]. Globally, one of the social determinants of health that has been universally identified is gender. In India, health inequalities between men and women have played a pivotal role in disease development, including eye disorders. With respect to eye care, women have been generally cited to have higher rates of blindness in India and are less likely to access appropriate eye services [7][8]. However, as we can see from the study which was focused on Telangana, this is not the case, as male patients exceeded female patients, and this could be for the reason that Telangana has been ranked as one of the top ten innovative and developed states in India according to the India Innovation Index 2019 [15] where access to healthcare is available and appreciated by both male and female.

India has been one of the countries where efforts to strengthen the evidence-base for blindness control has received significant attention from policy planners and program managers. Over the past four decades, a series of population-based blindness and visual impairment surveys have been undertaken in India, using different survey methods. This included detailed eye examination surveys, as well as rapid assessments [8].

B. Occupation and Clinical Visits

Second, to dig deeper into understanding the culture, we studied the occupation of the patients to determine what could be posing a high risk in terms of where someone

spends the day the most. The study showed that home makers, employees in the government and private sectors, and students make the top three categories who visited the clinic the most.

Percentage of Clinical Visits by Occupation

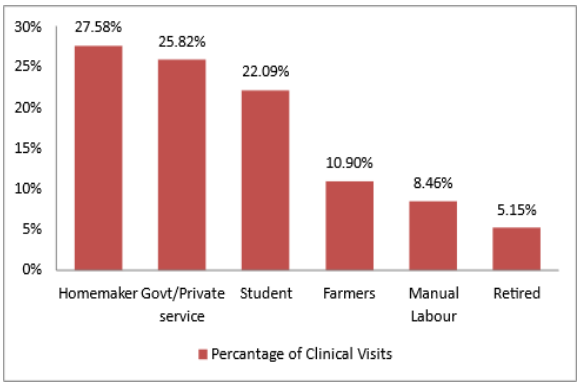


Figure 3. Clinical Visits by Profession (2011-2019)

Figure 3 depicts this analysis and portrays the top six professions taken from the analysis. We can also see that workers in Agriculture and manual laborers tend to present themselves with eye disorders as well, and that could be to the nature of the job, in which they are exposed to certain chemicals, dust, and usually work in heated environments. Recent estimates from the World Health Organization indicate that 90 per cent of all those affected by visual impairment live in the poorest countries of the world [9]. India is home to one-fifth of the world's visually impaired people and therefore, any strategies to combat avoidable blindness must take into account the socio-economic conditions within which people live [9].

Home makers could also translate to housewives, who are at higher-risk of visual disorders, and this is in line with a study that was conducted in 2009 on women in Indian culture, where it showed that housewives are more likely to suffer from heart diseases than working women, and that is due to lack of education, lifestyle that is based on obesityand cultural myths that do not focus on women's health. Having a similar study related to eye disorders and visual impairment, as per the study based on the sample of the population from Telangana, the same pattern can be seen and it can potentially be from these similar reasons [10].

C. Location and Eye Disorders (Older Age 41-70)

Third, we broke down the data to understand the location of the patient, and whether there are underlying reasons in a specific location for a higher rate of risk to developing eye diseases. The study was broken down into examining the location by different age groups.

Number of Visits by Location (Age 41-70)

Location	Number of Visits
Paloncha	5471
kothagudam	2930
kothagudem bazar	2684
Manuguru	2380
Bhadrachalam	2287
Yellandu	1867
Adilabad	1634
Tekulapalli	1432
Burgampahad	1309
Madhapur	1126

Figure 4. Location and Eye Disorders in Older Age Population

Figure 4 shows the locations that people with eye disorders come from and is focused on the older age population. Cataract seemed to be the most disease that has affected older age in Telangana, which is the clouding of the natural human lens. Cataract is a condition known to affect older age, and this study revalidates the information.

D. Location and Eye Disorders (Younger Age 11-20)

The second age group that was studied as per the relation to the location of the patient was the younger population, who we defined from the data as being between the ages 11-20.

Number of Visits by Location (Age 11-20)

Location	Number of Visits
Paloncha	601
Madhapur	325
Adilabad	256
Bhadrachalam	226
kothagudam	177
kothagudem bazar	168
Nagarkurnool	132
Manuguru	128
Gachiowll	127
Kondapur	116
Yellandu	116

Figure 5. Location and Eye Disorders in Younger Age Population

Figure 5 shows the location that people with eye disorders come from and is focused on the younger age population. Astigmatism, which is an irregularity of the shape of the cornea was present in younger age population. Astigmatism has been linked to being a hereditary condition in ophthalmology.

In both contexts, it appeared to be that eye disorders are mostly concentrated in residents from the district of Paloncha, and even though this district has a higher literacy rate than state average is 77%, 10% higher than that of the state average which is at 67%, it has been reported that it has been hit with pollution and contaminated water in 2015. The state-run thermal power plant installed in 2015 caused pollution and health disorders including eye disorders [11]. Residents complained of gray water, and doctors in Paloncha confirmed that the prolonged exposure to air and water pollution has led to higher incidences of respiratory diseases, tuberculosis, skin diseases, blurring of vision and irritation in the eyes, such as Cataract, Cornea, Anterior Segment, Retina, and Glaucoma [11].

E. Consistent Prevalence of Cataract in Rainfall

Fourth, we broke down the data to understand the correlation between the number of visits and rainfall. The sample data was huge, and therefore we limited the study to understand right-eye diseases, as the classifications in Eye Smart were according to right-eye diseases and left-eye diseases separately.

Number of Visits for Right-Eye Diseases by Cumulative Rainfall

Right eye Diseases	Cumulative Rainfall	Number of Visits
Cataract	2.11	4126
Myopia	2.11	8058
Emmetropia	2.10	15253
Pseudophakos	2.12	5837
Presbyopia	2.15	5820
Senile Cataract	2.05	8134
Intraocular Lens	2.03	3146
Hypermetropia	2.13	4164
Simple Myopia	2.12	7686
Astigmatism	2.12	4172

Figure 6. Right Eye Diseases in Relation to Rainfall

Figure 6 shows the diseases that affect the right eye the most when tested alongside rainfall. Globally, cataract is the single most important cause of blindness, and the second most common cause of Moderate and Severe Vision Impairment (MSVI) according to the Global Burden of Disease, Injuries and Risk Factors Study, and it is most predominant in Southeast Asia. Cataract contributed to a worldwide 33.4% of all blindness and 18.4% of all MSVI. Translating the same into actual numbers, cataract caused blindness in 10.8 million of overall 32.4 million blind and visual impairment in 35.1 million of 191 million visually impaired individuals [13].

The close relationship between climate, environment and the development of Cataract is crucial to understand for future preventative measures. In Telangana, it shows that Cataract is the disease most prevalent in rainfall.

F. Consistent Prevalence of Pterygium in Relation to Global Radiation

Fifth, we broke down the data to understand the correlation between right-eye diseases and global radiation. Similar to the previous study of number of visits and rainfall, the sample data was huge, and therefore we limited the study to understand right-eye diseases.

Right-Eye Diseases as Influenced by Global Radiation

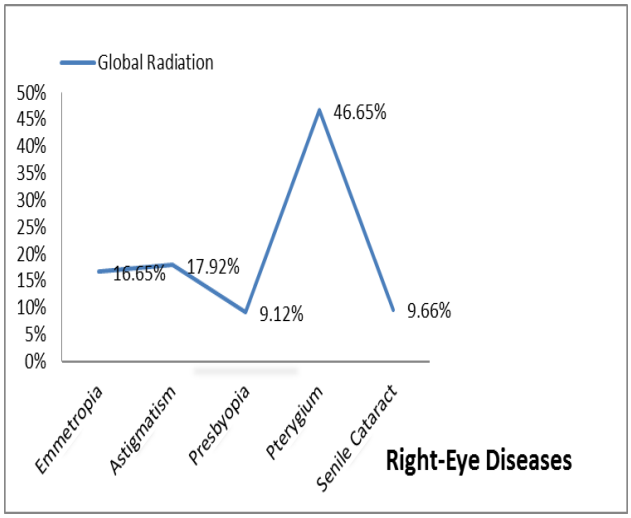


Figure 7. Right Eye Diseases in Relation to Global Radiation

Figure 7 shows the diseases that affect the right eye the most when tested alongside global radiation. We analyzed patients who presented with degeneration symptoms, and correlated the diagnosis to climatic factors, such as humidity, rainfall, temperature and global radiation. The above analysis shows the top 5 most prevalent degeneration right-eye diseases as impacted by global radiation. Pterygium shows to be most prevalent at over 46% of the total global radiation value. The analysis was done on a

patient basis and not a disease basis, as the data showed that one patient can develop more than one disease.

G. Consistent Prevalence of Pterygium in Relation to Windspeed

Finally, we broke down the data to understand the correlation between right-eye diseases and windspeed. Similarly again, the sample data was huge, and therefore we limited the study to understand right-eye diseases.

Right-Eye Diseases as Influenced by Windspeed

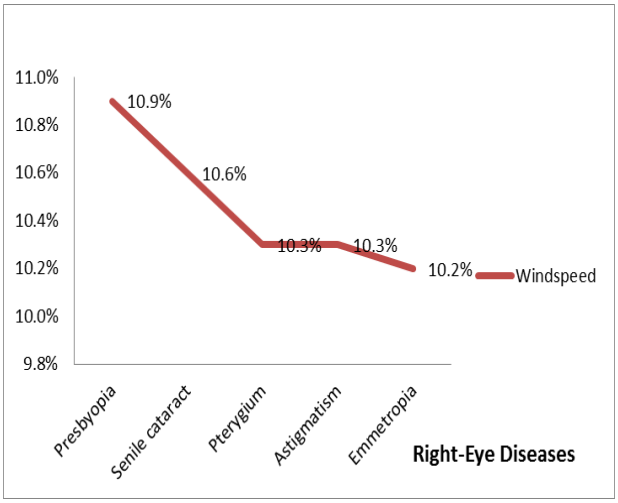


Figure 8. Right Eye Diseases in Relation to Windspeed

Figure 8 shows the diseases that affect the right eye when tested alongside windspeed. The analysis above shows the top 5 most prevalent right-eye diseases with degeneration as a symptom and how the diseases are influenced by maximum windspeed. Presbyopia, Cataract, and Pterygium were the most present among patients and concentrated at average maximum windspeed of between 10.2 and 10.9.

As concluded from the research, the healthcare EMR system is large and complex, one that does not naturally lend itself to easy analysis, design or even understanding. Therefore, the complexity and critical nature of the system beg for the development and use of good, representative models [25]. As per the case study of the research, to analyze the data in Eye Smart, an influence diagram was created to show the different co-factors that lead to cataract according to the findings generated. This was further analyzed using a systems- thinking approach, a method that allows consideration of the whole rather than individual elements of representation of the related co-factors. The influence diagram was useful to showcase the summary of the findings of the research, as it allows for us to see the connection of the variables in a picture format. This format of summary is useful for healthcare practitioners, and in the case of our study, for ophthalmologists, and decision-makers, in the field of eye-care management.

Influence Diagram - Taking Knowledge to Wisdom

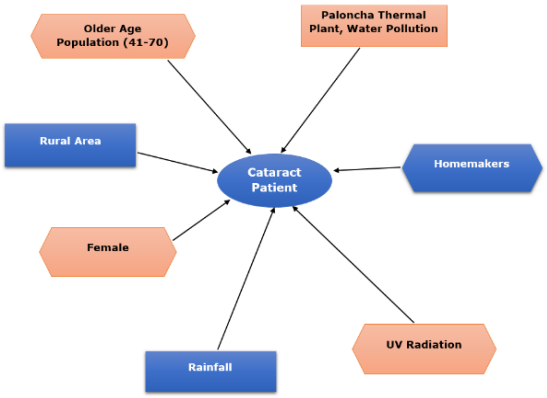


Figure 9. Influence Diagram of Cataract Patients and Co-Factors Affecting Cataract

Figure 9 depicts the influence diagram of the main study. Influence diagrams are closely related to decision trees and often used in conjunction with them. An influence diagram displays a summary of the information contained in a decision tree. It involves four variable types for notation: a decision (a rectangle), chance (an oval), objective (a hexagon), and function (a rounded rectangle). Influence diagrams also use solid lines to denote influence. In the case of the influence diagram generated below, this has been equivalent to the definition of creating “wisdom” in the DIKW pyramid explained earlier in previous chapters. Looking at the data analysis holistically creates the ease to depict the main causes of the development of cataracts. Figure 9 is a drawing of the influence diagram that explains the connection of the weather factors in relation to the development of the most right-eye disease found in the sample population of the study, which is cataract.

V. Conclusion and Future Work

This data analytics study provides an expanded exploration of how socio-demographic and climatic factors affect the prevalence of visual impairment and eye disorders in Telangana. Applying several statistical techniques, including pattern recognition, and generating other data visualizations, we were able to validate previously identified findings about gender’s relation to eye disorders in Telangana. We found the tools we used to be very useful for a discovery research to better understand the sample set of patients and to generate informative and understandable visuals.

Big data can serve to boost the applicability of clinical research studies into real-world scenarios, where population, race, and climate create a challenge. It equally provides the

opportunity to enable effective and precision medicine by performing patient stratification. This is indeed a key task toward personalized healthcare. A better use of medical resources by means of personalization can lead to well-managed health services that can overcome the challenges of a diverse population where poverty is high. Thus, creative featuring and data merging for health management of EMRs can have an impact on future clinical research.

From a systems perspective, we observe that a patient is influenced by several co-factors that result in the development of eye disorders, and that is significant in studying patient care from a holistic standpoint. AI tools create the pathway to merging publicly available data and aligning multiple variables as part of the overall influence. This technique is widely applied in decision-making and outcome assessment for an enhanced healthcare experience, in which modeling knowledge and expert experience are studied more thoroughly for new pattern recognition. However, variables must be minimized in order to capture the underlying knowledge, or otherwise patterns will be harder to spot. Thus, we attempt to apply this in the future with less variables to overcome the challenges in the first phase or data merging.

We recommend that the authorities spend more time and funds on creating awareness to educate individuals and families about the visual impairment crisis in Telangana. Creation of awareness is one of the most comprehensive approaches to sensitize communities concerning the consequences of eye disorders, but also one of the avenues to equip individuals with knowledge, skills, and correct attitudes towards a healthier lifestyle.

Besides creation of awareness, this study also recommends ophthalmologists' understanding of all factors that influence a disease other than medical history, and to look at each patient uniquely in terms of social income, cultural upbringing, and offer a more individualistic approach in educating a patient from the criticality of self-care, to help patients deviate away from high risk situations that can cause eye disorders, and to find ways from an earlier age for more effective preventative results that can reduce the number of affected individuals with vision impairment in Telangana.

Visual impairment has continually exhibited an escalating trend in underdeveloped countries over the past years, and in India, the burden of visual impairment is high in urban and rural areas. Eye-care services should be accessible and affordable to individuals in need. Future work will be on discovery research on how socio-demographic and climatic factors correlate to the development of other diseases. It is important to mention that hardly any investigation had been conducted with the application of AI tools to categorize huge data based on demographic and weather variables. Therefore, this study adds new insights to the field of ophthalmology.

REFERENCES

[1] A. Alalawi, A. Das, R. Derman, P. Gumpili, P. Lalakia, L. Sztandera, "A Data-Driven Approach for Eye Disease Classification in Relation to Demographic and Weather Factors Using Computational Intelligence Software," The Ninth International Conference on Data Analytics, IARIA, Proceedings of 2020. pp. 21-27.

[2] G. Rao, N. Khanna, and A. Payal, "The global burden of cataract," *Current Opinion in Ophthalmology*, *22* no. 1, pp. 4-9, 2011.

[3] A. Das, P. Kammari, and S. Ranganath Vadapalli, "Big data and the eyeSmart electronic medical record system-An 8-year experience from a three-tier eye care network in India," *Indian Journal of Ophthalmology*, *68* no. 3, pp. 427, 2020.

[4] S. Marmamula, R. Khanna, and G. Rao, "Unilateral visual impairment in rural south India–Andhra Pradesh Eye Disease Study (APEDS)", *International Journal of Ophthalmology*, *9* no.5, pp.763, 2016.

[5] M. V. Boland, "Big data, big challenges," Ophthalmology, *123* no.1, pp. 7-8, 2016.

[6] A. Clark, J. Ng, N. Morlet, and J. Semmens, "Big data and ophthalmic research," *Survey of Ophthalmology*, *61* no.4, pp. 443-465, 2016.

[7] E.M. Messmer, "The pathophysiology, diagnosis, and treatment of dry eye disease," *Deutsches Ärzteblatt International*, *112* no.5, pp. 71, 2015.

[8] D. Matthews, "How gender influences health inequalities," *Nursing Times*, *111* no.43, pp. 21-23, 2015.

[9] N. Diaz-Granados, et al. "Monitoring gender equity in health using gender-sensitive indicators: a cross-national study," *Journal of Women's Health*, *20* no.1 pp. 145-153, 2011.

[10] G.V.S Murthy, S.K. Gupta, D. Bachani, R. Jose, and N. John, "Current estimates of blindness in India," *British Journal of Ophthalmology*, *89* no.3, pp.257-260, 2005.

[11] S.K Angra, G.V. Murthy, S.K. Gupta, and V. Angra, "Cataract related blindness in India and its social implications," *The Indian Journal of Medical Research*, *106*, pp. 312-324, 1997.

[12] A. Janati, H. Matlabi, H. Allahverdipour, M. Gholizadeh, and L. Abdollahi, "Socioeconomic status and coronary heart disease," *Health Promotion Perspectives*, *1* no.2, pp. 105, 2011.

[13] S. Kumar, "National security environment," *India's National Security*, pp. 25-236, Routledge India, 2016.

[14] S.G. Honavar, "Eliminating cataract blindness: Are we on target?" *Indian Journal of Ophthalmology*, *65* no.12, pp. 1271, 2017.

[15] G. Bakshy https://www.jagranjosh.com/current-affairs/india-innovation-index-2019-karnataka-tops-the-list-followed-by-tamil-nadu-and-maharashtra-1571375106-1, 2019, [Online]. (Accessed on 07.09.2020).

[16] C.P. Ames, et al. "Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value," *Spine*, *44* no.13, pp. 915-926, 2019.

[17] F. Bertini, G. Bergami, D. Montesi, G. Veronese, G. Marchesini, and P. Pandolfi, "Predicting frailty condition in elderly using multidimensional socioclinical databases," *Proceedings of the IEEE*, *106*(4), 723-737, 2018.

[18] H. Chen, J. Kwong, R. Copes, P.J. Villeneuve, A. Van Donkelaar, and A.S. Wilton, "Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study," *The Lancet*, *389*(10070), pp. 718-726, 2017.

[19] E.P. Barracchia, G. Pio, D. D'Elia, and M. Ceci, "Prediction of new associations between ncRNAs and diseases exploiting multi-type hierarchical clustering," *BMC bioinformatics*, *21* no.1, pp. 1-24, 2020.

[20] B. Familiar and J. Barnes, "Data Visualizations, Alerts, and Notifications with Power BI. In Business in Real-Time Using Azure IoT and Cortana Intelligence Suite," pp. 397-473, *Apress*, Berkeley, CA, 2017.

[21] M. Lutz, "Programming Python: powerful object-oriented programming," *O'Reilly* Media, 2010.
[22] G. Marcus and E. Davis, "Rebooting AI: Building artificial intelligence we can trust," *Pantheon*, 2019.
[23] Russell L. Ackoff, "From data to wisdom," Journal of applied systems analysis, 16.1. pp. 3-9, 1989.
[24] A.P. Brown, "Qualitative method in applied research," 10 (2), pp. 229, 2010.
[25] M. Longman, London, Methods in Sociology, 1986.
[26] K. Patrick, and M. J. Schwandt. "Health systems: A dynamic system-benefits from system dynamics," Proceedings of the Winter Simulation Conference, 2005, IEEE.

.

Rule-Based Detection of Health-related Problems of People with Dementia from Lifestyle Wearables: The support2LIVE Approach

Lampros Mpaltadoros, Vasilis Alepopoulos, Antonios Pliatsios, Angelos V. Vasileiadis,
Thanos G. Stavropoulos, Spiros Nikolopoulos, and Ioannis Kompatsiaris
Information Technologies Institute (ITI)
Centre for Research & Technology Hellas (CERTH)
Thessaloniki, Greece
{lamprosmpalt, valep, apliatsios, angelvasileiad, athstavr, nikolopo, ikom}@iti.gr

Abstract— **In this paper we describe a rule-based framework for the detection of Health-related Problems of people with dementia, developed in the support2LIVE project. The framework combines a novel ontology for lifestyle data (steps, sleep duration and heart rate measurements) and Health-related Problem representation, a framework for IoT data collection from sensors, a Knowledge Ingestion component that links the data through meaningful relationships and a novel set of SPARQL Inferencing Notation (SPIN) Rules to infer problems from these data. Both the ontology and the rule set are designed based on clinical expert knowledge in the field of dementia. More specifically, lifestyle data is acquired from wearable devices available in the market, making the system affordable and convenient. A model based on Semantic Web technology, OWL (Web Ontology Language), is used to formally represent and integrate sensor measurements, promoting interoperability with other models and data exchange. SPIN rules offer the benefit of simplicity and flexibility as opposed to other rule representations in the domain. Our framework handles the incoming IoT data from the sensors with the use of suitable APIs and contains a Knowledge Ingestion component that transforms the received data to semantic knowledge in the form of RDFs. Finally, a clinician dashboard visualizes results to allow decision making, as shown in proof-of-concept scenarios from real participants in the support2LIVE pilots.**

Keywords- Ontology; Event Detection; Semantic Web; SPIN; Reasoning; Rule-Based Systems; Dementia.

I. INTRODUCTION

As the world population is rapidly aging, people living with dementia globally amount to 50 million in 2019 and are expected to triple to 150 million by 2050 [2]. Yet there is no silver bullet for dementia, such as a pharmacological solution. Only holistic and objective information about a patient's health status can drive tailored interventions to alleviate the ailments and slow down the progression of the disease. However, this imposes a huge burden to informal caregivers and healthcare professionals to manually monitor lifestyle, and Health-related Problems such as movement, sleep and stress.

Lifestyle sensors are a promising and affordable solution to objectively, continuously and affordably monitor patients, but a framework to map and extract clinical Health-related problems is needed. The acquisition of knowledge from continuous and heterogeneous data flows is a prerequisite for Internet of Things (IoT) applications [2][3][4]. Semantic technologies provide integrated tools and methods for representing data and producing new Knowledge from them. Smart environments are increasingly encountering in healthcare technologies at home in actions that create better living conditions for older people by using IoT technologies, such as Active and Healthy Ageing (AHA) and Ambient Assisted Living (AAL). In this context, human activity recognition plays a main role [4], because it could be considered as a starting point to facilitate assistance and care for people with dementia. Due to the nature of human behavior, it is necessary to manage the time and adhere to the spatial restrictions. In doing so, semantic technologies enable expressive reasoning over health data, allowing clinical decision support to be realized. Ontologies are used to describe the context elements of interest (e.g., persons, events, activities, location, time), their pertinent logical associations [6], as well as the background knowledge required to infer additional context information.

In this paper, we propose a Semantic framework for Health-related Problem detection that combines ontologies and SPARQL Inferencing Notation (SPIN) Rules [7]. Ontologies are used to provide the common vocabulary for representing activity related contextual information, whereas SPIN rules derive high-level activity interpretations. SPIN is used as a standardized declarative language able to address the limitations of the standard OWL Semantic Web technologies mentioned previously. More specifically, the temporal relations among activities are handled by SPARQL functions, whereas the derivation of new composite activities exploits the native capabilities of SPARQL to update the underlying activity model.

The SPIN language was chosen to implement this system because it combines concepts from object-oriented languages, query languages, and rule-based systems to describe the behavior of objects on the web of data and the Internet of Things [8]. In addition, it makes the rules accessible and easy to maintain, extend and share. A suitable Reasoner tool, such as the SPIN Application Programming Interface (API), can extract the extra information generated by the rules and reuse it, for example, in executing a SPARQL query, thus generating new knowledge. These rules apply using SPARQL CONSTRUCT or SPARQL UPDATE requests (INSERT and DELETE). SPIN standards also make it possible to define such rules in higher level domain specific languages, so that rule designers do not have to work directly with SPARQL.

In [1], we presented our SPIN/SPARQL rules for health problem detection and tested them in real data coming from sensors. In this paper, we expand our framework to include components such as the IoT data collection, which collects the data coming from various sensors with the use of APIS, the Knowledge Ingestion component, which "translates" the collected data to RDF triples and the visualization dashboard, which can be used by clinicians to monitor their patients

The approach is employed in support2LIVE [9], a project which aims to integrate wearable devices and smartphone apps to support timely assessment and intervention of elders in the spectrum of dementia. A proof-of-concept scenario on how visualizations of the detected problems aids clinical decision making is demonstrated using real-world data from the project's pilot deployments in Thessaloniki, Greece.

The rest of the paper is organized as follows. Section II presents related work in the domain of ontology-based reasoning architectures in the Healthcare field. Section III describes the proposed System architecture that combines OWL ontologies and SPARQL rules in order to derive high-level activity interpretations. Section IV presents the use case scenario that evaluates the proposed architecture. Finally, Section V provides our conclusions and proposed future work.

II. RELATED WORK

In previous related studies, Semantic Web technologies have been used to represent knowledge from home healthcare systems. Related projects incorporating these technologies are KnowSense [10], COSAR [11], ACTIVAGE [12], Dem@Care [13], Faber [6], and FallRisk [14]. Table 1 summarizes the Aim of these Projects and the Methodology of semantic technology used respectively.

Table 1 - Related Work

Project	Year	Aim	Methodology
KnowSense	2015	Activity Recognition in Healthcare system	Description Logic Reasoning, (DL) for activity detection and SPARQL queries to extract clinical problems
COSAR	2011	Activity Recognition in in context-aware environments	Ontological reasoning is also combined with statistics
ACTIVAGE	2017	Development of Smart Living solutions for active and healthy aging	Interoperable Ontologies, rule-based reasoning
Dem@Care	2015	Supporting independent life for elderly people with dementia	Interoperable Ontologies, Rules, Reasoning
FABER	2015	Detect abnormal behaviors for medical applications	Simple reasoning on an ontology
FallRisk	2015	Detect falls of elderly in smart homes	Semantic Reasoning techniques

KnowSense supports monitoring of the activities of elderly people with dementia in controlled and diffused environments. Semantic Web technologies, such as OWL 2, are widely used in KnowSense in order to display sensor and specific application observations, as well as to implement solutions for identifying activities and identifying problems in everyday life activities (Instrumental Activities of Daily Living, IADLs) with the aim of clinical evaluation, in various stages of dementia. Description (Logic Reasoning, DL) reasoning for activity detection and SPARQL questions are used to extract clinical problems. On the other hand, the semantic techniques that KnowSense uses cannot be easily extended and re-used.

COSAR provides a solution based on the use of ontologies and ontological reasoning combined with statistical conclusions. Simple patient activities are identified by statistical methods, such as selecting the most likely method compared to others. Ontological reasoning is also combined with statistics to identify complex activities that cannot be apparent only by statistical methods.

ACTIVAGE is a large-scale pilot project, which aims to develop Smart Living solutions that positively affect active and healthy aging. The ACTIVAGE IoT Ecosystem Suite (AIOTES) project, is a set of techniques, tools and methodologies (rule-based reasoning, interoperable ontologies, etc.) that increases semantic interoperability at different levels between heterogeneous IoT platforms. The approach uses different mechanisms of reasoning that can improve the understanding of patients' heterogeneous data and help generate new knowledge by providing services to end users.

Dem@Care provides a complete system consisting of heterogeneous sensors, designed to support independent living for the elderly with dementia or similar health problems. This approach incorporates a heterogeneous set of detection methods and technologies, including video, audio, in addition to normal, environmental and other measurements. Semantic technologies (e.g., rule-based reasoning) are used to process and analyze sensor data according to user requirements. This leads to feedback and decision support, which is communicated to end users through appropriately designed user interfaces. A variety of clinical scenarios and environments are supported, from short-term trials in hospital settings to long-term monitoring and support of daily living at home, for independent living.

FABER is a pervasive system designed to detect abnormal behaviors for medical applications. It first computes events and actions from the available context data by using simple reasoning on an ontology. Computed boundaries, actions and events are sent to the knowledge-based inference engine.

In **FallRisk**, the main objective is to detect falls of elderly in smart homes. It relies on a platform that uses multiple learning-based fall detection systems. The results of these systems are filtered and put into an ontology that carries the context knowledge. The knowledge, including contextual information about the user, is then used to refine the fall detection. The strength of this approach, besides the combination of both techniques, is the compatibility with any fall detection technique. However, it solely deals with fall detection.

The above-mentioned systems use semantic rule-based mechanisms and provide solutions for activity and event recognition based on the use of ontologies and ontological reasoning. However, most methods are quite sophisticated and complex to express and to maintain due to rich logic support. For this reason, the SPIN language was selected for this study to create semantic rules. SPIN offers multiple advantages [6] [7]. SPIN rules offer the benefit of simplicity and flexibility as opposed to other rule representations. Moreover, SPIN is based on SPARQL, a well-established query language and protocol, which is supported by numerous engines and databases. Therefore, SPIN rules can be directly executed on the databases and no intermediate engines with communication overhead need to be introduced. Moreover, it has an object-oriented model that leads to better maintainable solutions. Specifically, the SPIN rule engine does not have to check all rules at all times, but instead rules are checked incrementally when new instances of a certain class are inserted (or modified) in the ontology. This leads to better rule execution performance. Furthermore, SPIN is a more promising de-facto industrial standard for the future of combining ontologies and rules, as it builds upon the widespread use of SPARQL.

III. SEMANTIC REASONING APPROACH FOR HEALTH-RELATED PROBLEM DETECTION

This section presents the proposed Semantic System for Health-related Problem detection with the aim of recognizing the activities of people with dementia through the use of different sensors and producing new knowledge by offering new services to end users of the system such as doctors, health professionals and patients. As shown in Figure 2, the raw data are collected by users (i.e., the patients with dementia) using various wearable sensors and smart home sensors. Afterwards, raw data are modeled on Resource Description Framework (RDF) ontologies and stored in the Knowledge Database (GraphDB) for the purpose of creating the System Knowledge Base. Then, the semantic analysis, which will be presented in the next section (Spin Rule Engine, Ontology and Rule reasoner, etc.), processes and interprets the data, enriching the Knowledge Base of the system.

A. IoT Data Collection

The data are collected from sensors installed in smart homes as well as from wearables. The sensors track the activities of the users unobtrusively. In that way, participants were able to perform their daily tasks without any external interference. Thus, acquiring data that correspond to everyday tasks without introducing bias in the behavior of the user due to the presence of sensors, our framework would simulate the everyday tasks of the participant at the highest possible level.

The data collected through the available sensors can be classified into three distinct categories. These categories are:

1. Activity of the patient/participant
2. Sleeping behavior
3. Cardiac activity

As far as the activity of the user is concerned, the available sensors can provide measurements listed in Table 2.

Table 2 - Activity Metrics

Metric	Unit
Steps	Amount of steps
Calories	Amount of calories
Distance Covered	Amount of meters covered
Elevation	Amount of meters in height climb
Sedentary Active State	Minutes
Light Active State	Minutes
Fairly Active State	Minutes
Very Active State	Minutes

Regarding the sleep pattern of the user, the data acquired provide records and deliver information listed in Table 3.

Table 3 - Sleep Metrics

Metric	Unit
Duration of sleep	Minutes
Efficiency of sleep	Percentage
Asleep Stage	Minutes
Time to Fall Asleep	Minutes
Awake Stage	Minutes in stage
Light Stage	Minutes in stage
Deep Stage	Minutes in stage
REM Stage	Minutes in stage
Awake Stage Count	Count of stage in session
Light Stage Count	Count of stage in session
Deep Stage Count	Count of stage in session
REM Stage Count	Count of stage in session

Regarding the cardiac activity of the user, the data collected from the available sensors provide the metrics shown in Table 4.

Table 4 - Cardiac Activity Metrics

Metric	Unit
Heart rate	Beats Per Minute
Fat Burn Zone	Minutes in this Zone
Cardio Zone	Minutes in this Zone
Peak Zone	Minutes in this Zone
Out of Range Zone	Minutes in this Zone

The sensor data are communicated on our databases through APIs, in order to be collected and processed properly. In that way, the data are grouped together and can be further processed in order to have sanity checks run on them, assuring the validity of their values before proceeding into the next steps. This procedure ensures that the data process pipeline produces reliable results in each step.

B. Knowledge Ingestion

Having acquired all the necessary raw information from all the available sensors, it is imperative to translate the raw data into meaningful representations. This procedure is called Knowledge Ingestion (KI) and is of crucial importance because it links the data through meaningful relationships, that are human understandable. These representations will act as a knowledge base, upon which more sophisticated and complicated relations will be built, thus providing semantic data logical relationships as well as acting as a logic base for further reasoning.

The translation process from raw information to linked data representations, i.e., representations that provide "meaning", is realized through the creation of RDF (Resource Description Framework) statements. An RDF statement is composed of three elements, often called a triplet. Through the use of a well-defined syntax, the relation between two different data segments is determined, via the use of a subject, a predicate and an object. The subject and the object elements represent the two different data segments that are going to be linked together, while the predicate element defines robustly the relationship between the two. Thus, each RDF Statement that is created, by using the proper subject, predicate and object, supplies the Knowledge Base (KB) with the correct relationship between the provided data.

The creation of multiple RDF statements that characterize the relationships between the data, leads to an extensive graph that describes all the possible relations between all the available data. Each data segment is represented by a node, and the relationship between two data segments (i.e., nodes) is represented by a node-linking line. An example of an RDF statement that exists in our Knowledge Base is shown in the block of code below.

```
<rdf:Description
rdf:about="http://www.semanticweb.org/ITI/ontologies/2021/2/CARL
#Sleep_2020-05-237"><awake_minutes
xmlns="http://www.semanticweb.org/ITI/ontologies/2021/2/CARL#"
rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">9
</awake_minutes>
</rdf:Description>
```

The aforementioned example illustrates the way an RDF statement is structured in the Knowledge Base. This representation, written in an XML format, is robustly structured and well-defined, for each of the elements of the triplet that constitute an RDF statement. It can be seen from the input "Sleep_2020_05-237" that the participant has been awake for 9 minutes. The relationship between the subject and the object, that is the integer number 9, is defined by the predicate "awake_minutes", that signifies the relationship between these two data segments.

The corresponding graph of the aforementioned statement is shown in Figure 1 below.

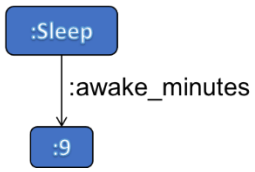


Figure 1 - Graph of an RDF statement in our Knowledge Base

C. Ontology and Knowledge Base

The proposed approach is built on top of emerging Semantic Web technologies. We started with the definition of system ontology for representing different elements of a healthcare system. The goal of the ontology creation is the semantic visualization of all concepts related to activity recognition in the healthcare system, as well as the ability to act as a semantic information integration model derived from the system's sensors. A common practice in the development of ontologies is the reuse of existing models, so we relied on already developed and valid ontologies for developing a part of the supporting ontology. The following are an overview of the existing entities used:

- **Dem@Care** [12]: An ontology to represent experimental protocols of diagnostic support and dementia diagnosis in a controlled environment.
- **SSN (Semantic Sensor Network)** [14]: Contains the ontology SOSA (Sensor, Observation, Sampler and Actuator). These ontologies describe semantic sensors, actuators, sampling and their actions. It is a W3C recommendation and OGC application.

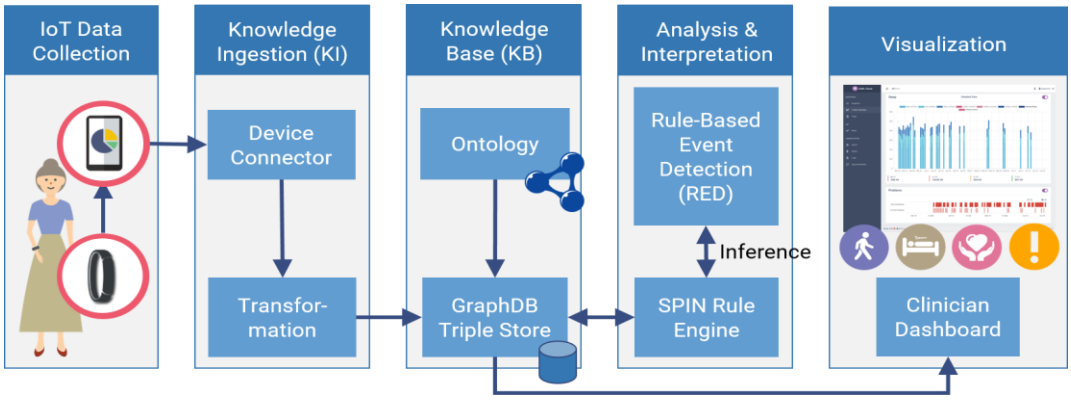


Figure 2 - Architecture of the proposed system

- **SmartHome** [15]: This ontology is an extension of SSN ontology and focuses on the representation of spatial and time aspects of entities included in spaces with devices belonging to the smart home category.

The system's ontology (Figure 3) is expressed in OWL 2 (W3C, 2012), which is a representation language commonly used in the semantic community for entity development shows the hierarchy of entity classes and the hierarchy of its properties. Object attributes are relationships that link classes together, and data attributes link classes to simple values (such as integers, alphanumeric, dates, etc.).

The main classes of ontology are Device, Event, HealthProblem, Person, and Profile. The Device represents the devices of the system. Event is a parent class for different Event-related classes. It has two subclasses Activity and Measurement. Activity contains the information of activities. Measurement includes instances, which represent information of measurements (Calories, Distance, Floor, HeartRate, Movement, Sleep, Steps). The HealthProblem is a parent class for different Health Problem - related classes. It consists of subclasses HeartProblem, MovementProblem, MultiProblem and SleepProblem. The class Person includes instances, which represents the type of Person of the system (Doctor, Patient). Finally, the class Profile includes information from users' profile (Age, Gender, etc.).

After adding Semantic Web technologies to the raw data and modeling them based on the system ontology, "Semantic Data" are stored in a semantic Graph Database, which constitutes the Knowledge Base of our system. For this purpose, we have chosen GraphDB, an enterprise ready Semantic Graph Database, compliant with W3C Standards. Semantic Graph Databases (also called RDF triplestores) provide the core infrastructure for solutions where modelling agility, data integration, relationship exploration and cross-enterprise data publishing and consumption are important. Querying and reasoning are performed over stored RDF graphs with SPARQL language.

Table 5 - A Priori Rule Base of the different semantic rules that describe the modeled activities

	Variables (number)	Rule	Problem
1	Duration in minutes	Time to fall asleep in a day > 180	Insomnia
2	Count of sleep interruptions	Number of interruptions in a day > 10	Restlessness
3	Duration in minutes	Sleep total duration in a day> 480	Too much sleep
4	Duration in minutes	Sleep total duration in a day < 300	Lack of sleep
5	Duration of "Nap" state in minutes	Asleep in Naps > 100 in a day	Increased Napping
6	Occurrence of "Nap" State, Occurance of "Night Sleep" state	Asleep in Naps end time < 2 hours from Sleep start time	Nap close to bedtime
7	Time Asleep / Time in bed	Sleep Efficiency < 85	Bad Quality Sleep
8	Step count, Heart Rate measure, Duration in minutes	Steps < 50 & Heart Rate > 90 (Fat Burn Zone) for duration > 300	Stress or Pain
9	Heart Rate measure	HR < 60	Low Heart Rate
10	Step count, Heart Rate measure, Duration in minutes	Steps <1000 & Heart Rate < 80 for duration > 300	Inactivity
11	Step count, Heart Rate measure, Duration in minutes	Steps < 500 & Heart Rate < 100 for duration > 800	Lack of Movement
12	Step count	Steps < 80	Lack of Exercise

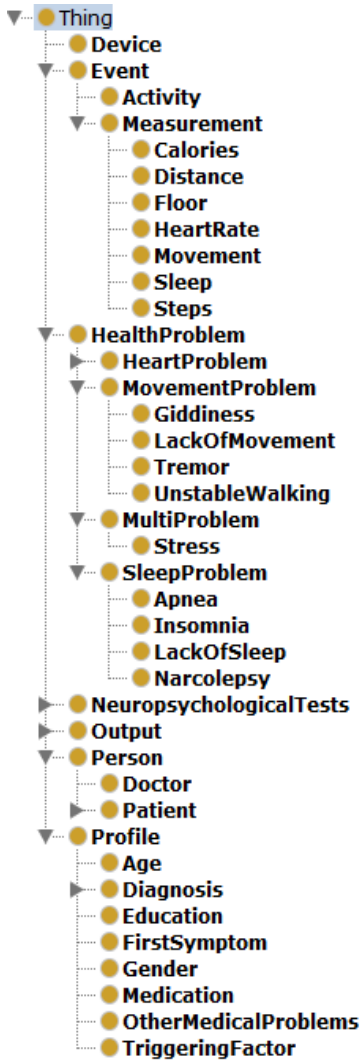


Figure 3 - Classes of the proposed ontology

D. Analysis & Interpretation with SPIN

We used the TopBraid composer [16], a tool for modeling and developing semantic data applications, to present the SPIN rules. TopBraid allows us to easily develop SPIN rules in the form of SPARQL queries, which is more readable than regular SPIN syntax. In practice, the following three code blocks act as examples in order to present in SPIN language three simple semantic rules that were applied to the system ontology.

The following code block shows the implementation of SPIN rule for Sleep problem "Lack of Sleep". Applying this rule produces the addition of a new property that represents the type of sleep problem "Lack of Sleep" in the objects of the ontology (users of a support system). If the patient's sleep duration is less than 300 minutes, then it is considered that there is a sleep problem (lack of sleep).

SPIN rule for sleep problem "Lack of Sleep".

```
CONSTRUCT {
?p owl:hasSleepProblem "Lack of Sleep "}
WHERE {
   ?p a :Person .
            ?p :duration ?d.
            FILTER (?d <300 )}
```

The following rule in SPARQL and SPIN adds new knowledge to the system Ontology. If the sleep duration of the patient with dementia is greater than 480 minutes then we conclude that there is a sleep problem (too much sleep).

SPIN rule for sleep problem "Too much Sleep".

```
CONSTRUCT {
?p owl:hasSleepProblem "Too much sleep "}
WHERE {
   ?p a :Person .
            ?p :duration ?d.
            FILTER (?d >480 )}
```

The following code block shows the implementation of the simple semantic rule "Lack of Exercise". If the steps of the participant with dementia are less than 80 and we conclude that there is a lack of exercise.

SPIN rule for problem "Lack of Exercise".

```
CONSTRUCT {
?p owl:hasProblem "Lack Of Exercise" }
WHERE {
?p a :Person .
            ?p :steps ?st1.
            FILTER (?st1<80)}
```

E. Visualization on Clinician Dashboard

The final step and ultimate purpose of our developed Semantics system, is to provide to the clinician experts the appropriate tools in order to monitor the patients and allow them, through the information that our framework provides, to make the proper decision regarding the treatment course of each patient. In order to give the clinicians the ability to have an overview of the activities or their derived problems, we created a visualization dashboard, where all the necessary analytical metrics are provided in a coherent, robust and user-friendly manner. As shown in Figure 4, the information about the patient's activities are depicted through eye-friendly tables and charts. Our implementation tried to evade complex and tiresome tables with numbers, and translated all the raw information into visualizations that are easy to understand.

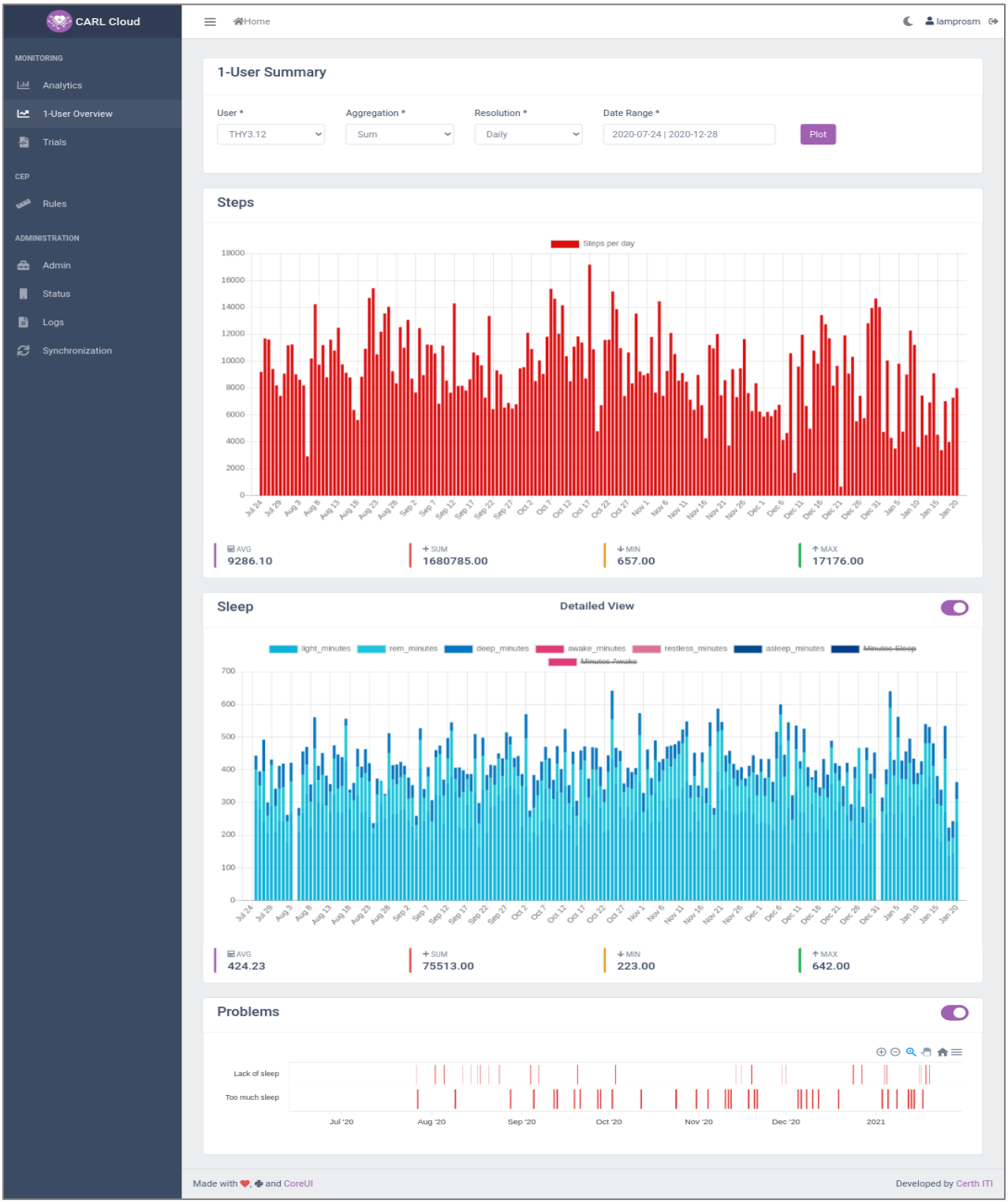


Figure 4 - Dashboard of Steps, Sleep and Problems generated

IV. USE CASE

For the evaluation of the proposed architecture we consider the following use case scenario performed on real-world data from participants recruited in the framework of the support2LIVE project. There were two patients with Dementia that were measured and evaluated. The duration of the first participant is for the full length of October 2020, while the measurements for the second participant is January 2021. The initial sensor data was modeled by using the system ontology and stored in the Knowledge Base. Then, the proposed semantic techniques were applied and in particular the semantic rules of the system were checked.

Figure 5 shows the measurement of sleep minutes of Patient 1. Likewise, Figure 6 shows the measurement of sleep minutes of Patient 2, while Figure 7 presents their measurements of steps per day. All data are processed, and results are shown, in the form of health-related problems that the patient experiences during this time.

Specifically, as shown in Figure 5 the first patient with dementia slept below the limit of 300 minutes (Table 5, Rule 4) on October 26. The results of these measurements are shown in Figure 8 with the creation and visualization of the "Lack of Sleep" problem. In addition, the patient slept above the limit set by Rule 3 (Table 5, 480 minutes) on October 3, and this resulted in the creation of the problem “Too Much Sleep”. This shows on example of how highlighting health-related problems from a dataseries would help the clinicians to efficiently and effectively detect issues that would otherwise take more time and examination.

Regarding the second patient, as shown in Figure 6, the user slept below the limit of 300 minutes (Table 5, Rule 4) on January 3 and 7. The problems generated by the system for these measurements are shown in Figure 9 with the creation and visualization of the "Lack of Sleep" problem. In addition, the patient slept above the limit set by Rule 3 (Table 5, 480 minutes) on January 2, and this resulted in the creation of the problem “Too Much Sleep”. Similarly to the previous case, our framework efficiently detects the underlying health-related problems in both activity and sleep.

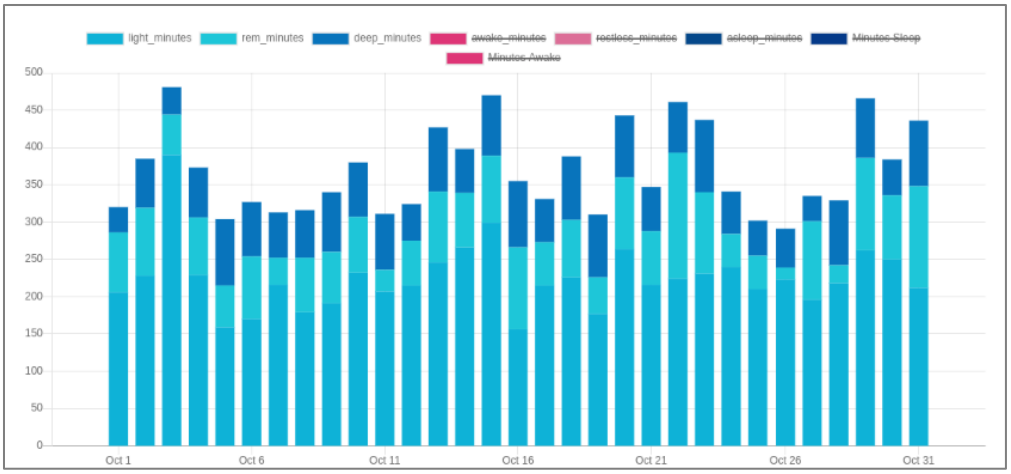


Figure 5 - Patient 1 Sleep

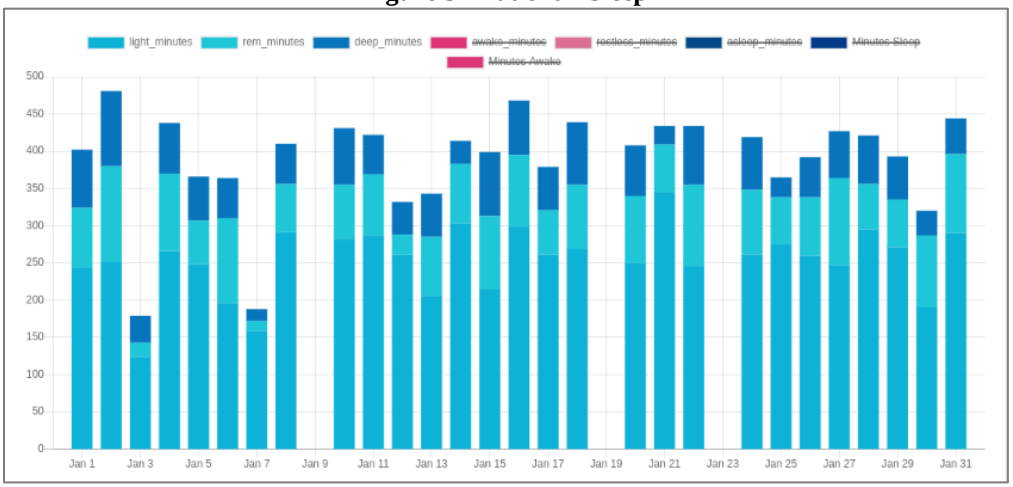


Figure 6 - Patient 2 Sleep

Finally, in Figure 7, it is observed that on January 8 and 20 the patient took a low number of steps. As a result, this measurement generates the problem (Rule 11) “Lack of Movement”, is shown in Figure 9.

This example showcases the ability of the system to highlight problems of relatively high importance to the users. Specifically, this user’s steps are generally low everyday, but detected problem occurs and emphasizes only on the days of exceptionally low activity (as shown in Figure 7 and Figure 9), something that the human eye would not that easily and efficiently detect.

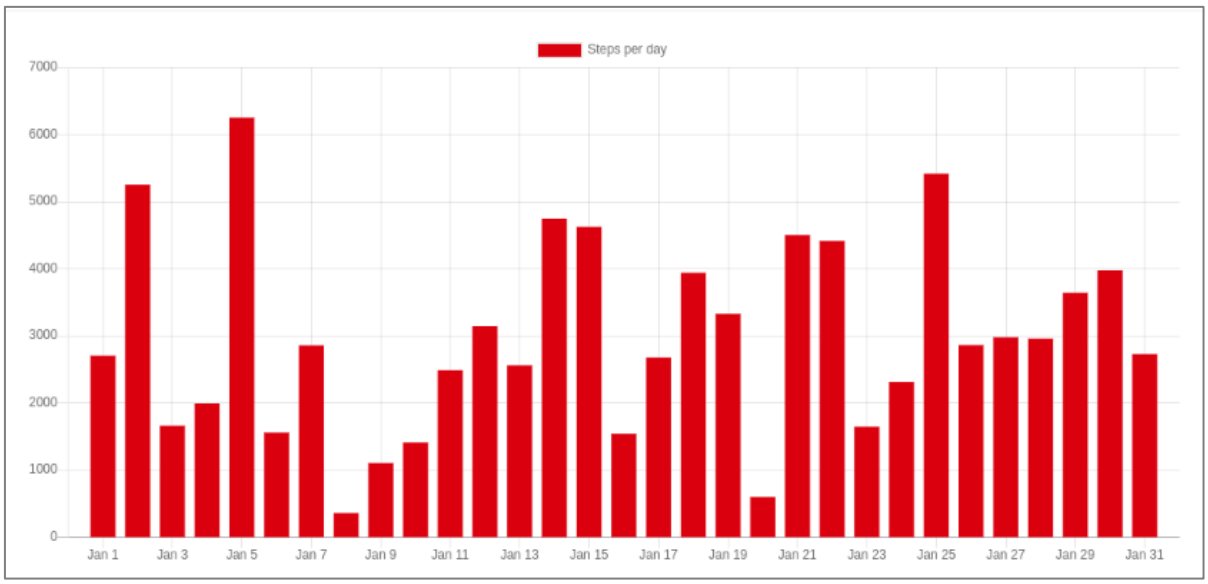


Figure 7 - Patient 2 Steps

Figure 8 - Patient 1, Problems Generated by the System

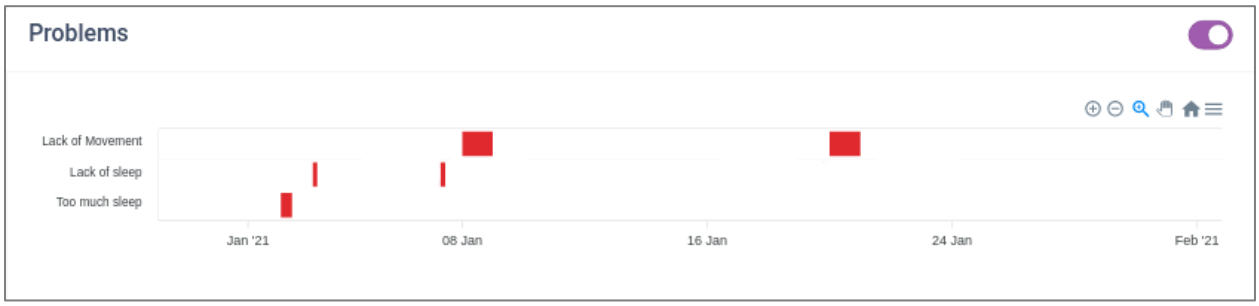


Figure 9 - Patient 2, Problems Generated by the System

V. Conclusion and Future Work

In this paper, we presented our approach towards the definition of a semantic system for Health-related Problem detection that combines ontologies and SPIN Rules. Architectures related to the proposed framework are listed, and the advantages of using the SPIN language to create semantic rules are presented. The main purpose of the proposed architecture is to generate new knowledge from the original raw data, especially recognition of healthcare problems for people with dementia. The system is validated through a proof-of-concept use case scenario where a wearable sensor gathers data from a real participant, and the framework extracts the expected Health-related Problems.

As future work, we plan to evaluate the framework in a formal clinical trial with real participants. Participants will be recruited in the spectrum of dementia, as well as healthy controls, and use the wearables for several months. The framework will be used to extract problems and clinical experts will evaluate its accuracy, usability and usefulness for the disease. In the long run, the system will be able to support decision making of the clinicians and facilitate them in adjusting non-pharmaceutical interventions.

Acknowledgment

This work was co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE, project code: T1EDK-02668.

References

[1] A. Pliatsios, T. G. Stavropoulos, D. Strantsalis, V. Kassiano, S. Nikolopoulos, I. Kompatsiaris, "Detection of Health-Related Problems of People with Dementia from Lifestyle Wearables: A Rule-Based Approach", The 12th International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED 2020, 21-25 November 2020, Valencia, Spain

[2] A. Pliatsios, C. Goumopoulos, and K. Kotis, "Interoperability in IoT: A Vital Key Factor to Create the 'Social Network' of Things," in *UBICOMM 2019*, pp. 63–69, 2019.

[3] A. I. Maarala, X. Su, and J. Riekki, "Semantic Reasoning for Context-Aware Internet of Things Applications," *IEEE Internet Things J.*, 2017.

[4] A. Pliatsios, C. Goumopoulos, and K. Kotis, "A Review on IoT Frameworks Supporting Multi-Level Interoperability. The Semantic Social Network of Things Framework," *Int. J. Adv. Internet Technol.*, vol. 13, pp. 46–64, 2020.

[5] M. Ziaeefard and R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognit.*, 2015,. 48(8), 2329-2345

[6] K. L. Skillen et al., "Ontological user modelling and semantic rule-based reasoning for personalisation of Help-On-Demand services in pervasive environments," *Futur. Gener. Comput. Syst.*, 2014, 34, 97-109.

[7] H. Knublauch, J. A. Hendler, and K. Idehen, "SPIN: SPARQL Inferencing Notation," *W3C Member Submission*, 2011.

[8] C. Fürber, & M. Hepp (2010, May). Using SPARQL and SPIN for data quality management on the semantic web. In International Conference on Business Information Systems (pp. 35-46). Springer, Berlin, Heidelberg.

[9] The support2LIVE project: https://www.ypostirizo-project.gr

[10] G. Meditskos, T. G. Stavropoulos, S. Andreadis, and I. Kompatsiaris, "KnowSense: A semantically-enabled pervasive framework to assist clinical autonomy assessment," in *CEUR Workshop Proceedings*, 2015.

[11] D. Riboni and C. Bettini, "COSAR: Hybrid Reasoning for Context-aware Activity Recognition," *Pers. Ubiquitous Comput.*, vol. 15, pp. 271–289, 2011.

[12] G. Fico *et al.*, "Co-creating with consumers and stakeholders to understand the benefit of internet of things in smart living environments for ageing well," in *IFMBE Proceedings*, 2017.

[13] T. G. Stavropoulos, G. Meditskos, S. Andreadis, and I. Kompatsiaris, "Dem@Care: Ambient sensing and intelligent decision support for the care of dementia," in *CEUR Workshop Proceedings*, 2015.

[14] F. De Backere *et al.*, "Towards a social and context-aware multi-sensor fall detection and risk assessment platform," *Comput. Biol. Med.*, 2015, 64, 307-320.

[15] M. Compton *et al.*, "The SSN ontology of the W3C semantic sensor network incubator group," *J. Web Semant.*, 2012, 17, 25-32.

[16] The TopBraid Composer Tool:
https://www.topquadrant.com/products/topbraid-composer/

A Rule-Based Framework for Object Localization, Spatial and Situational Awareness with Natural Language Feedback for the Deafblind

Vasileios Kassiano, Anastasios S. Kesidis, Thanos G. Stavropoulos,
Spiros Nikolopoulos, and Ioannis Kompatsiaris
Information Technologies Institute (ITI)
Centre for Research & Technology Hellas (CERTH)
Thessaloniki, Greece
{vaskass, akesidis, athstavr, nikolopo, ikom}@iti.gr

***Abstract*—Deafblindness is a debilitating condition that hinders communication, awareness of the surroundings and ultimately independent living. Technological advancements in image analysis and haptics promise to support patients. Yet, machine-interpretable representations, interpretation and feedback to bridge the gap between perceiving the environment and communicating this to the user, are lacking. This paper presents a rule-based framework that supports object localization, spatial and situational awareness and natural language feedback for the deafblind. The framework utilizes ontologies as an interoperable data model to represent knowledge about the environment and rules to extract object localization, spatial and situational awareness as well as to form appropriate natural language responses. The rule set is expressed in the SPARQL Inferencing Notation (SPIN), which enables simplicity and flexibility in rule definition. A dashboard web application visualizes the streaming detections perceived from the environment and allows real-time queries and responses. A proof-of-concept scenario is realized with synthetic data from a realistic environment, showing use cases in detecting surroundings, locating an object and being aware of a situation (e.g., people wearing a facemask as a COVID-19 precautionary measure). In the future, the platform will support connection to a data exchange message bus to receive detections from image analysis and communicate actions towards haptic hardware to enable testing by patients.**

Keywords-ontology; rules; natural language; SPIN; situational awareness; spatial awareness; localization; deafblindness.

I. INTRODUCTION

Communication with and between users with deafblindness is constrained by the debilitating nature of this disability, ranging from congenital to acquired deafblindness, including worsening eyesight, hearing or both over time, and, ultimately, symptoms of ageing as well. Technology promises to facilitate such communication problems with advancements in image analysis [2] [3] and haptic technology [4] [5]. Yet, applications that integrate a machine-interpretable understanding of the users' surroundings and give comprehensive feedback to them are limited.

This paper presents a framework that provides object localization, spatial and situational awareness for the deafblind. The framework is based on ontologies to interoperable represent context (i.e., the environment, surroundings and situations) and rules to interpret, transform and formulate appropriate responses to user queries. The SPIN rule-engine provides flexible and extendable rule sets for finding an object, realizing surroundings in relevance to the user and becoming aware of the situation (e.g., people wearing facemasks as a precautionary measure against COVID-19). While the previous study in [1] presented the data model (ontology) and spatial awareness rules, this study extends the framework with object localization and situational awareness as well as a dashboard for visualizing the system's streaming detections, performing user-defined queries in real-time and getting responses. Proof-of-concept use cases are realized using this dashboard, demonstrating suitable responses inspired by the SUITCEYES project [6], to improve the Quality of Life of the deafblind using interactive technology. In SUITCEYES, the platform will additionally support connection to a message bus to receive real detections from cameras and image analytics as well as haptics, in order to provide feedback.

The rest of the paper is structured as follows: In Section II, related work regarding natural language (NL) and spatial awareness is presented. In Section III, the architecture of our system is introduced, including data collection and storing processes, as well as inference methods incorporated. The methods used for Spatial Information to Natural Language, for Facemask Feedback and for Object Localization can also be found in this section. In Section IV, our experiments and results produced by using synthetic data are shown, inspired by the data received from the SUITCEYES platform. Finally, in Section V conclusion and motivation for future work can be found.

II. RELATED WORK

An increasing number of Internet of Things (IoT) applications are used for healthcare purposes due to their ability to support living with various ailments, such as ageing and dementia [7] [8] [9]. These types of applications acquire knowledge from multiple sources and from continuous and heterogeneous data flows [10] [11]. Semantic technologies provide comprehensive tools and methods for representing knowledge and using inference to produce new knowledge from data. IoT environments are increasingly found in home healthcare technologies in actions that create better living conditions for the elderly, through the use of IoT

technologies, such as Active and Healthy Ageing (AHA) and Home Ambient Assisted Living (AAL).

Furthermore, in the case of deafblind people, the simple and accurate representation of their surrounding environment is one of the most basic needs for their quality of life. This can be achieved by providing the nature of their surroundings in natural language. Some of the most relevant work for language processing with ontologies are [12] [13].

KnowSense [7] is an activity monitoring system for elder people with dementia, deployed in controlled and diffuse environments. Semantic Web technologies, such as OWL 2, are widely used in KnowSense to display sensor and specific application observations, as well as to implement solutions for identifying activities and problems in everyday life activities (Instrumental activities of daily living, IADLs) with the aim of clinical evaluation of various stages of dementia. Description Logic (DL) reasoning for activity detection and SPARQL questions are used to extract clinical problems.

ACTIVAGE [8] is a large-scale pilot project, which aims to develop Smart Living solutions that strengthen active and healthy aging. The ACTIVAGE IoT Ecosystem Suite (AIOTES) project consists of a set of techniques, tools and methodologies (rule-based reasoning, interoperable ontologies, etc.) that increases semantic interoperability at different levels between heterogeneous IoT platforms. The approach uses different mechanisms of reasoning that can improve the understanding of patients' heterogeneous data and help generate new knowledge by providing services to end users.

Dem@Care [9] is a system based on heterogeneous sensors that provides support for independent living for elder people with dementia or similar health problems. This approach incorporates a heterogeneous set of detection methods and technologies, including video, audio, in addition to normal, environmental, and other measurements. Semantic technologies (e.g., rules-based reasoning) are used to process and analyze sensor data according to user requirements. This leads to feedback and decision support, which is communicated to end users through appropriately designed user interfaces. The support includes various clinical scenarios, both short (trials in hospital settings) and long term (daily living at work), for independent living.

Furthermore, semantic mechanisms are created and applied that process patients' original data and use it to generate new knowledge by offering new services for the benefit of end users such as caregivers, health professionals and patients.

In [14], a system for healthcare in Smart Home environments is developed which considers social relationship-based contexts to provide a fully personalized healthcare service.

An ontology-based sensor selection for real-world wearable activity recognition is presented in [15], in which the use of ontologies is proposed to thoroughly describe the wearable sensors available for the activity recognition process. This enables the semantic selection of sensors to support a continuity of recognition.

In [16], an extended version of a linguistic ontology is presented that works particularly with space. Language regarding space, spatial relationships and actions in space is covered and an ontological structure that relates such expressions with ontology classes is developed. Finally, examples of the ontology's results based on natural language examples are presented.

In [17], a project in which ontologies are part of the reasoning process used for information management and for the presentation of information is presented. Both accessing and presenting information are mediated via natural language and the ontologies are coupled with the lexicon used in the natural language component.

An approach to transform natural language sentences into SPARQL is proposed in [18], with the use of background knowledge from ontologies and lexicons. The results of this approach show that the diagnosis process and the data search for a broad range of users is improved.

In [19], an evaluation is made on how efficient the SPARQL query language is and the SPARQL Inferencing Notation (SPIN) when utilized to identify data quality problems in Semantic Web data automatically, and within the Semantic Web technology stack

In the case of deafblind patients, the simple and accurate representation of their environment is one of the most important needs. In [15], [16] and [17] even though forms of natural language processing through ontologies are proposed, they do not involve healthcare or wearable sensors. For the healthcare related work [4] [5] [6] [8], [13] and [14], the natural language presentation of information component is absent.

Meanwhile, as the COVID-19 pandemic crises developed worldwide, in [20], the need for precautionary measures to consider and adapt to people with disabilities is emphasized, with tailored recommendations. That includes people with deafblindness, who find it harder to maintain social distancing and assess potential threats in the surroundings.

The framework proposed in this paper, combines existing approaches to represent data and execute rules using semantic technologies for healthcare, but extends them with: 1) a representation of the surrounding environment for deafblind patients; 2) a set of rules to extract knowledge for object localization, situational and spatial awareness and provide responses in natural language; 3) a dashboard to visualize inputs and outputs and integrate with image and haptic technology to enable pilots. Specifically, existing data models and rule sets are greatly extended to support the more varied toolbox of rules expressed in SPARQL and SPIN notation, which are considered to be the state-of-the-art notation when it comes to semantic data management. Particularly, inspired from the pandemic, situational awareness includes rules to respond to whether people are wearing a mask or not.

III. THE PROPOSED FRAMEWORK

A. Requirements and System Specification

The requirements for the framework were derived from the SUITCEYES project, aiming to improve the Quality of

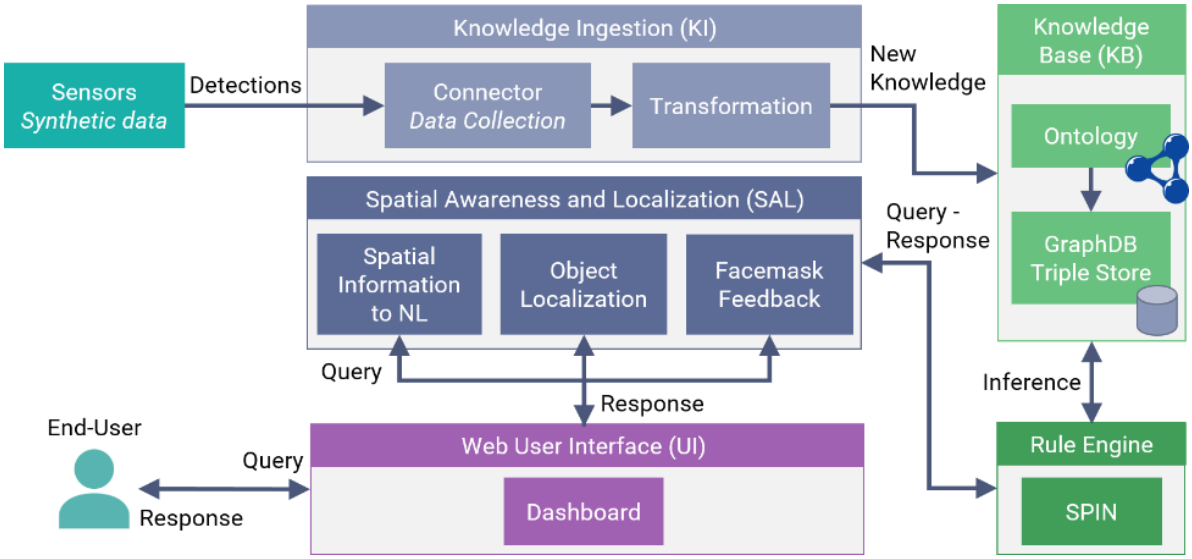


Figure 1: System Architecture

Life of people with deafblindness. As such, the project provided both functional and non-functional requirements such as the ability to exchange data coming from sensors, cameras and image analytics and respond back to haptics in a universal JSON format in real-time.

The "detections" from input devices contain information such as which object/person was detected, when the detection took place, what sensor made the detection and spatial information like where in the user's field of view is the object/person, e.g., left, right, in front of, etc. Finally, the measured distance is reported, if the sensor has this capability. Additionally, the system should receive and respond to queries, like "Where is my book?" and respond in natural language, comprehensively, in a manner that the user would understand (through translation to haptic devices).

As such, there is a need for a "brain" of the system, a toolbox of methods containing the appropriate rules for extracting knowledge from the incoming data, a knowledge/data base for storage and the methods for handling the overall procedure and the input/output.

The input that our system receives come from the end-user who sends queries like "Where is my PC?" to the dashboard. After examining the knowledge contained in the ontology, they receive the appropriate response, either in a natural language message or comprehensive directions and pointers, which can later be translated in haptic devices.

B. Architecture

According to the aforementioned requirements, the proposed framework was designed using a modular, extensible and service-oriented architecture shown on Figure 1. It consists of the following components:

- The sensors/synthetic data which provide information in the form of detections to be stored in the ontology.
- The Knowledge Ingestion (KI) component of our framework is responsible for receiving the incoming data from the sensors/synthetic data source and make the appropriate transformations (e.g., differential population) so that the data are ready to be uploaded to the ontology stored in GraphDB.
- The component that stores all data and rules, and executes the queries is the Knowledge Base (KB). In our KB the ontology with all the new knowledge is stored, in the GraphDB Triple Store database.
- The Spatial Awareness and Localization (SAL) component contains all methods that were developed to extract information from the ontology and to answer the user queries. These methods are: a) the Spatial Information to NL, which answer simple queries like: "Where is my laptop?", b) Object Localization (OL), which answer queries like "Locate my book" and c) Facemask Feedback, which answer queries like: "Is the person next to me wearing a facemask?".
- A Dashboard implemented as a Web User Interface (UI) application, which is responsible for visualizing the latest streaming detections as they arrive in the KB, and enables users to input queries and receive answers. The dashboard is intended for researchers and integrators to visualize results before integrating with their technology for data input (detections) or output (haptic messages).
- Finally, we use the SPIN rule engine to execute rules automatically, each time new data gets uploaded to our KB.

C. The Ontology – Data Model

As a primary data model, the framework utilizes the SUITCEYES ontology [4], presented in [1] and available online [24]. The ontology integrates heterogeneous, multimodal input from different sensors in a formal and semantically enriched basis, thus, user context-related

information, that can provide enhanced situational awareness and augment the user's navigation and communication capabilities. The ontology already includes representation of objects and activities from the Dem@Care ontology [9], sensors and sensor observations from the SOSA/SSN ontologies [21], persons and social associations from the Friend-Of-A-Friend (FOAF) specification [22]. In addition, core topological concepts of a building from SEAS (Smart Energy Aware Systems) [23] was integrated, which is a schema for describing the core topological concepts of a building, such as buildings, building spaces and rooms.

In this work, concepts and relations are extended to support new functionality. Some of the basic classes of the SUITCEYES ontology represent objects, spaces and people that can be detected as raw data form sensors, such as cameras, or processed data, through a visual analysis component. For example, some of the objects that can be found in the ontology are: computer, laptop, alarm clock, mug, table, chair and other everyday objects, while some of the spaces include: bedroom, bathroom, living room and other spaces that can usually be found in a home environment. We extended these concepts further from the ontology presented in [1], by adding more classes to the ontology that refer to objects, spaces and rooms that could be useful to any deafblind user. We also added a Boolean data property named "facemask" to the Person entity.

D. Knowledge Ingestion (KI) Component

The data containing the detections from sensors or from synthetic data arrive in our KI component. Each message that arrives represents an image frame that contains information about the scene, people and objects detected. An example with the kind of data that the KI component receives can be seen below:


```
{
 "image": {
            "target": [
                      {
                       "height": 176,
                       "top": 96,
                       "type": "person_unknown",
                       "width": 115,
                       "left": 407,
                       "confidence": 0.725,
                       "distance": 1322.01
                        },
                       {
                        "height": 134,
                        "top": 329,
                        "type": "computer",
                        "width": 195,
                        "left": 123,
                        "confidence": 0.597,
                        "distance": 1822.01
                        }
                                ],
  "timestamp": "2020-02-19T08:18:27.649",
  "scene_score": 1.0,
  "width": 640,
  "scene_type": "office",
  "height": 480,
  "name": "19_02_20_08_18_27_649204"
 }
```


In this example, we can see that an unknown person and a computer are detected in an office. We can also acquire information about the time of this detection, the distance from the source and various information about the position of the bounding boxes of the detected entities (height, top, width, and left) in pixels.

The KI component receives these messages and using a simple function prepares them to be stored in our ontology by creating a simple INSERT SPARQL query. This query contains all the meaningful information about the entities and scenes detected.

Before executing an INSERT query, another step of data transformation is performed. We developed an efficient approach to store the incoming data to the ontology. We call this approach "differential population" of the ontology, which means that instead of populating the ontology with every detection data received, a method is applied that checks if the incoming detection data is already included in the ontology. If true, then only the timestamp of the existing detection instance is changed, otherwise we add the new detection to the ontology. This algorithm of the differential population procedure is presented in Figure 2.

By using this technique, we limit the volume of data that are inserted in the ontology by only applying a simple check each time we receive a detection. This increases efficiency, considering that in a home environment the same objects could be detected in the same place multiple times (e.g., a TV almost never changes place in a home) and that many sensor systems continuously send data via their sensors.

E. Spatial Awareness and Localization (SAL) component

The SAL component is responsible for performing all the processes using semantic technologies to answer the user queries. It contains 3 main methods that answer different type of queries. These methods are the following:

Algorithm: **Differential Population**
Input: D (Detection type object)

for each D_i stored in ontology ***do***:
{
If D.detects_object == D_i.detects object and
D.spatialContext == D_i.spatialContext then
D_i.timestap = D.timestamp
}
End

Figure 2: Differential Population Algorithm

```
CONSTRUCT{
    ?detection sot:producesOutput ?output.
    ?output sot:refersToDetection ?detection.
    ?output a sot:Output.
    ?output sot:hasTextualDescription ?description .
}
WHERE{
    BIND (sospin:Function_SelectLatesObjectDetection() AS ?detection) .
    ?detection a sot:Detection .
    ?detection sot:detectsObject ?object .
    ?object a ?object_type.
    FILTER (?object_type = sot:Laptop).
    ?object rdfs:label ?objectName .
    ?object sot:hasSpatialContect ?spatialContext.
    ?spatialContext a ?spatialContextType .
    ?spatialContextType rdfs:SubClassOF sot:Spatialcontext .
    FILTER ((?spatialContextType - sot:RightSpatialCOntext) || (?spatialContextType =
sot:LeftSPatialContext)).
    BIND (sospin:Function_LeftRightContext(?spatialContext) AS ?leftright_annotation).
    FILTER ((?spatialContextType - sot:RightSpatialCOntext) || (?spatialContextType =
sot:LeftSPatialContext)).
    BIND (sospin:Function_LeftRightContext(?spatialContext) AS ?leftright_annotation).
    BIND (sospin:Function_CloseFarContext(?spatialContext) AS ?closefar_annotation).
    FILTER ((?spatialContextType - sot:CloseSpatialCOntext) || (?spatialContextType =
sot:farSpatialContext)).
    BIND (sospin:Function_CloseFarContext(?spatialContext) AS ?closefar_annotation).
    BIND (BNODE() AS ?output) .
    BIND (CONCAT(?objectName, "is on your", ?leftright_annotation, " side.", ?closefar_annotation, "to you.
") AS ?description).
}
```

Figure 3: SPARQL/SPIN rule for query #1

1) Spatial Information to Natural Language:

In Table 1, we present a list of indicative queries that are used in the ontology to provide the user with a natural language output regarding spatial information of their surroundings. In these queries, variables are used to cover a broad number of objects and spatial contexts. This kind of inference is achieved by using a set of ontological rules, written in SPARQL/SPIN notation, that run on top of the ontology, whenever a specific query is triggered by the user. Within the context of this scenario, a list of indicative queries were created that can dynamically change on specific aspects, i.e., to cover different entities of interest, different spatial relations (with respect to the distance of position left/right), etc.

Table 1: List of rules and their output

#	Query	Nat. Lang. Output
1	Where is my <object >?	Your <object > is on your < right/left spatial context> side, <close/far spatial context> to/from you.
2	How many <objects> are on my <right/left spatial context> side?	<# counted> objects, an <object1> and an <object 2> are on your < right/left spatial context > side.
3	Which <objects> are <spatial context> to/from me?	An <object1>, <object2> and object3> are located <spatial context> to/from you.

The variables in the above queries are given a specific value, depending on what the user wants to ask. For example, the first query can be transformed in natural language to: “Where is my laptop” and its output can be: “Your laptop is on your right side, close to you”. The implementation of this query is presented as an ontological rule written in SPARQL/SPIN syntax in Figure 3, using synthetic data, which we created, that include various objects and spatial contexts. Most of our ontological rules use SPARQL CONSTRUCT and DELETE/INSERT commands, in order to create new triples in the ontology and thus enrich the knowledge stored in the schema.

In Figure 4, the implementation of the #3 query is presented, which in natural language translates to “Which objects are close to me?” which using our synthetic data produces the output: “A laptop, a TV and a chair are located close to you”. For this query we have skipped the construct rule, which is the same as the #1 query.

2) Facemask Feedback

During the COVID-19 pandemic, it would be extremely useful for a deafblind user to ask if the person/people next to them wear a facemask. For this reason, we developed the ontology and implemented a SPIN query to perform this task.

From the ontology side, we added a Boolean data property to the Person entity called “facemask” which has value true if the person is detected wearing a facemask and

```
WHERE{
    BIND (sospin:Function_SelectLatesObjectDetection() AS ?detection) .
    ?detection a sot:Detection .
    ?detection sot:detectsObject ?object .
    ?detection sot:detectsObject ?object2 .
    ?detection sot:detectsObject ?object3 .
    ?object a ?object_type.
    ?object2 a ?object_type.
    ?object3 a ?object_type.
    ?object rdfs:label ?objectName .
    ?object2 rdfs:label ?objectName2 .
    ?object3 rdfs:label ?objectName3 .
    ?object sot:hasSpatialContect ?spatialContext.
    ?object2 sot:hasSpatialContect ?spatialContext2.
    ?object3 sot:hasSpatialContect ?spatialContext3.
    ?spatialContext a ?spatialContextType .
    ?spatialContext2 a ?spatialContextType2 .
    ?spatialContext3 a ?spatialContextType3 .
    ?spatialContextType rdfs:SubClassOF sot:Spatialcontext .
    ?spatialContextType2 rdfs:SubClassOF sot:Spatialcontext .
    ?spatialContextType3 rdfs:SubClassOF sot:Spatialcontext .
    FILTER ((?spatialContextType - sot:CloseSpatialCOntext) &&
        (?spatialContextType2 = sot:CloseSPatialContext)  &&
        (?spatialContextType3 = sot:CloseSPatialContext)).
    BIND (sospin:Function_CloseFarContext(?spatialContext) AS ?closefar_annotation).
    BIND (BNODE() AS ?output) .
    BIND (CONCAT(?objectName, ", ", ?objectName2, "and a ", ?objectName3, ", are ", ?closefar_annotation, "to you. ")
        AS ?description).
}
```

Figure 4: SPARQL/SPIN rule for query #3

false if not. In Figure 5, we present the facemask data property of the Person class.

If the query returns a person wearing a facemask, we offer this information in the form of an output file with the field "facemask: yes/no". The query fired to return the information of whether the person next to the user wears a facemask is presented in Figure 6.

As its name suggests, the variable "?facemask" will contain the Boolean value of whether the person nearest the user wears a facemask or not. We order the results in descending order regarding to the timestamp, as we want to take into account only the most recent detection and in ascending order regarding the distance, as the query refers to the person nearest to the user. The limit value on the last line of the query shown in Figure 6 can be changed to include more people wearing a facemask around the user.

3) *Spatial Awareness and Localization (SAL)*

For the purpose of locating an object, it would be useful if our ontology could answer queries of a more continuous nature. For example, in the previous iteration, if the user asked "Where is my laptop", they would get a single answer containing the laptop's location if it was found, or the information that the laptop was not detected. Then the user would have to repeat the query again and again, until the object is finally detected, i.e., when the object comes in the field of view of the camera that is attached on the user.

For this reason, we implemented SPIN/SPARQL queries that support SAL. These queries are continuously executed and the process ends only when the object that the user is searching for is found and it is very close to them, or when

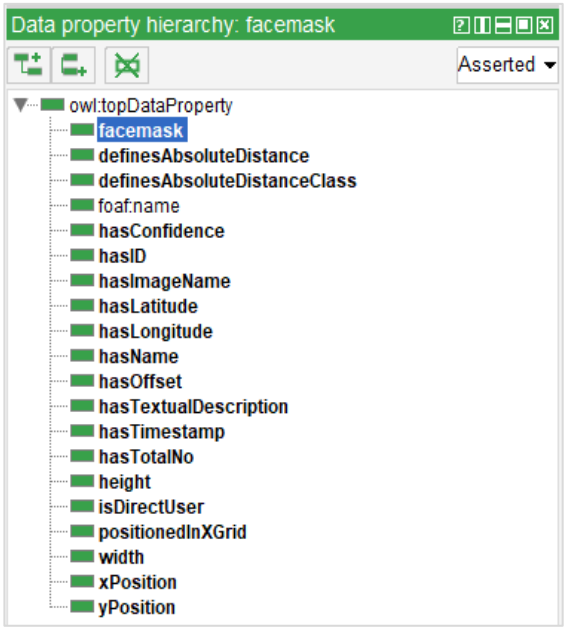


Figure 5: Facemask Data Property

```
SELECT ?facemask
WHERE {
    ?detection a sot:Detection .
    ?detection sot:hasTimestamp ?timestamp .
    ?detection sot:providedBy ?sensor .
    ?detection sot:detectsSemanticSpace ?scene .
    ?detection sot:detectsPerson ?person .
    ?scene rdf:type sot:SemanticSpace .
    ?person sot:hasSpatialContext ?sc.
    ?person sot:facemask ?facemask.
    ?sc sot:definesAbsoluteDistance ?distance.
    }
    ORDER BY desc(?timestamp) asc(?distance)
    LIMIT 1
```

Figure 7: Facemask Feedback Query

the user manually stops the search. The process of SAL is described below.

When the user tries to find a specific object, i.e., when a "Locate my <object>" query is received, the system performs the SAL procedure. During this procedure, the SPIN rules that try to look for this object in the database are fired consequently, every second that passes. The information that is received from the KB is then stored in an output file to be shared with the appropriate component of the system, which gives directions to the user. The directions given are based on an offset value that our query returns. This value that refers to pixels, ranges between -200 to 200 in a 400x400 picture that a sensor captures. For example, if the offset value is 100, it means that the object is on the right side of the user and the directions should guide the user to the right. The process stops when the distance variable that our query returns is less than 0.3 meters which is the value we assume that the user can reach the desired object with their hands. This threshold could be changed throughout the development of the system to correspond to specific objects, by experimenting with different setups and users.

Algorithm: **Spatial Awareness and Localization**
Input: User query in JSON format
Output: JSON file with object information
uq = user query
distance = 1;

```
While (distance > 0.3 || object != found)
{
        query response = perform query(uq);
        if (query_response has object detected)
                object = found;
        distance = query_response.distance;
        output = prepare_response_Json(query
        response);
}
End
```

Functions:
perform_query: sends the query to be executed in our GraphDB instance
prepare_response_JSON: takes the dataset that the query returns and creates the response JSON file

Figure 6: SAL Algorithm

```
SELECT ?detection ?timestamp ?slabel ?elabel ?leftright ?dis
WHERE {
    ?detection a sot:Detection .
    ?detection sot:hasTimestamp ?timestamp .
    ?detection sot:providedBy ?sensor .
    ?detection sot:detectsSemanticSpace ?scene .
    ?detection sot:detectsObject ?entity .
    ?entity rdfs:label ?elabel.
    ?scene rdfs:label ?slabel.
    ?entity a sot:Object.
    ?scene rdf:type sot:SemanticSpace .
    ?entity sot:hasXGridSpace ?x.
    ?x sot:positionedInXGrid ?leftright.
    ?entity sot:hasSpatialContext ?sc.
    ?sc sot:definesAbsoluteDistance ?distance.
    }
ORDER BY desc(?timestamp)
LIMIT 1
```

Figure 8: SAL query for Object Entity search

In SAL, the information that is sent to the system contains the following four kinds of data:

- Timestamp of the detection that the object was last detected.
- If the object was found in the latest detection, the detection field contains the field "detection" which holds a Boolean variable with value "True" in the response file. If the object is not found in the last detection, the value is "False".
- Object's orientation from the user in the form of pixel offset in the detection image. For example, if the offset is negative 100, the object is located on the left of the side of the field of view of the user.
- Object's distance from the user in meters.

The form of the response output can be seen below:

```
{
  "timestamp": "2021-05-17T08:42:00",
  "detection": True,

  "offset": -100,
  "distance": 2.0
}
```

The algorithm that performs the SAL procedure is presented in Figure 7.

The rules that are fired during SAL are presented below. Once again, we use SPIN/SPARQL notation to implement these rules and the result set is transformed into an output file with the appropriate information with the use of a simple JAVA function that we created. In Figure 8, we present the query fired when the user searches for an entity.

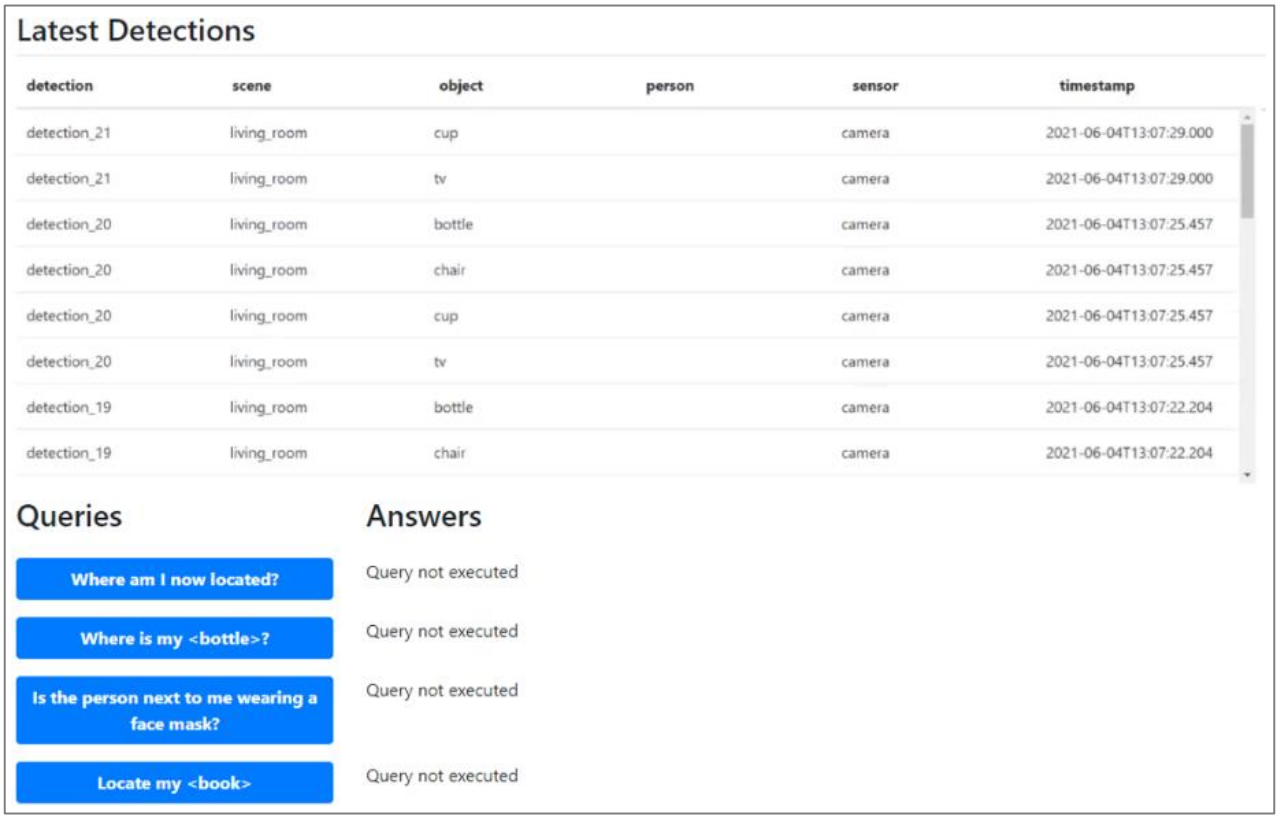


Figure 9: Dashboard showing latest detections as they stream into the platform and allowing queries and responses

The variable "?timestamp" contains the timestamp information that is needed for the output file. The "?elabel" variable contains the string value that describes the object entity. For example, the label of a Cup object of the ontology is "cup". This field helps us group some objects in the same category, e.g., when the user searches for a mug we assume that a cup would also suffice. This is why we have put the label "cup" in the Mug objects of the ontology as well, so when a query with the rdfs:label "mug" is fired, objects of both Mug and Cup classes would be returned.

The "?leftright" variable contains the value of the sot:positionedInXGrid data property. This value is measured as pixels in an integer, where 0 is the center of the picture, negative values mean left and positive right from the center. The "?leftright" variable gives us the offset value of the response.

The "?distance" variable, as the name suggests, contains the distance of the object to the user. The distance is measured in millimeters, which are then translated to meters, using a JAVA function.

Finally, we use the "ORDER BY desc (?timestamp)", to order them by descending timestamp, and the LIMIT 1 commands to get only the most recent detection of the searched entity. In the event of the entity not found, by using once again a JAVA function, we return the Boolean value "False" to be contained in the output, otherwise the value is "True".

The notion behind this procedure is that when an object is detected and it is very close to the user (e.g., <0.3m), the user can reach and find the object by touch. This is why our loop also checks the distance value of the detection, to decide whether to continue or not.

F. Web User Interface (UI) – Dashboard

For visualization purposes, along with making easier the process of testing and experimenting with our system, we developed a dashboard, which contains the most recent information about our KB and visualizes the queries that are sent by the user and the answers that they receive from our system. The dashboard is presented in Figure 9.

The dashboard contains two main components:

- The Latest Detections table, which visualizes the most important information about the latest detections stored in the ontology such as detection id, scene, object and person detected, sensor that made the detection and timestamp of the detection. The detections table is updated live with the latest detections received from the sensors or, in our case, the synthetic data.
- The Query and Answer component, which in this iteration of the dashboard contains 4 of the main queries that a user can send to our system. These queries are: 1) Where am I now located? 2) Where

is my <bottle>? 3) Is the person next to me wearing a facemask? and 4) Locate my <book>.

Note that the entities in <> are parameters, i.e., can be changed to include other entities.

IV. PROOF OF CONCEPT USE CASES

Utilizing the system and the ability to visualize data and input queries on the dashboard, we performed a set of use case scenarios to validate its correctness. The scenarios are inspired from the SUITCEYES project. Based on its requirements, we constructed a set of synthetic data in the form of input messages streaming in the platform and with a realistic sequence of objects and people entering, moving and leaving the scene, as well as scenes changing, emulating a person moving in a room with a wearable camera. The use case scenarios are the following:

i) A developer/integrator wants to test the KB platform for receiving detection data from their component connected to the message bus

ii) A developer/integrator wishes to receive natural language output from user queries to integrate into their haptic system

iii) A deafblind user wishes to know whether the person or people around them wears a facemask

iv) A deafblind user wishes to find an object and navigate to it

Each of the following subsections presents one scenario performed through the system.

A. Incoming Detection Data stored Visualization

For the purpose of visualizing the process of receiving data, we developed a simple dashboard interface which shows in real time the data stored in the ontology.

As we can see in the snapshot presented in Figure 9 from the synthetic data we received in the last three detections the scene is a living room, the objects detected are a cup, a tv, a bottle and a chair, no people were detected and the sensor was a camera. Note that this snapshot will be used as our basis for the state of the ontology during our queries experimentation.

B. Spatial Information to Natural Language Experiments and Results

We present the output of the SPARQL/SPIN rule shown in Figure 3 by using the Protégé software [25]. In Figure 10, we present the object and description returned by the query. The description will be used as the output to the user.

In Figure 11 and Figure 12, we present the objects returned from the query “Which objects are close to me” with their spatial context, and the output that the user will get as textualDescription.

We also use the queries and answer component of the dashboard we developed to receive queries and provide answers in real time. In Figure 13, we can see the queries sent and the answers received from our framework, regarding the state of the ontology described in Figure 9.

By receiving the expected results regarding our synthetic data, we validated the correctness of our queries.

We also used the same synthetic data to test the differential population procedure. In the example below, we have stored in our KB two similar detections: The first detection contains a laptop in the living room a detection with the same information, arrives from the SUITCEYES system with a more recent timestamp. This information, if stored twice in the ontology would be duplicated with the exception of the different times. In Figure 14, we present what the outcome would be without using our differential population method, i.e., the storing of 2 similar detections.

In Figure 15, the outcome of the same incoming detection is presented, but with using the differential population procedure. Here, only the first detection is stored, with its timestamp field changed to match that of the second, incoming detection.

This absence of almost identical detections could save a lot of space in real applications that continuously receive and store data.

C. Facemask Feedback experiments and results

In this section we present the process that is executed when a user asks if the person next to them wears a facemask. For this reason, we created synthetic data that contain people

object	description
Laptop	"Your Laptop is on your right side, close to you."^^<http://www.w3.org/2001/XMLSchema#

Figure 10: Query output of "Where is my laptop"

obj	spatialContext	obj2	spatialContext2	obj3	spatialContext3
Laptop	Close	Chair	Close	TV	Close

Figure 11: Objects and their Spatial Context from query#3

queryOutput	textualDescription
QueryOutput	"A Laptop, a Chair and a TV are close to you"^^<http://www.w3.org/2001/XMLSchema#string>

Figure 12: Query output of "Which objects are close to me"

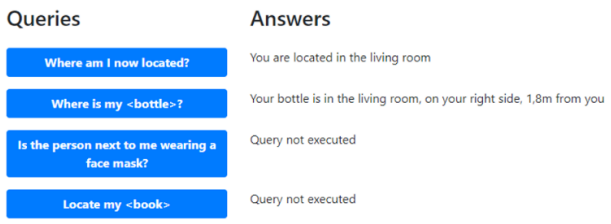


Figure 13: Query and Answer component of the dashboard

detection	obj_detected	where	timestamp
Detection_1	Laptop	Living_Room	"2020-08-25T13:18:30Z"^^<http://www.w3.org/2001/XMLSchema#dateTimeStamp>
Detection_2	Laptop	Living_Room	"2020-08-25T13:20:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTimeStamp>

Figure 14: Detections saved without the differential population procedure

detection	obj_detected	where	timestamp
Detection_1	Laptop	Living_Room	"2020-08-25T13:20:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTimeStamp>

Figure 15: Detections saved with the differential population procedure

```
sd:detection_5
  rdf:type sot:Detection ;
  sot:detectsPerson soa:john ;
  sot:detectsPerson soa:unknown_person_1 ;
  sot:detectsSemanticSpace soa:kitchen_1 ;
  sot:hasConfidence "0.77"^^xsd:float ;
  sot:hasTimestamp "2021-05-20T09:32:00.233"^^xsd:dateTime
  sot:providedBy soa:camera_1 ;
.
sd:john
  rdf:type sot:KnownPerson ;
  rdf:type sot:Person ;
  rdf:type foaf:Person ;
  sot:locatedInSemanticSpace soa:kitchen_1 ;
  sot:definesAbsoluteDistance 1500.00;
  sot:facemask "True";
  rdfs:label "John" ;
.
sd:unknown_person_1
  rdf:type sot:Person ;
  rdf:type sot:UnknownPerson ;
  rdf:type foaf:Person ;
  sot:definesAbsoluteDistance 2500.00;
  sot:locatedInSemanticSpace soa:kitchen_1;
  sot:facemask "False";
.
```

Figure 16: Latest detection received for facemask query

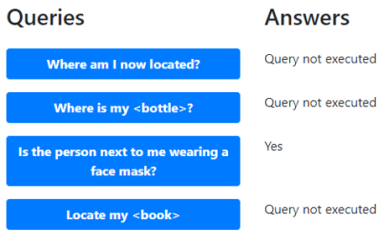


Figure 17: SAL dashboard Facemask Feedback Response

```
sd:detection_1
  rdf:type sot:Detection ;
  sot:detectsObject soa:laptop_1 ;
  sot:detectsObject soa:book_1 ;
  sot:detectsObject soa:cup_1 ;
  sot:detectsSemanticSpace soa:living_room_1 ;
  sot:hasConfidence "0.95"^^xsd:float ;
  sot:hasTimestamp "2021-06-15T09:40:00"^^xsd:dateTi
  sot:providedBy soa:camera_1 ;
.
sd:livingroom_1
  rdf:type seas:LivingRoom ;
  rdfs:label "living room" ;
.
sd:laptop_1
  rdf:type sot:Object ;
  rdf:type sot:Laptop ;
  sot:locatedInSemanticSpace soa:living_room_1 ;
  sot:definesAbsoluteDistance 3400.00;
.
sd:book_1
  rdf:type sot:Object ;
  rdf:type sot:Book ;
  sot:locatedInSemanticSpace soa:living_room_1 ;
  sot:definesAbsoluteDistance 5600.00;
.
sd:cup_1
  rdf:type sot:Object ;
  rdf:type sot:Cup ;
  sot:locatedInSemanticSpace soa:living_room_1 ;
  sot:definesAbsoluteDistance 4500.00;
.
```

Figure 18: Synthetic Database snapshot

that are either wearing or not wearing a facemask, in various distances from the user. The file that contains this synthetic data is presented in Figure 16 below. We have named this ontology file "synthetic data", so we use the preface "sd".

We notice that in this detection there are two people detected in the kitchen, John and an unknown person. John wears a facemask but the unknown person is not. Now, we assume that the user sends us a query asking whether the person closer to him wears a facemask. This query has been presented in Figure 8. Using a simple function, we produce a simple response file with the field "facemask" and the value "yes". We present the Query and Answer component of our dashboard in Figure 17.

We can modify the query and the output according to the user's needs. For example, if the user would like to know about the facemask status of all the people in his field of view, we can simply omit the LIMIT part of the query presented in Figure 6. Then by using another simple function we produce the message output shown below.

```
{
  "person": "John",
  "facemask": "yes",
  "person": "Unknown_Person_1",
  "facemask": "no"
}
```

D. Spatial Awareness and Localization queries experiments and results

For the purpose of experimenting with SAL queries we used a synthetic database that we know would accurately represent a series of detections in the SUITCEYES platform. Our synthetic data contains rooms and objects in those rooms that a user could choose to search for. In Figure 18, we can see part of this synthetic database. This part describes a detection which contains a living room as the recognized scene and a laptop, book and cup as the recognized objects in various distances.

Now, we are going to present the process that is being executed when a SAL query is asked, i.e., "Locate my <book>, by showing the queries and responses we receive from the GraphDB database. For this experiment, we receive a query from a user searching for his book. Our framework initiates SAL process by entering the loop and executing the query presented in Figure 19, on our GraphDB stored ontology. Note that this query is the query presented in Figure 8, with only the "?entity" variable changed to contain an object of the class Book.

The execution of the aforementioned query in combination with the function for the message output produces as a result the file shown in below:

```
{
  "timestamp": "2021-06-15T09:40:00",
  "detection": True,
  "offset": -13,
  "distance": 5.6
}
```

We present the Queries and Answer component of our dashboard where we verified the correctness of our responses in Figure 20.

As we can notice, our method reports back the data that are stored in our ontology database. As the distance value is not lower than 0.3, which we assume is the reaching distance for an object, after 1 second of the initial query execution, we perform again the same query. As our framework is designed to always receive new detections, we receive a detection in which the user has gone closer to the object after they receive directions based on our initial information. The new state of the detections in our ontology can be found in Figure 21.

In this latest detection we notice that the distance of the book object that the user is searching is decreased to 2.5 meters. We also notice that the cup object that was detected previously has not appeared in this detection, meaning that it is not in the field of view of the user. The output that we produce for this iteration of the OL procedure is presented below:

```
{
  "timestamp": "2021-06-15T09:40:10",
  "detection": True,
  "offset": 33,
  "distance": 2.5
}
```

In this output we can see again that we produce the correct information for the book object. We can also notice that the offset has changed into a positive value that means that the object is now on the right side of the field of view of the user. Once again, we verify our results using the dashboard in Figure 22.

The same procedure is repeated until the distance value becomes less than 0.3 meters. Then, we assume that the user can use his hands to find the object that they look for.

If in any time during this procedure the detection value becomes “False”, we just re-execute the query until we receive a detection which contains the object that the user is searching for.

Our SAL component is designed to work in combination with some module that gives some kind of directions to the user so that they can approach the object that they look for.

V. CONCLUSION AND FUTURE WORK

In this paper, we present a framework to support deafblind people using a rule-based toolbox of methods for object localization, spatial and situational awareness. The framework utilizes ontologies for uniform context data representation and the SPIN/SPARQL notation to extract knowledge and construct natural language responses. A dashboard visualizes inputs and outputs streaming over a universal message bus message exchange so as to enable developers integrate advanced detection methods from wearable cameras and user interfaces on haptic devices. The framework is validated through use cases scenarios showcased on its dashboard.

```
SELECT ?detection ?timestamp ?slabel ?elabel ?leftright ?distance
WHERE {
    ?detection a sot:Detection .
    ?detection sot:hasTimestamp ?timestamp .
    ?detection sot:providedBy ?sensor .
    ?detection sot:detectsSemanticSpace ?scene .
    ?detection sot:detectsObject ?entity .
    ?entity rdfs:label ?elabel.
    ?scene rdfs:label ?slabel.
    ?entity a sot:Book.
    ?scene rdf:type sot:SemanticSpace .
    ?entity sot:hasXGridSpace ?x.
    ?x sot:positionedInXGrid ?leftright.
    ?entity sot:hasSpatialContext ?sc.
    ?sc sot:definesAbsoluteDistance ?distance.
    }
ORDER BY desc(?timestamp)
LIMIT 1
```

Figure 19: Book SAL query

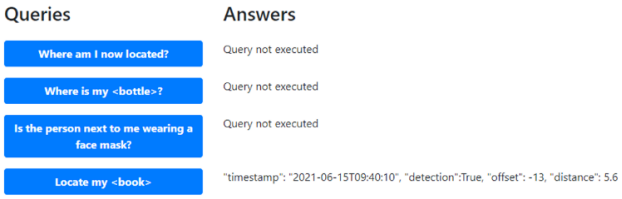


Figure 20: SAL response on dashboard

```
sd:detection_4
  rdf:type sot:Detection ;
  sot:detectsObject soa:laptop_1 ;
  sot:detectsObject soa:book_1 ;
  sot:detectsSemanticSpace soa:living_room_1 ;
  sot:hasConfidence "0.95"^^xsd:float ;
  sot:hasTimestamp "2021-06-15T09:40:10"^^xsd:dateTime
  sot:providedBy soa:camera_1 ;
.
sd:book_1
  rdf:type sot:Object ;
  rdf:type sot:Book ;
  sot:locatedInSemanticSpace soa:living_room_1 ;
  sot:definesAbsoluteDistance 2500.00;
.
```

Figure 21: Latest detection received

Figure 22: SAL updated response on dashboard

Such systems can greatly improve the quality of life of people with deafblindness especially regarding spatial awareness, but also situational awareness in the pandemic era, when people with disabilities need tailored measures and means to recognize their surroundings. Future work, includes integration of the framework with real data providers, such as sensors, cameras and the visual analysis. Moreover, a pilot phase includes deafblind users interacting with the system to

validate its acceptance and usefulness and to incrementally optimize the system for them.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreements No. 78880814 SUITCEYES.

REFERENCES

[1] V. Kassiano, T. Stavropoulos, S. Nikolopoulos, I. Kompatsiaris, M. Riga, "Spatial Awareness for the Deafblind in Natural Language Presentation using SPIN Rules: A Use Case in the SUITCEYES Platform," The 12th International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED 2020, 21-25 November 2020, Valencia, Spain

[2] N. Olson, Nasrine, et al. "Sensor Technology, Gamification, HapticInterfaces in an Assistive Wearable," Journal on Technology and Persons with Disabilities 7 (2019): 79-87.

[3] O. Korn et al. "Empowering persons with deafblindness: Designing an intelligent assistive wearable in the SUITCEYES project," Proceedings of the 11th PErvasive technologies related to assistive environments conference. 2018.

[4] S. Darányi et al. "Static and Dynamic Haptograms to Communicate Semantic Content: Towards Enabling Face-to-Face Communication for People with Deafblindness." SEMAPRO 2019, The Thirteenth International Conference on Advances in Semantic Processing. International Academy, Research and Industry Association (IARIA), 2019.

[5] O. Korn, J. Gay, R. Gouveia, L. Buchweitz, A. S. Schulz, and M. Umfahrer. "Tactile navigation with checkpoints as progress indicators? only when walking longer straight paths." In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1-8. 2020.

[6] The SUITCEYES project: https://suitceyes.eu/, Last access date 12/12/2021

[7] G. Meditskos, T. G. Stavropoulos, S. Andreadis, and I. Kompatsiaris, "KnowSense: A semantically-enabled pervasive framework to assist clinical autonomy assessment," in CEUR Worshop Proceedings, 2015.

[8] G. Fico et al., "Co-creating with consumers and stakeholders to understand the benefit of internet of things in smart living environments for ageing well: The approach adopted in the madrid deployment site of the activage large scale pilot," in IFMBE Proceedings, 2017, doi: 10.1007/978-981-10-5122-7_272.

[9] T. G. Stavropoulos, G. Meditskos, S. Andreadis, and I. Kompatsiaris, "Dem@Care: Ambient sensing and intelligent decision support for the care of dementia," in CEUR Workshop Proceedings, 2015.

[10] I. Maarala, X. Su, and J. Riekki, "Semantic Reasoning for Context-Aware Internet of Things Applications," IEEE Internet Things J., 2017, doi: 10.1109/JIOT.2016.2587060.

[11] A. Pliatsios, C. Goumopoulos, and K. Kotis, "Interoperability in IoT: A Vital Key Factor to Create the 'Social Network' of Things," in The Thirteenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2019), 2019, pp. 63–69.

[12] J. A. Bateman et al. "A linguistic ontology of space for natural language processing," Artificial Intelligence 174.14 (2010): 1027-1071.

[13] D. Estival, C. Nowak, and A. Zschorn. "Towards ontology-based natural language processing," Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology. 2004.

[14] L. Haesung and J. Kwon. "Ontology model-based situation and socially-aware health care service in a smart home environment," International Journal of Smart Home 7.5 (2013): 239-250.

[15] C. Villalonga et al. "MIMU-Wear: Ontology-based sensor selection for real-world wearable activity recognition." Neurocomputing 250 (2017): 76-100.

[16] M. Sander et al. "Ontology-based translation of natural language queries to SPARQL." 2014 AAAI fall symposium series. 2014.

[17] G. Meditskos, S. Dasiopoulou, and I. Kompatsiaris, "MetaQ: A Knowledge-driven Framework for Contextaware Activity Recognition Combining SPARQL and OWL 2 Activity Patterns," Pervasive and Mobile Computing, vol 25, pp. 104-124, 2016.

[18] G. Meditskos and I. Kompatsiaris, "iKnow: Ontologydriven Situational Awareness for the Recognition of Activities of Daily Living," Pervasive and Mobile Computing, Vol 40, pp. 17-41, 2017.

[19] C. Fürber and M. Hepp. "Using SPARQL and SPIN for data quality management on the semantic web." International Conference on Business Information Systems. Springer, Berlin, Heidelberg, 2010.

[20] M. Mörchen, H. Kapoor, and S. Varughese. "Disability and COVID-19." Community Eye Health 33, no. 109 (2020): 10.

[21] K. Janowicz, A. Haller, S. JD Cox, D. Le Phuoc, and M. Lefrançois. "SOSA: A lightweight ontology for sensors, observations, samples, and actuators." Journal of Web Semantics 56 (2019): 1-10.

[22] M. M. A. Al-Mukhtar and A. T. A. Al-Assafy. "The implementation of foaf ontology for an academic social network," International Journal of Science, Engineering and Computer Technology, vol. 4.1, p. 10, 2014.

[23] M. Lefrançois, J. Kalaoja, T. Ghariani, and A. Zimmermann. "The SEAS Knowledge Model." PhD diss., ITEA2 12004 Smart Energy Aware Systems, 2017.

[24] The SUITCEYES Ontology:

https://mklab.iti.gr/results/the-suitceyes-ontology/, Last access date 12/12/2021

[25] The Protégé Software: https://protege.stanford.edu/, Last access date 12/12/2021

Cognitive and Behavioral Data for Decision Tree-based Diagnosis of Attention-Deficit/Hyperactivity Disorder

Anderson Martins Silva
Graduate Program
in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 71953371@mackenzista.com.br

Luiz Renato Rodrigues Carreiro
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: luizrenato.carreiro@mackenzie.br

Mayara Miyahara Moraes Silva
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 72054298@mackenzista.com.br

Maria Cristina Triguero Veloz Teixeira
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: mcris@mackenzie.br

Leandro Augusto da Silva
Graduate Program
in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: leandroaugusto.silva@mackenzie.br

***Abstract*—Attention-Deficit/Hyperactivity Disorder (ADHD) presents in children and adolescents as a persistent pattern of inattention, hyperactivity, and impulsivity that interferes with their development. Computational studies on ADHD focus on measures of brain activity of the participants and a few use standardized cognitive tests or behavioral inventories to assess objective indicators for diagnosis. The paper presents a computational proposal in which the combination of two artificial intelligence methods is used to aid the identification of diagnostic indicators for ADHD. The proposal is to combine a neural network of self-organizing maps to group factors from standardized tests and inventories, and a decision tree to classify the most relevant factors. The study included 127 children and adolescents from 6 to 16 years old, 48 with ADHD diagnosis and 79 without ADHD (control group). The most relevant result of the study was the strong contribution of the Child and Adolescent Behavior Inventory results in the diagnosis of the disorder with great performance in prediction when compared to real data and reliability by Kappa statistics.**

***Keywords*—Self-Organizing Maps (SOM); Decision Tree; Attention Deficit/Hyperactivity Disorder (ADHD).**

I. Introduction

This work is complementary to and based on the published article BRAININFO 2021 [1] and according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition - DSM-5 [2]. Attention-Deficit/Hyperactivity Disorder (ADHD) is a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. The disorder is characterized by inattention involving, for example, difficulty sustaining attention in tasks or playing activities, a state in which the mind seems elsewhere, even in the absence of any obvious distraction, difficulty to follow through with instructions and failing to finish schoolwork, often forgetful in daily activities, chores, or duties in the workplace, losing things, expressing excessive activity or restlessness, and inability to wait one's turn, always in ways that are excessive for one's age or developmental level. ADHD has its initial expressions in childhood and usually persists into adulthood, resulting in impairments in social, academic, and occupational functioning.

The diagnosis of ADHD is clinical, based on the individual's history and expression of symptoms. Because this diagnosis is often based on reports of symptom severity and because these symptoms are also part of other clinical conditions, the diagnostic difficulty is present in the daily lives of the interdisciplinary teams responsible for the evaluation process [3] [4]. Because of the complexity of the diagnostic evaluation, the American Association of Pediatrics recommends the use of an algorithm, both for evaluation and treatment of children and adolescents with ADHD [5]. To support clinical decision making, neuropsychological, behavioral, and adaptive functioning assessment procedures have often been used in conjunction with neurological assessments [6]. Considering the social importance involved in properly issuing a correct diagnosis of ADHD, in both children and adolescents, studies must be proposed that discuss which are the best indicators of clinical-neurological, neuropsychological, and behavioral-adaptive diagnostic evaluation when children and adolescents present with complaints of inattention and hyperactivity. Furthermore, for appropriate assessments and interventions to be implemented, differential criteria are needed to correctly characterize and identify attention-deficit/hyperactivity among children and adolescents. Comprehensive assessments in this regard allow a better understanding of the complexity of each case for appropriate guidance, design of the therapeutic intervention, and evaluation of the need for educational and

emotional support for patients and families [6].

Computational studies can help professionals in diagnostic assessments, especially using machine learning algorithms, Kam et al. [8] used an artificial intelligence algorithm called decision tree for screening ADHD, by monitoring the school activities of 153 children using 3-axial actigraph and obtained results consistent with previous studies. In turn, Lee et al. [9] analyzed the classification of ADHD in children through brain activity measurements. In their work, they used a neural network algorithm called self-organizing maps allowing categorizing characteristics of children with and without clinical indicators of ADHD.

Unlike previous proposals presented in the literature, this work aims to combine two artificial intelligence techniques. In the first step, standardized test results are grouped by means of Self-Organizing Maps (SOM) and, in a second step, the groups with a high level of overlap are analyzed using a decision tree algorithm. This helps discover which attribute is discriminative in the diagnosis of children and adolescents with suspected ADHD.

Besides Section I, that aims to contextualize the work and present the objective, the work is organized into six parts. Section II presents the theoretical framework and justification of the study. Section III presents the proposed use of two artificial intelligence algorithms to aid in the diagnosis. In Section IV, the procedures for developing the study are described, including the computational development with the application of two artificial intelligence techniques. In Section V, the contribution of standardized cognitive tests or behavioral inventories is described, as well as the proposal to solve the diagnostic doubt within the self-organizing maps and then the classification by the decision tree for understanding the characteristics of the diagnosis of the disorder. Finally, in Section VI, we present the conclusion and recommendations for further studies.

II. Related Work

A. Elements of Attention Deficit/Hyperactivity Disorder (ADHD)

ADHD is part of the group of neurodevelopmental disorders beginning in childhood, but a substantial proportion of children with ADHD remain relatively impaired into adulthood [9]. From a cognitive-behavioral point of view, it is characterized by deficits in several cognitive functions, such as attention, especially selective, sustained, alternating, and divided attention, deficits in inhibitory control, processing speed, organization, ability to inhibit distracting information, deficits in cognitive flexibility, hyperactivity behaviors, restlessness, and impulsivity. ADHD affects 5.29% of the world's child population. Of this population, 30% up to 70% maintain symptoms into adulthood [10] [11]. According to DSM-5 [1], ADHD can be classified according to the predominance of symptomatic axes as predominantly inattentive, predominantly hyperactive-impulsive, or combined presentations.

Behavioral patterns are important in the diagnosis of ADHD. Here are some difficulties related by parents regarding the children: listening, obeying, following routine rules, often postponing and forgetting daily activities, following direct instructions, regulating feelings of frustration, exacerbation of motor activity, maybe impulsive in changing activities before they are completed, having difficulty waiting their turn, may have impairments in social relationships. These behaviors may contribute to high-stress in family or school environments [12].

The inequality of symptom axes within the predominantly inattentive subgroup compromises its validity when compared to the combined subgroup. However, individuals considered only inattentive, but with a subclinical level of hyperactivity-impulsivity symptoms (4 or 5 symptoms), have their classification without intensity of the combined ADHD subtype. However, there is still little evidence on the qualitative difference between the subtypes presented, even when inattentive classification is established to individuals with three or fewer hyperactivity-impulsivity symptoms [13]. At different stages of life, ADHD can show itself at about 2:1 in the case of children and 1.6:1 in adults. This shift in prevalence from children to adults hypothetically occurs because children and adolescents can create methods that suppress the disorder as they develop, making them more functional, and hiding some difficulties and symptoms [9].

According to the new classification of the A.P.A [2], the highlights are the changes in the various forms of symptoms, trying to contextualize the criteria diagnosed throughout the individual's life; change in the age of onset of symptoms, from 7 to 12 years of age; change of the term "subtype" for "current presentation"; and removal of autistic spectrum disorders as excluding factors in the diagnosis.

Children who are usually in the preschool phase may present high levels of motor activity, attention deficit, and poor inhibitory control that are behavioral manifestations of ADHD. However, in clinical cases, they are more significant and result in considerable impairment, with high accident rates and poor school performance that can persist into school age in 60 to 80 percent of cases [14]. Approximately 70% of school-age children experience worsening in school activities and impairment in family life and relationships with other children. Generally, inattention symptoms relative to hyperactivity symptoms decelerate with age, i.e., decline slowly between the passage of the child, adolescent, and adult age stages with greater persistence [15]. About 33% of children with ADHD in adulthood no longer have the symptoms, as opposed to the rest who continue to have the disorder or episodes that result in loss. Over a prolonged period, adults with ADHD experience worsening academic and work performances, as well as increased traffic offenses, motor vehicle accidents, and

sexual behavior with high risk and the concomitant onset of various psychiatric illnesses, such as anxiety disorder, mood disorders, and substance abuse [6].

Genetic factors and environmental conditions contribute greatly to a complex etiology of ADHD with neurobiological bases established by research. For example, study conducted by Williams [16] points to unusual patterns in the central nervous system of people with ADHD [17]. According to Carreiro et al. [6], the study of neuropsychological endophenotypes is motivated by the etiological complexity and clinical inhomogeneity of ADHD. Endophenotypes, sometimes called "intermediate phenotypes", are inherited and gauged traits that can be found in the pathway linking a genotype to complex neuropsychiatric disorders. However, individuals with ADHD may or may not display different forms of patterns following a complex cognitive profile and multiple variations, such as loss or low attentional focus, causing a lack of flexibility, difficulty dealing with distractors, difficulty with self-regulation, behavioral impulsivity, lack of motor coordination, and shuffling of mental information due to difficulty with organization. The signs presented show that the child may have an impairment in the intellectual development in different areas of the brain and are shown by playful interactions, by observation, and by standardization of instruments, becoming fundamental in measuring the complaints of inattention and hyperactivity [18].

Behavioral patterns are important in diagnosing ADHD, parents report that children have difficulty listening for inattention when someone speaks directly, obey, and follow rules and routines, often postpone and forget daily activities, have difficulty following direct instructions, do not accept frustration, have an exacerbation of motor activity, may show some form of impulsivity in an activity before it is completed, are unable to wait, have impaired social relationships, have high-stress in the family or school environments. Observation of several people can increase the accuracy of the diagnosis, enabling in cases of comorbidities and inconsistent symptoms, the differentiated possibility of diagnosis. This creates the need to report in a standardized way, to avoid bias and the possibility of building a representative behavioral profile for evaluation. Currently, the psychology literature has several instruments that are applied to parents and teachers, to extract as much information as possible from the two environments of children [18]. These instruments, which are seen in this work as attributes, have a protocol-based direction to assess ADHD complaints, and the attributes are analyzed by cognitive assessment parameters used, such as the Cancel Attention Test, Trail Making Test, Continuous Performance Test, Wisconsin Letter Test, Wechsler Abbreviated Intelligence Scale and Wechsler Intelligence Scale for Children, as well as the behavior inventories for children and adolescents (CBCL/6-18) and for teachers (TRF/6-18).

Given the importance of collecting various pieces of information in cognitive neuropsychology and behavior analysis, the treatment and multivariate analysis of the data can help us obtain relevant information in understanding ADHD complaints, and the artificial intelligence techniques used become key elements in diagnostic discrimination.

B. Self-Organizing Maps (SOM)

According to Merényi et al. [19], a SOM network provides clustering and visual representation of data in low dimensions. This technique preserves the topological structure of the data in a lattice of neurons. The grid can be defined as a rectangular or hexagonal grid, as in Figure 1, usually two-dimensional, in an ordered manner such that the most similar neurons are grouped with neurons that are close in the grid, and the opposite is true for less similar neurons that are far apart in the grid, providing a topological view of the data. All neurons in the grid must undergo exposure to different realizations of the input dataset to ensure that the self-organization process matures. The algorithm then proceeds to choose synaptic weights initially randomly with small values. Once the grid has been initialized, we have the presence of three essential processes used to construct the self-organizing map.

With the need to understand the characteristics of the combined attributes and facing data with non-linear distribution, it was chosen in this work an unsupervised model, developed by Kohonen in 1982, called SOM, being especially suitable for data assimilation because it has visualization properties such as highlighting [20], but the dataset allowed the use of principal component analysis or cluster analysis of the data. The goal of unsupervised methods is to identify clusters in unlabeled sets of data vectors sharing similarities. This helps to build a cognitive model that realizes the interrelationship of the data [21]. As proposed by Kohonen [22], capturing the features of an input data set, with a nonlinear distribution, is effected by building a complex neural network around a one- or two-dimensional grid of neurons. The ordering of the input data is structured by weight vectors of neurons called prototypes, and is inspired by neurobiology. They were summarized by Kohonen [22] and Kubat [23] as follows:

1) Competition: The generalist model can serve to describe the phenomena of the initial data in a way that collectively orders a complex system, which shows representative statistical features, even with the fully disordered input space [23].

However, the SOM model is automatically associated with the nodes of the rectangular or hexagonal grid, Figure 1, usually two-dimensional, in an ordered fashion and in a way to which the most similar models join closer nodes in the grid. The opposite being true for less similar models that are equidistant, thus this process enables a topological view of the data [24].

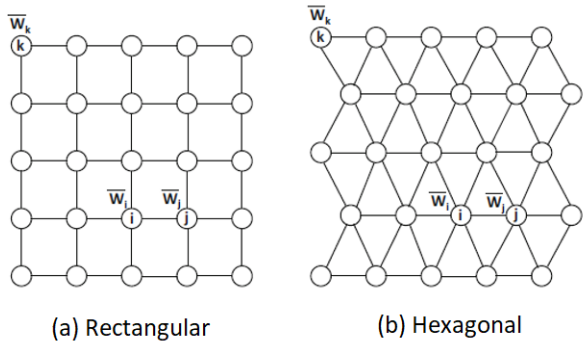


Figure 1. Topological map - rectangular and hexagonal grids ([25, p.451])

The synaptic weight vector is calculated for each *j* neuron of the grid with the same dimension as the input dataset through the inner product between the synaptic weight vector and the input data vector, this function being the basis for choosing the winning neuron. The maximization of this function has mathematical equivalence with the minimization of the Euclidean distance between the synaptic weight and input data vectors.

$$X = [x_1, x_2, ..., x_m]^T \tag{1}$$

X is the input vector of the space *m* transposed.

$$W_j = [w_{j1}, w_{j2}, ..., w_{jm}]^T, \ j = 1, 2, ..., l \tag{2}$$

W_j is the synaptic weight vector of each neuron in the grid. In the competition step, according to Kubat [23], an input vector is randomly selected along with the synaptic weight vector of the neuron *j*, that by making the inner product of the two vectors, being $W_j^T X$ for *j* = 1, 2,..., *l*. Transposing the input vector it is possible to make the selection of the largest product. The topological neighborhood of the excited neurons is centered and since maximizing the inner product of the vectors X and W_j is minimizing their Euclidean distance, by creating an index *i*(X) to identify the neuron that best relates to the input vector X, one can define *i*(X) by:

$$i(X) = \arg\ \min_j \left\| X - W_j \right\|, \ j = 1, 2, ..., l \tag{3}$$

i(X) is the index that summarizes the competitive process between neurons.

This process summarizes the competition between neurons. Equation (3) shows the *i*(X) is the goal of this process because in this step the identity of the neuron *i* is important and the neuron that satisfies the condition is called the winning neuron for the input vector X.

2) Cooperation: The cooperation process starts when the winning neuron is updated around a topological neighborhood of the nearest neurons, being similar around a radius *r*. However, it is necessary to define the topological neighborhood so that only adjacent neurons are updated while having a way that the neighborhood decays smoothly with lateral distance [23].

Given $h_{j,i}$ the topological neighborhood centered around the winning neuron *i* that contains a set of excited neurons, one neuron of this set being represented by *j* and $d_{j,i}$ the lateral distance of the excited neuron *j* by the winning neuron *i*, then one can assume that the topological neighborhood. Equation (4) is a unimodal function of the distance $d_{j,i}$, provided it satisfies the symmetry conditions with respect to the maximum point with $d_{j,i}$=0, and the amplitude decreases with increasing lateral distance $d_{j,i}$ [23].

The basis for cooperation between neighboring neurons is provided by the winner neuron that shows the spatial location of the topological neighborhood of neurons neighboring the winner $h_{j,i(X)}$:

$$h_{j,i(X)} = \exp\left(-\frac{d^2{}_{j,i}}{2\sigma^2}\right) \tag{4}$$

$d_{j,i}$ is side distance and σ is the effective width of the topological neighborhood.

Since the topological neighborhood has some dependence with the lateral distance, as seen in Equation (4), then a one-dimensional grid $d_{j,i}$ is an integer equal to $|j - i|$. However, when it is two-dimensional it is defined by Equation (5) where the discrete vector $\mathbf{r}_j$ is the position of the excited neuron *j* and $\mathbf{r}_i$ is the position of the winning neuron *i*, both being measured in the discrete output space. An interesting feature of SOM is that the size of the topological neighborhood decreases with time and this is done by having the width σ of Equation (4) decrease with time [26].

$$d_{j,i}^2 = \left\| \mathbf{r}_j - \mathbf{r}_i \right\|^2 \tag{5}$$

$$\sigma(n) = \sigma_0 \ \exp\left(-\frac{n}{\tau_l}\right) \ n = 0, 1, 2, ..., \tag{6}$$

σ_0 is the value of σ when starting the SOM and τ_l a time constant. Thus, we have the topological neighborhood defined as a time variable according to Equation (7):

$$h_{j,i(X)}(n) = \exp\left(-\frac{d^2{}_{j,i}}{2\sigma^2(n)}\right) \ n = 0, 1, 2, ..., \tag{7}$$

So one can conclude that when the time *n* increases, the topological neighborhood decreases exponentially, as does the

width $\sigma(n)$. This topological neighborhood $h_{j,i(X)}(n)$ was used in the study and will be reference from this point on.

3) Adaptation: In the adaptation phase, for the grid to be self-organizing, the synaptic weight vector W_j of the neuron *j* of the grid must be changed relative to the input vector **X** [23]. However, there is a problem of saturation of the weights at the end of the process, but it can be corrected by changing the Hebbian assumption with a forgetting term $g(y_i)\mathbf{W}_j$ where $\mathbf{W}_j$ is the synaptic weight of the neuron *j* and $g(y_i)$ is the positive scalar function of the response y_i, the constant term of the Taylor series expansion of function $g(y_i)$ being zero [23].

Neighboring neurons to the winner increase their discriminant function values based on the input dataset, and as appropriate adjustments applied to their synaptic weights improve a subsequent input dataset. As Kohonen [22], the Equation (8) of updating is defined by:

$$W_j(n+1) = W_j(n) + \eta(n)h_{j,i(X)}(n)(X - W_j(n)) \quad (8)$$

n equals epoch, $\eta(n)$ is the learning rate, and $h_{i,j(x)}(n)$ is the neighborhood function.

According to Kohonen [22] there are two phases in the adaptive process: a sorting phase, which organizes the topology of the weight vectors, and a convergence phase, which adjusts the feature map, producing a statistical quantization of the input space. In the sorting phase, the learning rate parameter $eta(n)$ starts with a value of 0.1 and decreases to a value close to 0.01. The neighborhood function must contain almost all the neurons of the grid around the winning neuron *i* with a slow reduction over time and may require 1000 or more iterations. In the convergence phase the number of iterations should be at least 500 times the number of neurons in the lattice, and for good statistical accuracy, the learning rate parameter $eta(n)$ should be close to 0.01 and never zero. The neighborhood function has only the nearest neighbors of the winning neuron, which can be either one or zero neighboring neurons [23].

The originality of the SOM learning presented in the study may contribute to improving the diagnosis of ADHD patients by presenting a set of similar prototypes that represent the combination of attributes and then using a classification algorithm, such as the decision tree, which directly and quickly separates patients prone to the disorder. With this, the study can help health professionals and researchers to develop more accurate tools and reduce the cost of diagnostic tests that currently make their application unfeasible in the public health system.

C. Decision Tree

A decision tree is an Artificial Intelligence algorithm capable of organizing attributes from a dataset in priority, so that it can generate a path that leads to a decision for a classificatory attribute [27] [28]. A decision tree is built using algorithms that variously split a data set into branch-like segments. These segments create an inverted decision tree beginning at a root node at the top of the tree to the leaf at its end. Each object in the study is reflected in the root node and is a simple, one-dimensional display in the decision tree view. The name of the attribute is displayed along with a spread of values that are contained in the attribute. Its display shows all records in the dataset, attributes, and their values are viewed on the analysis object. The development of the decision rule for forming branches under the root node has, in the extraction method, a relationship between the object of analysis and one or more attributes that are used as input attributes to create the branches or segments, being used to estimate the likely value of the target, or outcome attribute or dependent attribute [29].

According to Castro [30], the structure of the decision tree is composed of internal nodes that correspond to an attribute test, each branch represents the test result, and the classes or class distributions are represented by the leaf nodes. The start node represents the root node, and the path from this node to the leaf node is called the classification rule. The decision tree construction can be used to classify an unknown class object, and the estimation is done by testing the attribute values in the tree and traversing until it reaches the leaf node. In this way, the result of the decision tree algorithm becomes a process that is easy to understand and visualize.

The decision tree algorithm as presented by Linoff and Berry [31] is a hierarchical collection of rules that describes how to divide a large collection of objects into successively smaller groups of objects. With each successive split, the members of the resulting followings become more similar to each other. For the selection of attributes to be chosen for division, one can use the information gain (*Gain*), which is one of the best-known methods and has as its generating base the entropy that is a measure of purity [32]. Thus, the gain is the expected reduction of entropy and has as its main function the division of attributes in the data set. Shannon entropy, on the other hand, measures the purity of the dataset [33], being a measure of the heterogeneity of the input data (*S*) relative to its classification (*m*). The expected value of the entropy of the attribute *E(A)* is given by Equation (9) and the expected information of S is given by Equation (11). Thus, the key factor is the use of a gain function that allows the attributes (A) to be compared to select the most relevant one. The chosen attribute is the one that maximizes the information gain which is calculated as being [30] [32]:

With n_{ij} being the number of objects in class C_i in a subset S_j, then the expected attribute information can be:

$$E(A) = \sum_{j=1}^{v} \frac{n_{ij} + ... + n_m j}{n} * I(n_{ij}, ..., n_{mj}) = \sum_{v \in values(A)} p(A_v) Entropy(A_v) \quad (9)$$

The entropy (Shannon's) [33] measures the impurity of the dataset, being a measure of the heterogeneity of the input dataset (S) relative to its classification (c). The *Gain(S,A)* is given by the Equation (10) and the entropy of *S* is given by the equation 11. Thus, the key factor is the use of a gain function that allows the attributes (A) to be compared to select the most relevant one. The attribute chosen is the one that maximizes the information gain which is calculated as being [33]:

$$Gain(S, A) = I(S) - E(A) \quad (10)$$

S is the input dataset, *A* is the attributes, and Θ represents the probability of *A* multiplied by its entropy.

$$I(S) = I(C_1, C_2, ..., C_m) = \sum_{k=1}^{m} -p_i log_2 p_i \quad (11)$$

The information Gain is given by the Equation (10) and represents the expected reduction in entropy when the value of the attribute *A* is known, since the process calculates the gain for each attribute, choosing the attribute with the highest gain to be tested in the set *S*. This process creates the division of objects to form the decision tree, giving rise to the node, labeling the attribute, and creating branches for each attribute value.

III. Proposed Method

The work presents a proposal for an unsupervised learning model as a method used in the identification of the neurons of the grid with greater diagnostic doubt of ADHD, that is, the diagnostic doubt in the neuron shows that it is difficult for both a machine learning algorithm and an expert to make a diagnosis. Thus, the paper brings a proposal to apply a decision tree on the neurons that show overlap to suggest which attributes are more discriminative. To understand this overlapping, the entropy (of Shannon) was calculated with the purpose of measuring the impurity of the neuron with its dataset, that is, the closer the entropy is to one, the greater the impurity of the neuron's dataset. Given this fact, a combination of SOM with the decision tree algorithm, which is a supervised model used in data classification to help identify one or more attributes from standardized assessment tools, such as cognitive tests and behavioral assessment inventories were sought. These tools were used to test the learning of ADHD characteristics.

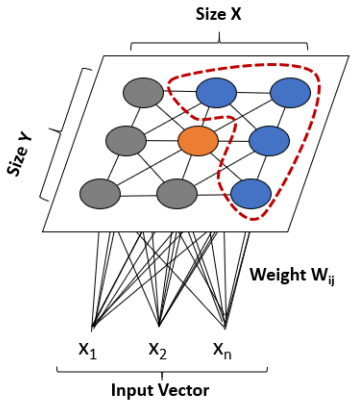


Figure 2. Segmentation of neurons in the self-organizing map

The objective of this decision tree algorithm was to verify the accuracy of the model for the confirmation of cases with ADHD diagnosis by identifying which assessment tools best contributed to this ADHD confirmation.

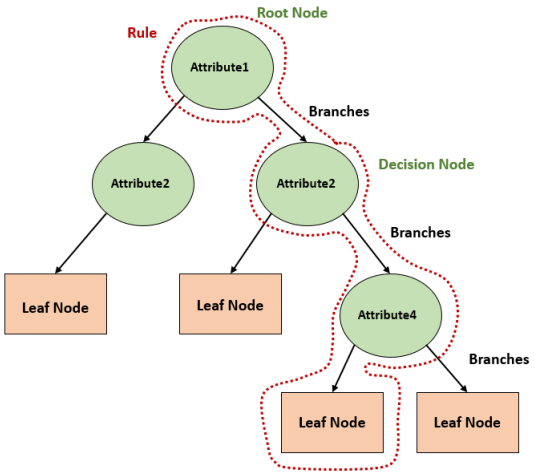


Figure 3. Decision tree structure

Then Kappa statistics is applied to measure the agreement and reliability of the observed diagnostic attributes with the expected diagnostic attribute, which is the measurement of the agreement of an expert's diagnosis compared to the algorithm's diagnosis. The next section will present the methodology developed with the data set and application of the artificial intelligence algorithms to arrive at the results of the work.

IV. Materials and Methods

The study sample consisted of 127 children and adolescents between 6 and 16 years old, 48 with a clinical diagnosis of ADHD and 79 from the control group, with no diagnosis of ADHD. The attributes that make up the neuropsychological tests and behavioral inventories applied in this study are

Attention Cancellation Test (TAC), Trail Making Test (TMT), Wechsler Intelligence Scale for Children (WISC-III), Wechsler Intelligence Scale for Children (WISC-IV), Wechsler Abbreviated Scale of Intelligence (WASI), Child Behavior Checklist for ages 6-18 (CBCL/6-18) and Teacher's Report Form for ages 6-18 (TRF/6-18).

These attributes were normalized by the z-score method [34] to standardize the different scales of the attributes. The normalized data property is used to train the network SOM using the package available in R language [35]. In this library, the functions *somgrid* and *som* are used to parameterize and train the map, respectively. According to Rubbo [36], the map size can be described according to Equations (12) and (13), with n being the number of objects and the constant Cm varying from [-3,3] to generate different map sizes:

$$l_{SOM} = \sqrt{n}/2 + Cm \tag{12}$$

$$Map\ Size = (l_{SOM})^2 \tag{13}$$

For the size of the map topology, the dimension 4x4 was chosen. With this, the hypothesis of the study was to find neurons with a representative density of objects and with a significant class distribution.

With the trained map, the analyses made were the density of objects in each neuron, the distance between neurons, the quality of adjustment of the neurons, the contribution of the attributes in the formation of neurons, and the distribution of the label of each object in each neuron. In addition to the outputs analyzed, the representativeness of the number of objects contained in each neuron with the label attribute was sought in the table generated by the SOM. In this way, the neurons of greater relevance were identified, that is, with larger numbers of objects generated by the SOM algorithm.

From this point on, the entropy algorithm (Shannon's) was used on each neuron in the network to select the neuron with the highest class overlap along with the representativeness of objects that are difficult cases to diagnose. By identifying neurons with overlapping classes, their objects are selected from the database generated by the SOM network for training and validation of the decision tree algorithm. The result of the decision tree brought a hierarchy of attributes in order of discrimination for cases of diagnostic doubt, and the validation of the algorithm shows the performance of the classification.

A. Rating Performance Evaluation

Table I shows the confusion matrix that was used to analyze the classification performance of the decision tree. The table indicates the prediction of the positive and negative scenarios, as well as current true and false scenarios [37]:

- TN is the correct number of negative predictions;
- FP is the number of false positive predictions;
- FN is the number of false negative predictions;
- TP is the correct number of positive predictions.

Table I. Confusion matrix.

	Predicted Negative	**Predicted Positive**
Current False	TN	FP
Current True	FN	TP

From the confusion matrix, it is possible to measure the performance of the algorithm by calculating the accuracy, as follows:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{14}$$

$$Error = (FP + FN)/(TP + TN + FP + FN) \tag{15}$$

B. Kappa Statistics

The measurement of agreement or Kappa coefficient (K) was used to comparatively measure the ADHD diagnosis of children submitted to the inventories corresponding to the disorder by the ADHD diagnosis predicted in the decision tree. The Kappa coefficient can be calculated with the results of the confusion matrix by Equation (16) [38], [39]:

$$kappa = P(O) - P(E)/1 - P(E) \tag{16}$$

Where P(O) is the observed probability of agreement (sum of the concordant answers divided by the total); P(E) is the expected probability of agreement (sum of the expected values of the concordant answers divided by the total). According to Silva [38], Kappa is a measure of interobserver agreement that evaluates the degree of agreement, as well as whether it is beyond what is expected given chance. It is a maximum unit value measure that corresponds to the absolute agreement and values close to zero or negative, indicating no or lack of agreement between the attributes being judged.

To simplify the methodological description developed in the study, we present a flowchart that briefly describes the process through the programming created to obtain the result and can be seen in Figure 4.

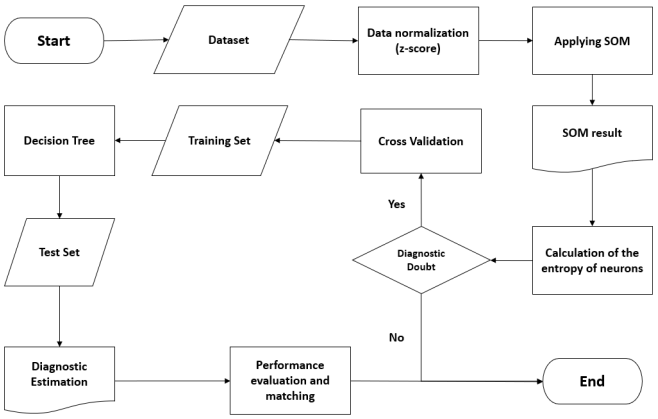


Figure 4. Dataset modeling flowchart

After this stage, it is possible to better understand the model's contribution to the understanding of Attention Deficit Hyperactivity Disorder, as well as to the diagnostic evaluation of patients. The next section presents the results obtained in this work.

V. Results

The training result of the SOM network can be seen in two different visualizations, depicted in Figures 5 and 6. Figure 5 presents the attributes, common to the trials, graphically distributed in each neuron. The sizes indicate the contribution that each attribute has to the formation of the neuron. Note that neighboring neurons have similarities among the attributes. In Figure 6, the diagnosis, an attribute that is not used in training the SOM, is projected on the map, allowing visualization of which neurons have the overlap of class 1 (group diagnosed with ADHD) and 2 (control group without ADHD). The network could not separate the diagnosed cases in neuron 4.

Table II presents for each neuron the percentage of objects of each class. Neuron 4 is the one with the highest concentration of objects (40%) and overlapping classes in the whole dataset.

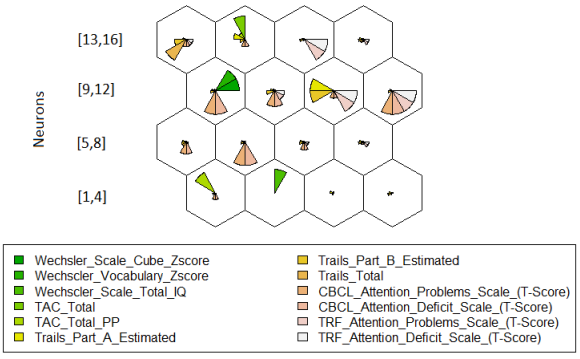


Figure 5. Contribution of the attributes in the formation of the neuron

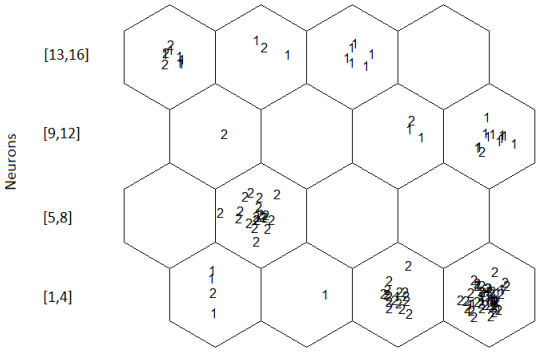


Figure 6. Scattering of objects diagnostic within neurons

Table II. Comparative diagnosis by the neuron dimension 4X4.

	Diagnostic	1	2	Total
neuron	1	3 (6.2%)	1 (1.3%)	4 (3.1%)
	2	1 (2.1%)	0 (0.0%)	1 (0.8%)
	3	1 (2.1%)	16 (20.3%)	17 (13.4%)
	4	16 (33.3%)	35 (44.3%)	51 (40.2%)
	6	0 (0.0%)	20 (25.3%)	20 (15.7%)
	9	0 (0.0%)	1 (1.3%)	1 (0.8%)
	11	2 (4.2%)	1 (1.3%)	3 (2.4%)
	12	11 (22.9%)	1 (1.3%)	12 (9.4%)
	13	6 (12.5%)	3 (3.8%)	9 (7.1%)
	14	2 (4.2%)	1 (1.3%)	3 (2.4%)
	15	6 (12.5%)	0 (0.0%)	6 (4.7%)
Total		48 (100.0%)	79 (100.0%)	127 (100.0%)

The result of the decision tree with the data mapped onto neuron 4 can be seen in Figure 7. The result shows that the Child Behavior Checklist for ages 6-18 attribute, specifically the probability of attention problems scale (T-score) [40] [41] neurons had the highest discrimination.

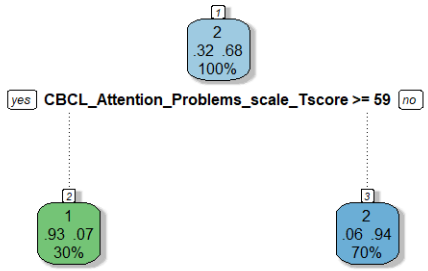


Figure 7. Decision tree of neuron 4

Finally, Figure 8 allows you to visualize all six attributes with greater discrimination for complex cases relative to the integration of clinical evaluation and evaluation using tests for a confirmation of the diagnosis.

The decision tree is a supervised algorithm, has as response attribute an estimated of the class attribute and the result can

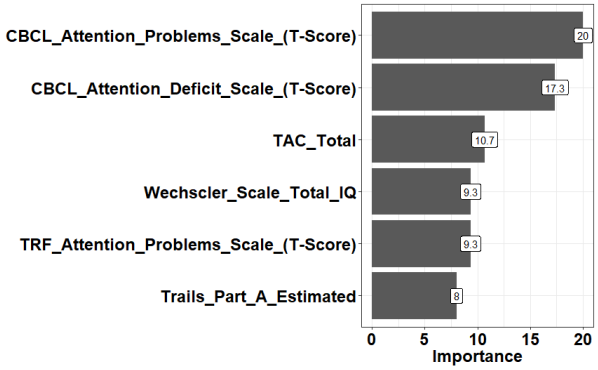


Figure 8. Importance of the attributes in neuron 4 by decision tree

be seen in Table III. Based on the confusion matrix generated, it is observed that of the 35 positive diagnoses for ADHD collected by applying the inventories, there are 33 positive diagnoses estimated by the algorithm. Based on the numbers presented, one can calculate the accuracy of the algorithm, as well as use the Kappa statistic to measure the agreement between the observed objects and the estimated objects [39]. The result with the 4x4 grid generated an accuracy of 88% with a good Kappa reliability of 72%, with the p-value equal to $2.68e^{-7}$.

Table III. Confusion matrix

	Predicted Negative	**Predicted Positive**
Current False	12	4
Current True	2	33

The results presented by the SOM and the decision tree corroborate the three dimensions analyzed and bring a new perspective to cases of doubt in diagnosing ADHD. However, the next section will discuss how the study could help through the mathematical understanding of the interpretation of neurodevelopmental disorder and the possibility of future work using newly supervised or/and unsupervised computational approaches to improve on groups with overlapping diagnoses of ADHD.

VI. Conclusion and Future work

Data from the behavioral assessment inventory [6] can generally be more susceptible to respondent bias because it is based on the answers of the subject. This bias is less so when using cognitive tests which are assessment measures applied directly to the person. Mathematical understanding and model generation is likely to become more difficult using only behavioral inventories. Since ADHD demands the use of both types of measures, in this study both tools were used to apply the decision tree. In the study, it was possible to group the children with and without ADHD by SOM, which made it possible to understand from the perspective of each grouping what was most important in their formation. The self-organizing map contributed especially to the formation of groups and the understanding of clusters with class overlapping, which is the proposal of this work. In this case of overlapping to diagnose a disorder, the decision tree was used to classify the attributes that contributed to the formation of the ADHD group. With this, the predominance of characteristics that helped in the understanding of ADHD in children and adolescents in the study was observed.

The application of the decision tree identified six attributes, namely two of cognitive assessment and four of behavioral assessment, that showed relevant discrimination to make the diagnosis. The Child Behavior Checklist for ages 6-18 attribute the one that showed the highest discriminative power. However, the incidence of low T-scores on the attention problems scale and attention deficit scale does not necessarily imply that the child has ADHD. The results presented showed the difficulty and complexity of finding indicators that define ADHD, as already signaled by some authors [6] [7] [9] [42] [43]. Importantly, the diagnosis of ADHD is a clinical diagnosis that considers the measurement of behavioral correlates of attentional deficits and indicators of hyperactivity and impulsivity in more than one environment. With the Child Behavior Checklist for ages 6-18 attribute being a parent-reported measure, the validity of these two scales for identifying ADHD will likely be confirmed. However, when disregarding the scales, one should consider the evaluations made with the cognitive tests that directly make cognitive measurements and are essential to decide the diagnosis of ADHD. In this study, the tests that contributed the most to this decision tree were the Attention Cancellation Test (ACT) and the Trail Making Test (TMT).

The study presented as a relevant factor the case of overlapping diagnoses of neurons when using the SOM and, in conjunction with the decision tree, was able to separate 88% of the cases. This way, future works can collaborate with the technique addressed in the study through supervised data procedures. These tools can help in making comparisons between results of standardized tests aiming to reduce possible biases of behavioral evaluations based on informants' reports. Future studies can test the same decision tree on larger samples to see if the attributes that showed higher accuracy are maintained. By doing so, the best indices of cognitive and behavioral assessment instruments that contribute to the increased accuracy of ADHD diagnosis may be identified. Since this study controlled for no comorbidities in the ADHD group, it is recommended for future studies to use sample groups with and without ADHD comorbidities from other psychiatric and neurodevelopmental conditions. This type of sample may allow the testing of new and more complex models due to the natural overlap of signs and symptoms between ADHD and some of these comorbidities.

VII. Acknowledgment

We thank the Academic Excellence Program of the Coordination for the Improvement of Higher Education Personnel (CAPES-PROEX), Process number 1133/2019, (CAPES-PROSUC) in mode II, the National Council for Scientific and Technological Development (CNPq, Cases 307730/2017-4 and 307443/2019-1), the Mackenzie Research Fund (MACKPESQUISA) of Mackenzie Presbyterian University, and the Foundation for Research Support of the State of São Paulo (FAPESP, Cases 2018/01063-0 and 2019/20757-1) for financial support and this study was financed in part by the Academic Excellence Program of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES-PROEX) - Finance Code 88887.581390/2020-00.

References

[1] A.Silva, L.Carreiro, M. Silva, M. Teixeira and L. Silva, "Combining Self-Organizing Maps and Decision Tree to Explain Diagnostic Decision Making in Attention-Deficit/Hyperactivity Disorder" The Sixth International Conference on Neuroscience and Cognitive Brain Information BRAININFO 2021. IARIA, Jul. 2021, pp. 5-10.

[2] American Psychiatric Association, DSM-5: Diagnostic and Statistical Manual of Mental Disorders. Artmed publisher, 2014.

[3] E. M. Mahone and M. B. Denckla, "Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective," Journal of the International Neuropsychological Society: JINS, Oct. 2017, vol. 23, no. 9-10, pp. 916-929, doi: 10.1017/S1355617717000807.

[4] E. Aretouli, "How neuropsychology can inform our understanding of preschool ADHD: Clinical and research implications," Applied Neuropsychology: Child, Apr-Jun. 2019, vol. 8, no. 2, pp. 174-181, doi: 10.1080/21622965.2017.1421463.

[5] M. L. Wolraich, J. F. Jr. Hagan, C. Allan, E. Chan, D. Davison, M. Earls, S. W. Evans, S. K. Flinn, T. Froehlich, J. Frost, J. R. Holbrook, C. U. Lehmann, H. R. Lessin, K. Okechukwu, K. L. Pierce, J. D. Winner and W. Zurhellen, "Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents," Pediatrics, Mar. 2020, vol. 145, no. 3, doi: 10.1542/peds.2019-2528.

[6] L. R. R. Carreiro, J. S. Schwartzman, C. N. Cantiere, A. F. Ribeiro, N. A. Silva, M. A. F. Martin, C. M. Chiquetto, G. S. Baraldi, M. M. C. Mariani, M. F. F. Seraceni and M. C. T. V. Teixeira, "Interdisciplinary neuropsychological, behavioral and clinical assessment protocol for children and adolescents with complaints of inattention and hyperactivity," Psychology: theory and practice, [online], Dec. 2014, vol. 16, no. 3, pp. 155-171. Available from: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872014000300012&lng=pt&nrm=iso

[7] S. Chandana and K. Vijayalakshmi, "An Approach to Measure and Improve the Cognitive Capability of ADHD Affected Children Through EEG Signals," 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, Jul. 2018, pp. 314-318, doi: 10.1109/ICALT.2018.00079.

[8] H. J. Kam, Y. M. Shin, S. M. Cho, S. Y. Kim, K. W. Kim and R. W. Park, "Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity," Applied clinical informatics, Nov. 2010, vol. 1, no. 4, p. 377, doi: 10.4338/ACI-2010-05-RA-0033.

[9] S. H. Lee, B. Abibullaev, W. Kang, Y. Shin and, J. An, "Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps," ICCAS 2010, Oct. 2010, pp. 2439-2442, doi: 10.1109/ICCAS.2010.5670255.

[10] G. Polanczyk, M. S. Lima, B. L. Horta, J. Biederman and L. A. Rohde, "The worldwide prevalence of adhd: a systematic review and metaregression analysis," American journal of psychiatry, Am Psychiatric Assoc, Jun. 2007, vol. 164, no.6, pp. 942–948, doi: 10.1176/ajp.2007.164.6.942.

[11] V. Simon, P. Czobor, S. Bálint, A. Mészáros and I. Bitter, "Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis," The British Journal of Psychiatry, Cambridge University Press, Mar. 2009, vol. 194, no.3, pp. 204–211, doi: 10.1192/bjp.bp.107.048827.

[12] E. B. P. Benczik and E. B. Casella, "The impact of Attention Deficit Hyperactivity Disorder (ADHD) on family relationships," Journal of Psychopedagogy, Jan-Mar. 2015, vol. 32, no. 97, pp. 93-103, 2015.

[13] E. G. Willcutt, J. T. Nigg, B. F. Pennington, M. V. Solanto, L. A. Rohde, R. Tannock, S. K. Loo, C. L. Carlson, K. McBurnett and B. B. Lahey, "Validity of dsm-iv attention deficit/hyperactivity disorder-symptom dimensions and subtypes," Journal of abnormal psychology, AmericanPsychological Association, Nov. 2012, vol. 121, no. 4, pp. 991-1010, doi: 10.1037/a0027347.

[14] M. Cherkasova, E. M. Sulla, K. L. Dalena, M. P. Pondé and L. Hechtman, "Developmental course of attention deficit hyperactivitydisorder and its predictors," Journal of the Canadian Academy of Child & AdolescentPsychiatry, Feb. 2013, vol. 22, no. 1, pp. 47-54.

[15] S. Dalsgaard, M. K. Humlum, H. S. Nielsen and M. Simonsen, "Common danish standards in prescribing medication for childrenand adolescents with adhd," European child & adolescent psychiatry, Springer, Sep. 2014, vol. 23, no. 9, pp. 841–844, doi: 10.1007/s00787-013-0508-5.

[16] N. M. Williams, I. Zaharieva, A. Martin, K. Langley, K. Mantripragada, R. Fossdal, H. Stefansson, K. Stefansson, P. Magnusson, O. O. Gudmundsson, O. Gustafsson, P. Holmans, M. J. Owen, M. O'Donovan and A. Thapar, "Rare chromosomal deletions and duplications in attention-deficithyperactivity disorder: a genome-wide analysis," The Lancet, Elsevier, Oct. 2010, vol. 376, no. 9750, pp. 1401–1408, doi: 10.1016/S0140-6736(10)61109-9.

[17] D. Puper-Ouakil, N. Ramoz, A. M. Lepagnol-Bestel, P. Gorwood and M. Simonneau, "Neurobiology of attention deficit/hyperactivity disorder," Pediatric research, Nature Publishing Group, May. 2011, vol. 69, no. 8, pp. 69–76, doi: 10.1203/PDR.0b013e318212b40f.

[18] L. R. R. Carreiro, M. C. T. V. Teixeira, M. M. M. Silva, I. T. Paes, R. A. C. B. Novaes and A. P. R. Micieli, "Neuropsychological Assessment in ADHD: contributions to the identification of cognitive difficulties and school counseling," In: C. A. H. Amato, D. Brunoni and P. L. Boggio, (Ed.). DEVELOPMENTAL DISORDERS: Interdisciplinary Studies. 1.ed. São Paulo: Memnon, 2018. vol. 1, c. 10, pp. 142–153. ISBN 978-85-7954-146-9.

[19] E. Merényi, M. J. Mendenhall, and P. O'driscoll, "Advances in self-organizing maps and learning vector quantization," Advances in Intelligent Systems & Computing, Jan. 2016, vol. 295, pp. 370, https://doi.org/10.1007/978-3-319-28518-4.

[20] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," IEEETransactions on neural networks, IEEE, May. 2000, vol. 11, no. 3, pp. 586–600, doi: 10.1109/72.846731.

[21] A. Rauber, D. Merkl and M. Dittenbach, "The growing hierarchical self-organizingmap: exploratory analysis of high-dimensional data," IEEE Transactions on NeuralNetworks, IEEE, Nov. 2002, vol. 13, no. 6, pp. 1331–1341, doi: 10.1109/TNN.2002.804221.

[22] T. Kohonen, "Self-organized formation of topologically correct feature maps," Biologicalcybernetics, Springer, Jan. 1982, vol. 43, no. 1, pp. 59–69, doi: 10.1007/bf00337288.

[23] M. Kubat, "Neural networks: a comprehensive foundation by Simon Haykin," Macmillan, 1994, ISBN 0-02-352781-7. The Knowledge Engineering Review, Feb. 1999, vol. 13, no. 4, pp. 409-412.

[24] T. Kohonen, "Essentials of the self-organizing map.Neural networks," Elsevier, 2013, vol. 37, pp. 52–65, doi: 10.1016/j.neunet.2012.09.018.

[25] C. C. Aggarwal, "Deep Reinforcement Learning. In: Neural Networks and Deep Learning," Springer, Cham, 2018, pp. 373-417, doi: 10.1007/978-3-319-94463-0.

[26] H. Ritter, "Neural computation and self-organizing maps: an introduction," Reading, MA: Addison-Wesley, 1992, pp. 141-161.

[27] M. J. Berry and G. S. Linoff, "Data mining techniques: for marketing, sales, and customer relationship management," John Wiley & Sons, 2004.

[28] I. H. Witten, E. Frank, M. A. Hall, and C.J. Pal, "Data Mining: Practical machine learning tools and techniques," 2005, vol. 2, pp.4.

[29] B. D. Ville and P. Neville, "Decision Trees for Analytics: Using SAS Enterprise Miner," SAS Institute Cary, NC, 2013, pp. 1-14.

[30] L. N. D. Castro and D. G.Ferrari, "Data Mining Introduction," Saraiva, 2016, pp. 344–360.

[31] G. S. Linoff and M. J. Berry, "Data mining techniques: for marketing, sales, andcustomer relationship management," John Wiley Sons, 2011.

[32] A. Colin, "Building decision trees with the ID3 algorithm," Dr Dobbs Journal, 1996, vol. 21, no. 6, pp. 107-+.

[33] T. Mitchel, "Machine learning," mcgraw hill education (ise editions), 1997, pp. 870-877.

[34] A. E. Curtis, T. A. Smith, B. A. Ziganshin and J. A. Elefteriades, "The mystery of the Z-score," AORTA Journal, Aug. 2016, vol. 4, no. 4, pp. 124-130, doi:10.12945/j.aorta.2016.16.014.

[35] R. Wehrens and J. Kruisselbrink, "Flexible self-organizing maps in kohonen 3.0," Journal of Statistical Software, Nov.2018, vol. 87, no. 1, pp. 1-18.

[36] M. Rubbo and L. A. Silva, "Filtering-based instance selection method for overlapping problem in imbalanced datasets," J, Jul. 2021, vol. 4, no. 3, pp. 308–327, ISSN 2571-8800, https://doi.org/10.3390/j4030024.

[37] S. Ruuska, "Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle," Behavioural processes, Aug. 2018, vol. 148, pp. 56-62, doi: 10.1016/j.beproc.2018.01.004.

[38] R. Silva and A. Paes, "Inside Statistics: Kappa concordance test," publication/uuid/3E5F4C37-E639-43D6-89C9-96597CA6AB40, Educ. Contin. health Einstein, 2012, vol. 10, no. 4, pp. 165–166.

[39] H. C. Kraemer, "Kappa coefficient," Wiley StatsRef: Statistics Reference Online, WileyOnline Library, Jun. 2014, pp. 1–4, doi: 10.1002/9781118445112.stat00365.pub2.

[40] H. Budak, S. E. Tasabat, "A modified t-score for feature selection," Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik, 2016, vol. 17, no. 5, pp. 845-852, doi: 10.18038/aubtda.279853.

[41] T. M. Nolan, L. Bond, R. Adler, L. Littlefield, P. Birleson, K. Marriage, A. Mawdsley, R. Salo and B. J. Tonge, "Child Behaviour Checklist classification of behaviour disorder," Journal of paediatrics and child health, Oct. 1996, vol. 32, no. 5, pp. 405-411, doi: 10.1111/j.1440-1754.1996.tb00939.x.

[42] R. A. Barkley and K. R. Murphy, "Attention-deficit hyperactivity disorder: A clinical workbook," Guilford Press, 2006.

[43] F. L. Cibrian, G. R. Hayes, and K. D. Lakes, "Research Advances in ADHD and Technology," Synthesis Lectures on Assistive, Rehabilitative and Health-Preserving Technologies, 2020, vol. 9, no. 3, pp. i-156, doi: 10.2200/S01061ED1V01Y202011ARH015.

A Symmetry-based Hybrid Model to Improve Facial Expressions Prediction in the Wild During Conversational Head Movements

Arvind Kumar Bansal
Department of Computer Science
Kent State University
Kent, OH, USA
email: arvind@cs.kent.edu

Mehdi Ghayoumi
eCornell and Department of Computer Science
Cornell University and University of San Diego
San Diego, CA, USA
email: mg948@cornell.edu

***Abstract*— Human emotion prediction is an important aspect of conversational interactions in social robotics. Conversational interactions involve a combination of dialogs, facial expressions, speech modulation, pose analysis, head gestures, and hand gestures in varying lighting conditions and noisy environment involving multi-party interaction. Head motions during conversational gestures, multi-agent conversations and varying lighting conditions cause occlusion of the facial feature-points. Popular Convolution Neural Network (CNN) based predictions of facial expressions degrade significantly due to occluded feature-points during extreme head-movements during conversational gestures and multi-agent interaction in real-world scenarios. In this research, facial symmetry is exploited to reduce the loss of discriminatory feature-point information during conversational head rotations. CNN-based model is augmented with a new rotation invariant symmetry-based geometric modeling. The proposed geometric model corresponds to Facial Action Units (FAU) for facial expressions. Experimental data show hybrid model comprising a CNN-based model, and the proposed geometric model outperforms the CNN-based model by 8%-20%, depending upon the type of facial-expression, beyond partial head rotations.**

Keywords-Artificial Intelligence; conversation; deep neural network; emotion analysis; facial expression analysis; facial occlusion; facial symmetry; head movement; multimedia.

I. INTRODUCTION

Due to an aging population in the developed world and limited workforce, there is a growing need of social robotics for elderly care and healthcare [1]-[3]. To show empathy, interact, and converse with humans, social robots need to understand human emotions and pain in the wild [4]-[9].

Predicting emotions in the wild is complex and requires multimodal media analysis involving dialogs, voice-modulation (including timed silence), gestures (including postures, gaze, conversational head and hand gestures, and haptic gestures), facial expressions, pain and tears [10]-[22]. Many desirable human-robot interactions, such as conversational gestures, including human warmth and affection, frustration, irritation, encouragement, impatience and pain shown by a combination of voice-modulation, speech-phrases, gestures, facial expressions and haptic touch are yet to be achieved [13]. Compared to emotions exhibited in dialogs, utterances and gestures, facial expressions are exhibited more involuntarily and express a major subset of expressed human emotions and acute pain [5]-[7], [10]-[20].

Interpreting facial expressions is context sensitive and augmented with other modalities such as speech or scene analysis [11], [12], [14]. Facial expressions and their intensity vary by gender, age and culture. A subset of facial expressions has universally accepted interpretations, and current-day research on facial expression analysis of basic emotions and pain assumes universally accepted meanings [5]-[7], [15].

In real-life scenarios, a face continuously moves during a conversation based upon 1) conversational gestures, such as argumentation, interrogation and denial; 2) intensity of emotion; 3) multi-party interactions, changing ambient lighting and shadows with head-movements [13]. Head rotations stochastically occlude feature-points causing information loss hindering accurate facial-expression classification as illustrated in Fig. 1.

In cognitive psychology, two approaches study facial expressions: 1) valence and intensity based Plutchik's eight emotion classes and their subcategories; 2) basic six emotions (*anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*) popularized by Ekman and others [5], [6]. Computational recognition of facial expressions is based on analysis of facial video modeled by Facial Action Unit System (FACS) [6], [15]-[17].

Figure 1. An example of recognizing facial expression in the wild. [Image source: Wikimedia Commons, public domain; Credit: US Navy, Bureau of Medicine and Surgery, 1945; Available at: https://commons.wikimedia.org/wiki/File:Navy_nurse_signing_cast_--_WWII.jpg]

Many Facial Action Units (FAUs) are also associated with acute pain that interferes with emotion analysis [18]-[20]. In this paper, we focus on the recognition of facial expressions under the assumption that conversing agents do not suffer from pain.

Previous studies are mostly limited to the frontal facial view or statically aligned poses using curated databases showing nonoccluded facial expressions in proper lighting conditions [15], [23]-[36]. Recent augmentation of CNN-based modeling with *Long Short-Term Memory* (LSTM), transfer learning and multiple feed-forward neural networks (FNN) improve the prediction of facial expression during head movements [37]. However, the model does not handle extreme information loss beyond partial occlusion and does not exploit facial symmetry. Experiments with CNN-based model show that facial expression prediction drops by 10-20% for partial occlusion (less than 45° rotation) and by 30-50% beyond 45° rotation as described in section V.

CNN-based models need to restore occluded discriminatory feature-points for beyond the partial occlusion in conversational head-gestures, such as emotional disagreement, interrogation, argumentation or denial; multi-party interaction that involves significant occlusion of one part of the face. Luckily, even during extreme head-rotations, only one side of the face is occluded. Hence, facial symmetry can be used to reconstruct the occluded discriminatory feature-points by estimating the angle of facial rotation and knowing the coordinates of their counterparts on the nonoccluded side.

This research improves facial-expression analysis under head-motion by utilizing facial symmetry along the vertical major axis [38]-[40]. Prediction uses 1) estimation of the angle of facial rotation using nonoccluded feature-points; 2) inherent facial symmetry around the vertical axis of the face; 2) differences between the symmetrical points and the actual geometric feature-points from the previous frames.

The proposed hybrid model integrates CNN-based classification for partially occluded space and the symmetry-based geometric model classification beyond partially occluded space. The proposed symmetry-based geometric model also provides motion continuity and temporal context to the CNN classifier for selecting the nearest static alignment.

The major contributions in this research are:

1. Development of a symmetry-based geometric model corresponding to *Facial Action Units* (FAUs) to recover discriminative feature-points during conversational head-rotations in real-time scenarios;
2. Augmentation of the CNN-based model with the proposed symmetry-based geometric model to improve the temporal context and the facial expression prediction beyond partial occlusion.

The overall roadmap is as follows. Section II describes background concepts. Section III describes the related work. Section IV describes the proposed symmetry-based geometric model. Section V describes an overall architecture. Section VI describes the implementation and discusses experimental results. Section VII discusses the limitations and concludes the paper.

II. BACKGROUND

A. Facial Muscles and FACS System Correspondence

A combination of facial muscles expressing facial expressions and pain, is shown in Fig. 2. The associated muscles and their functions are described in Table I. Italicized descriptions in Table I mark the FAUs involved in both pain and facial expressions. The compression of muscles is externally visible through facial feature-points, as illustrated in Fig. 3. A combination of movement of the feature-points forms the basis of geometric modeling and *Facial Action Unit System* (FACS).

Facial expression analysis is based on mapping a subset of FAUs to basic facial expressions. Tables II describes the FAUs associated with the simulations of the six basic facial expressions [6], [15]-[17].

B. Facial Feature-points and Symmetry

There are two types of facial feature-points: *fixed points* and *active points*. *Fixed points* act as a reference, and *active-points* move during facial-expressions, altering x and z-coordinates of feature-points [16].

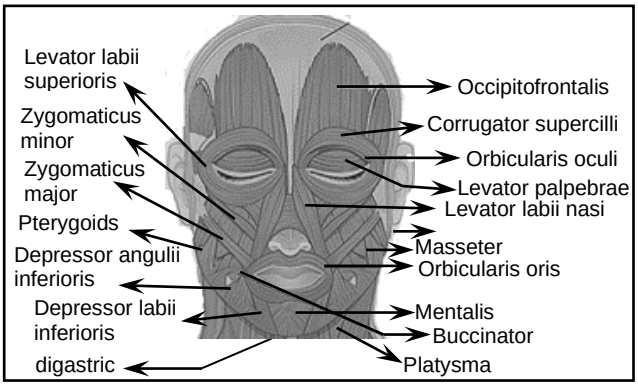


Figure 2. Facial muscles used in facial-expressions of emotion and pain. [Image adopted from Wikimedia Commons, Credit: CNX anatomy 2013]

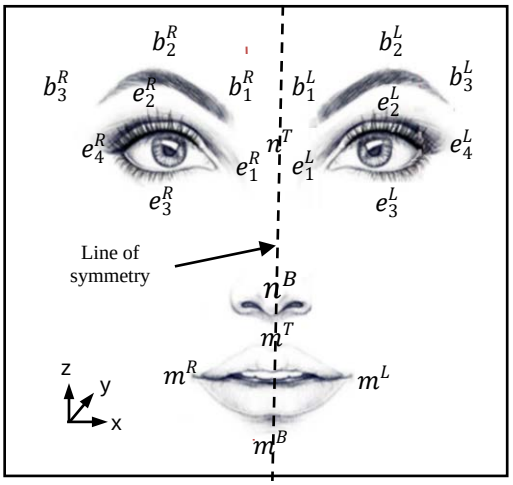


Figure 3. Facial feature points with symmetry

TABLE I. RELATED FAUS AND ASSOCIATED FACIAL MUSCLES. 'E' DENOTES EMOTION, AND 'P' DENOTES PAIN. THERE IS AN OVERLAP BETWEEN TWO SUBSETS.

FAU	Function	Facial Muscle	Type
1	inner brow raiser	occipitofrontalis	E
2	outer brow raiser	occipitofrontalis	E
4	*brow lowerer*	*corrugator supercilli*	E + P
5	Upper eyelid raiser	levator palpebrae superioris	E
6	*Cheek raiser*	*Orbicularis oculi, pars orbitali*	E + P
7	*Lid tightener*	*orbicularis oculii*	E + P
8	Lips towards each other	orbicularis oris	E
9	*Nose wrinkler*	*levator labii superioris nasi*	E + P
10	*Upper lip raiser*	*levator labii superioris*	E + P
11	Nasolabial deepener	zygomaticus minor	E
12	Lip corner puller	zygomaticus major	E
14	Dimpler	buccinator	E
15	Lip corner depressor	depressor anguli inferioris	E
16	Lower lip depressor	depressor labii inferioris	E
17	Chin raiser	mentalis	E
20	*Lip stretcher*	*platysma*	E + P
23	Lip tightener	orbicularis orbis	E
25	Lips part	depressor labii inferioris	P
26	*Jaw drop*	*masseter*	E + P
27	*Mouth stretcher*	*pterygoids, digastric*	E + P
41	*Lid droop*	*Relaxation of levator palpebrae superioris*	E + P
43	Eyes Closed	relaxation of levator palpebrae superioris	P

TABLE II. SIX BASIC FACIAL EXPRESSIONS AND FAUS

Basic facial expressions	FAU subset
Surprise	1, 2, 5, 10, 16, 26
Fear	1, 2, 4, 5, 15, 20, 26
Disgust	2, 4, 9, 15, 17
Anger	2, 4, 5, 7, 9
Happiness	6, 12, 14, 20, 27
Sadness	4, 8, 11, 15, 23, 41

Feature-point denotations use 'e' for eye; 'b' for brow; 'm' for mouth. A subscript enumerates feature-points for the same organ. A superscript denotes left-side by 'L'; right-side by 'R'; top by 'T'; bottom by 'B'.

A face has six major *fixed points* and 14 major active points. Fixed points are two ends of the left and right eyes (e_1^L, e_4^L, e_1^R, e_4^R); bottom of a nose (n^B); middle point between two eye-brows above the nose-tip (n^T). Active major points are: 1) three points on each brow (b_1^L, b_2^L, b_3^L, b_1^R, b_2^R, b_3^R); 2) two middle points of lips (m^T and m^B); 3) two endpoints of the mouth (m^R, m^L); 4) two middle points in each eye (e_2^L, e_3^L, e_2^R, e_3^R).

Facial features are symmetrical around the vertical axis as illustrated in Fig. 3: left-side of the vertical line $n^T n^B$ is symmetrical to the corresponding feature-points on the right-hand side. This symmetry causes similar changes on both sides of a face for most facial-expressions at the muscle level.

C. Occlusion and Head Movement

In a real-world situation, the head rotations are observed every 5 - 7 degrees [41]. The angle of rotation maps to one of the internal states based upon an identifiable resolution in the feature-points. Distances between the symmetry-axis and the feature-points on the nonoccluded side are used to estimate the coordinates of occluded feature-points using facial-symmetry.

D. Deep Learning Models for Facial Expression Analysis

Convolution neural network comprises a *cascade of convolution-layers*: *convolution filter*, *Rectified Linear Unit* (*RELU*) and *subsampler* (*pooling layer*) followed by a fully connected feed-forward neural network (FNN) [10], [37], [42]. In order to model a continuously moving head, multiple FNNs, one for each desired orientation, are used for the classification [37]. Due to the large number of angles, the classification is approximate, lowering the accuracy.

Long Short Term Memory (LSTM) is a variation of *Recurrent Neural Network* (RNN) where the previous outputs are fed back and gated to regulate the output [43], [44]. LSTM significantly reduces the problem of vanishing gradient and instability in RNN, and improves long range contextual dependency [44].

Transfer learning adapts meta-level learning rules to a similar but somewhat different domain or task [45]. A *domain* is a pair (ψ, $P(X)$) where ψ is the feature-space, $P(X)$ is the marginal probability-distribution, and X is a feature-vector (x_1, …, x_N) $\in \psi$. A *task* is modeled as a pair (Y, f), where Y is a label space $\{y_1, \ldots, y_M\}$, and f is a function that maps a feature-value x_i in X to a label $y_j \in Y$. The differences could be in feature-vector, probability distribution, label space, or mapping of feature-value to label space.

For facial-expression recognition, feature-vectors change due to angular variations. Thus, the feature-vector has to be approximated and modified based upon proximity and similarity analysis with input feature-vector values of an FNN to reduce classification-error.

E. Notations

Line-segments are denoted by two end feature-points or their intuitive description. For example, eye-width is denoted as EW or $e_1^L e_4^L$. Lip-width is denoted by LW or $m^T m^B$. Given a line-segment LS, magnitudes of the x-axis, y-axis and z-axis component are denoted respectively by $|LS|_X$, $|LS|_Y$, and $|LS|_Z$. In this paper, parameterization is illustrated using left-side of a face. The technique applies also to the right-side of the face.

III. RELATED WORK

In recent years, many researchers have suggested techniques to handle information loss caused by partial occlusion and multiple orientations due to head movements. Related work can be classified as: 1) handling occlusion for improper lighting conditions, hand-gestures and external objects such as eye-glasses, hats, scarfs, and medical masks; hand gestures; hair and mustaches; ambient lighting conditions; 2) analyzing emotions in the wild; 3) mapping motion as a set of fixed alignments; 4) a combination of CNN,

LSTM and transfer learning to map continuous motion to the corresponding CNN [9], [15], [23]-[31], [36], [37], [46].

Most occlusion handling schemes handle occlusion of static faces in frontal pose based on the reconstruction of small patches of partially occluded parts using the corresponding nonoccluded (or global) facial texture [23]-[31]. The schemes combine information from previous images, texture-patterns, or other features from nonoccluded space, and dynamic weighting of texture-patterns to reconstruct occluded patches using well curated datasets.

To handle the occlusion caused by external objects, researchers have trained hybrid models on occluded and nonoccluded samples and used nonoccluded features-space as a guidance to predict texture of occluded patches, combination of sparse representation and maximum likelihood estimation, a combination of Gabor filter and co-occurrence matrix of local features, a combination of Gabor descriptors and deep structure recognition based upon deep belief network, a combination of local binary pattern, feature histogram, dimension reduction and support vector machine (SVM), a fusion of information derived using Weber Local Descriptors and histograms of multiple features using SVM, combining global and local textures with CNN and attention, the use of LSTM auto-encoders and transfer learning, and Bayesian networks. These schemes do not analyze occluded facial expressions in the wild during conversational head-motion.

Zong et al. use a combination of transduction transfer learning and linear discriminant analysis to map the trained data using curated dataset to the data in the wild [9]. However, the scheme does not: 1) handle conversational head movements and the resulting occlusion; 2) does not use symmetry to recover occluded feature-points.

T-H. S. Li et al. integrate CNN with LSTM to provide the temporal context required for analyzing facial expression during head rotation [37]. They use transfer learning to map a position to the corresponding static alignment of CNN for improved accuracy. The scheme is limited by the number of fixed domains for transfer learning and does not exploit symmetry. Besides, LSTM cannot estimate the coordinates of the occluded feature-points explicitly.

These techniques are not suited for extreme loss of discriminatory feature-points during extreme head-rotations in argumentation, denial or multi-party interactions where a significant part of face is occluded for a longer period.

The proposed geometric model exploits facial symmetry to recover occluded feature-points during extreme head rotations [38]-[40]. The correspondence of the line-segments joining discriminatory feature-points to *Facial Action Units* (FAUs) relates the proposed hybrid model with *Facial Action Coding System* (FACS) based analysis and CNN-based analysis. In addition, the changes in line-segment ratios with head-movements provide temporal context even beyond partial occlusion. In our scheme, the discriminative feature-points also support multimodal analysis of head-gestures and provide explanation capability.

IV. PROPOSED SYMMETRY-BASED GEOMETRIC MODEL

Facial expression analysis requires: 1) removal of the distortions caused by camera zooming; 2) removal of the distortions in the line-segments caused by head-rotations, and 3) correspondences of parameters to the changes in FAUs.

The identification of parameters invariant to head-rotations requires the use of *fixed feature-points* that act as a reference to measure the changes in orientation and lengths of the line-segments with varying facial expressions.

The motions of *active-points* that contribute to the facial expressions are: 1) vertical and horizontal motion of b_1^L, b_2^L, b_2^L on an eyebrow; 2) vertical motions of $\{e_2^L, e_3^L\}$ in the center of an eyelid, 3) vertical and horizontal motions of m^L (lip-endpoints), and 4) vertical motions of m^T, m^B and $\{m_1^L, m_2^L\}$ (lip-midpoints). Fig. 4 shows left side of the face with the required feature-points and line-segments used in the facial expression classification.

The line-segments for the facial expression analysis are: $n^B b_1^L$, $n^B b_2^L$, $n^B b_3^L$, $b_1^L b_3^L$, *EH* ($e_2^L e_3^L$: eye-height), *LH* ($m^T m^B$: lip-height), *LW* ($m_1^L m^L$: lip width), *EL* ($m^L e_c^L$: lip segment to the eye (e_c^L is the left center of an eye given by $\frac{e_2^l + e_3^l}{2}$). The line-segments $n^B b_1^L$, $n^B b_2^L$, $n^B b_3^L$, $b_1^L b_3^L$ and *EL* have x-magnitudes and z-magnitudes.

The line-segments *LH* and *EH* have z-magnitudes; the line-segment *LW* has x-magnitude. With no rotation and zooming, changes in the x-magnitudes and z-magnitudes of these line-segments correspond to different facial expressions. In an actual scenario, these line-segments vary with head-rotations and image scaling due to the camera-zooming. These line-segments are mapped to parameters invariant to head-rotations and camera zooming, such that the resulting parameters vary with facial expressions only.

Four line-segments, joining fixed-points, $n^B n^T$, $e_1^L e_4^L$, $n^T e_1^L$, and $n^T e_4^L$ have been used to derive parameters invariant with respect to head rotation. The effect of zooming is removed by dividing the z-magnitudes by the magnitude of the line-segment $n^B n^T$.

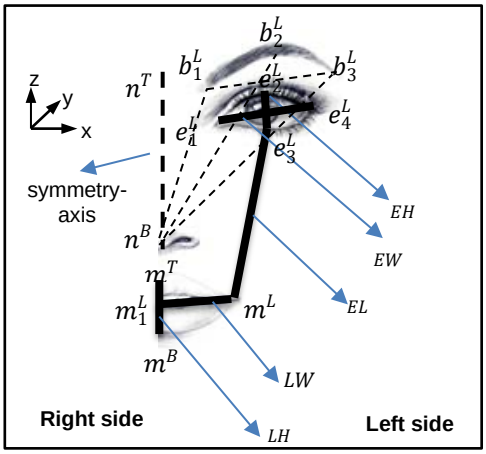


Figure 4. Facial feature-vectors

To minimize the effect of variation of x-coordinates during a head-rotation, the most *aligned fixed segments* are chosen that are affected similarly by the head-rotation compared to line-segments involving *active points*.

A division of line-segments by the x-magnitude of the line-segments involving the nearest fixed-points parallel to the same axis minimizes the effect of rotation and preserves the changes due to facial expressions.

The division by the segment $e_1^L e_4^L$ provides invariance for the eye-brow area. The division by x-magnitude $|\, n^T e_1^L \,|_X$ cancels the effect of head-rotation on the magnitude $|\, n^B b_I^L \,|_X$. The division by the x-magnitude $|\, n^T e_4^L \,|_X$ cancels the effect of the head-rotation on the magnitude $|LW\,|_x$.

A. Frontal Pose Estimation

The fixed feature-points nose-bottom n^B, left inner-eye e_1^L and right inner-eye e_1^R are used to establish frontal pose (see Fig. 3 and Fig. 4). The ratio $|\, n^T e_1^L \,| \,/\, |\, n^T e_1^R \,| = 1$ for the frontal-pose, only altering during head-rotation. The overall estimate for the frontal pose is given by (1) where ϵ is an experimentally derived value slightly greater than zero to take care of involuntary and random head-movements.

$$1 - \epsilon \leq |\, n^B e_1^L| \,/\, |\, n^B e_1^R| \leq 1 + \epsilon \qquad (1)$$

Estimation of rotation angles is based on missing landmarks on the rotated side of the face. The landmarks n^t and n^b become invisible in the complete occlusion and are visible between partial and complete occlusion. For rotation to the left or right, the ratio changes beyond $1 \mp \epsilon$.

The proposed line-segments (LH, LW, EL, EH, $|\, b_1^L\ b_3^L \,|_x$, $|n^B b_1^L|_z$, $|n^B b_2^L|_z$, $|n^B b_3^L|_z$, $n^B n^T$, EW, $|n^T n^B|$) cover all FAUs to express six basic emotions, as described in Table II. The overall correspondence is summarized in Table III.

Variations in the line-segment *LH* reflect tightening or opening of lips and mouth, and jaw-drop. It is associated with FAU 8 (lips towards each-other), FAU 10 (upper lip-raiser), FAU 16 (lower lip-depressor), FAU 17 (chin-raiser), FAU 23 (lip-tightener), FAU 26 (jaw-drop) and FAU 27 (mouth-stretcher).

TABLE III. LINE-SEGMENTS

Line-ratio	Normalized ratio	Description
R^{LH}	$\|LH\| \,/\, \|n^B n^T\|$	lip height ratio
R^{LW}	$\|LW\|_x \,/\, \|EW\|$	lip-width ratio
R^{EL}	$\|EL\|_z \,/\, \|n^B n^T\|$	eye-to-lip ratio
R^{BW}	$\|\, b_1^L\ b_3^L \,|_x \,/\, EW$	brow-width ratio
R^{IBH}	$\|\, n^B b_1^L \,|_z / \|n^B n^T\|$	inner brow-height ratio
R^{MBH}	$\|\, n^B b_2^L \,|_z / \|n^B n^T\|$	mid-brow height ratio
R^{OBH}	$\|\, n^B b_3^L \,|_z / \|n^B n^T\|$	outer-brow height ratio
R^{EH}	$\|EH\| \,/\, \|n^B n^T\|$	eye-height ratio

Variations in the line-segment *LW* reflect compression and stretching of a mouth. It corresponds to FAUs 6, 12, 14, 20, 23 and 27. These FAUs are involved in *happiness* (lip-corner and cheek-stretching obliquely up), and *sadness* (lip-corner stretching obliquely downwards).

Variations in the z-component $|EL|_Z$ (eye-to-lip vertical component) measure compression and stretching of cheek muscles. The decrease in $|EL|_Z$ corresponds to FAU 6 (cheek-raiser) associated with *happiness*. The increase in $|EL|_Z$ corresponds to FAU 15 (lip-corner depression) associated with negative emotions *fear, disgust* and *sadness*. The change in the magnitude of the line-segments *EW* (eye-width) and *EH* (eye-height) correspond to FAU 7 associated with *anger*. The magnitude $|EH|$ increases during *anger* due to the raising of the upper eyelid and middle eye-brow point.

Variations in eye-brow length $|\, b_1^L\ b_3^L \,|$ (brow compression and stretching) correspond to FAU 1 (inner brow raiser), FAU *2* (upper brow raiser) or 4 (brow lowerer). However, only the x-component $|\, b_1^L\ b_3^L \,|_x$ is used because vertical variations in eye-brow are processed by $|n^B b_1^L|_z$, $|n^B b_2^L|_z$ and $|\, n^B b_3^L \,|_z$. The increase in $|\, b_1^L\ b_3^L \,|_x$ corresponds to FAU 4 (brow-lowerer) associated with negative emotions: *fear, disgust*, *anger*, and *sadness.*

The z-component $|n^B b_1^L|_z$ corresponds to inner-eyebrow raising or lowering. The increase in magnitude $|n^B b_1^L|_z$ corresponds to FAU 1 associated with *surprise*. The decrease in $|n^B b_{i(1\leq i\leq 3)}^L|_z$ corresponds to FAUs 4 and 9 associated with negative emotions: *fear, disgust, sadness,* and *anger*. The increase in the magnitude $|n^B b_3^L|_z$ corresponds to FAU 2 associated with *fear*.

B. Normalized Ratios

In the beginning, the frontal pose is recorded to derive the original coordinates of feature-points and the original length and orientation of line-segments. The zooming distortion and head-rotation distortions in the x-direction are removed from the feature-points and the corresponding line-segments.

Vertical segments $|n^B b_1^L|_z$, $|n^B b_2^L|_z$ $|n^B b_3^L|_z$, EH and $|EL|_Z$ are divided by $|n^B n^T|$ to derive the corresponding normalized ratios. Horizontal line-segment $|LW|_X$ and $|\, b_1^L\ b_3^L \,|_X$ are divided by $|n^T e_1^L|$ and EW, respectively. The normalized ratios are summarized in Table IV.

C. FAU Correspondence

Table V describes conditions by combining the normalized ratios across the same or different video-frames that are sampled periodically because facial expressions alter after few seconds.

The increase in the ratio R^{LH} corresponds to FAU 10 (upper lip raiser), FAU 26 (jaw-drop), and FAU 27 (mouth-stretch). The decrease in the ratio R^{LH} corresponds to FAU 8 (lips towards each other), FAU 16 (lower lip-depressor), FAU 17 (chin-raiser), and FAU 23 (lip-tightener).

The increase in the ratio R^{LW} corresponds to FAU 6 (cheek-raiser), FAU 12 (lip-corner puller), FAU 15 (lip-corner depressor), FAU 16 (lower lip-depressor), and FAU 20 (lip-stretcher). The decrease in the ratio R^{LW} corresponds to FAU 23 (lip-tightener). The increase in the ratio R^{EL} corresponds to FAU 15 (lip-corner depressor); the decrease in the ratio R^{EL} corresponds to FAU 6 (cheek-raiser). The increase in the ratio R^{EH} corresponds to FAU 5 (upper lid raiser); the decrease in the ratio R^{EH} corresponds to the FAU 7 (lid tightener) or FAU 41 (lip-stoop). The increase in the ratio R^{BW} corresponds to FAU 4 (brow-lowerer).

The increase in the ratio R^{IBR} corresponds to FAU 1 (inner eye-brow raiser); the decrease in the ratio R^{IBR} corresponds to FAU 4 (brow-lowerer). The increase in the ratio R^{OBR} corresponds to FAU 2 (outer eye-brow raiser); the decrease in R^{OBR} corresponds to FAU 4 (eye-brow lowerer).

TABLE IV. LINE-SEGMENTS AND FAU CORRESPONDENCE

Line-seg.	FAUs	Basic emotions
LH	8, 10, 16, 17, 23, 26, 27	anger, disgust, fear, sadness, surprise
LW	6, 12, 15, 16, 20, 23	happiness and sadness
EL	6, 15	disgust, fear, happiness, sadness
EH	5, 7	anger
$\lvert b_1^L\ b_3^L \rvert_x$	4	anger, disgust, fear, sadness
$\lvert n^B b_1^L \rvert_z$	1, 4, 9	anger, disgust, fear, sadness, surprise
$\lvert n^B b_2^L \rvert_z$	4, 5	fear and surprise
$\lvert n^B b_3^L \rvert_z$	2	fear
$n^B n^T$	used for vertical normalizations	
$EW, \lvert n^T n^B \rvert$	invariant with head-rotation	

TABLE V. FAUS AND NORMALIZED RATIO CONDITIONS

FAUs	Condition ($n = m + k$ and $k > 0$)
#1	$R_n^{IBR} < R_m^{IBR}$
# 2	$R_n^{OBR} > R_m^{OBR}$
#4	$R_n^{IBR} < R_m^{IBR} \wedge R_n^{MBR} < R_m^{MBR} \wedge R_n^{OBR} < R_m^{OBR}$
#5, 27	$R_n^{EH} > R_m^{EH}$
#6, 12	$R_n^{LH} < R_m^{LH} \wedge R_n^{EL} < R_m^{EL}$
#7, 41	$R_n^{EH} < R_m^{EH}$
#8	$R_n^{LH} < R_m^{LH}$
#10	$R_n^{LH} > R_m^{LH}$
#15	$R_n^{EL} > R_m^{EL} \wedge R_n^{EW} > R_m^{EW}$
#16	$R_n^{LH} < R_m^{LH} \wedge R_n^{EL} > R_m^{EL}$
#17	$R_n^{EL} < R_m^{EL}$
#20	$R_n^{LW} < R_m^{LW}$
#23	$R_n^{LW} > R_m^{LW}$
#26	$R_n^{EL} > R_m^{EL}$

A simultaneous decrease in the ratio R^{LH} and an increase in the ratio R^{EL} correspond to the activation of FAU 16 (lower-lip depressor). Simultaneous decreases in the ratios R^{LH} and R^{EL} correspond to the activations of FAU 12 (lip-corner puller) and FAU 6 (cheek-raiser). Simultaneous increases in the ratios R^{EL} and R^{EW} correspond to FAU 15 (lip-corner depression). Simultaneous decreases in the ratios R^{IBR}, R^{MBR} and R^{OBR} and increase in the ratio R^{BW} corresponds to the activation of FAU 4 (eye-brow lowerer).

V. AN OVERALL ARCHITECTURE

The proposed architecture, as illustrated in Fig. 5, comprises: 1) preprocessing and denoising module; 2) hybrid classifier module; 3) angle-based output selector module.

The *preprocessing and denoising module* denoises the data, reduces dimensions of facial images using *Locality Sensitive Hashing* (LSH) and uses Gabor filter to preserve directionality in images.

The *hybrid classifier module* comprises: 1) CNN-based classifier; 2) the proposed geometric-classifier; 3) facial orientation based output selector either from the CNN-based classifier or the proposed geometric-classifier.

There are two types of connectivity: 1) based on data flow, as shown by the solid arrows; and 2) time-stamped stream of angles emanating from the second submodule of the geometric-classifier as shown by the dashed arrows.

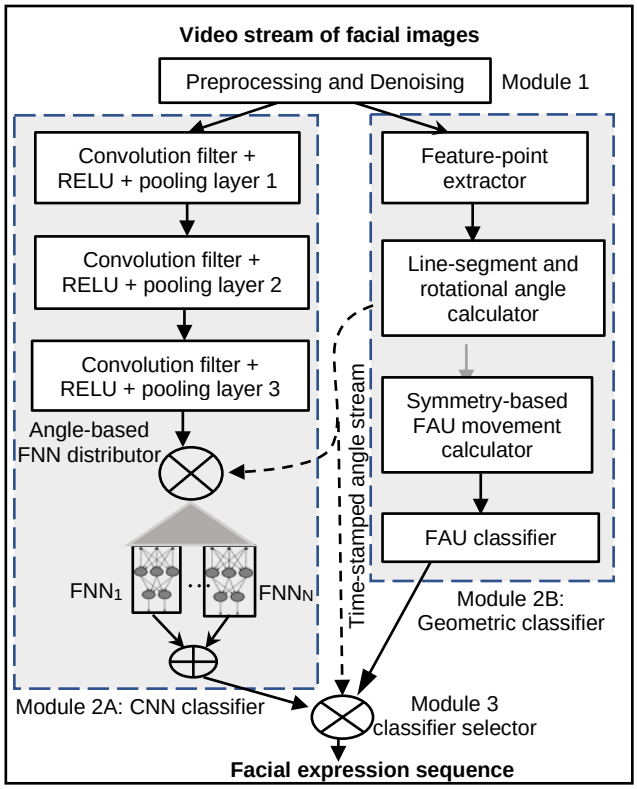


Figure 5. An overall architecture

The time-stamped angle-stream is used for: 1) selecting the output from one of the two classifiers; 2) maintaining sequentiality of the derived facial-expression labels coming from different classifiers; 3) selecting the optimal FNN in the CNN-classifier. Only one FNN is selected at a time based upon the value of the time-stamped angle.

The symmetry-based geometric-classifier has four submodules: 1) feature-point extractor; 2) invariant ratio and rotational angle calculator; 3) symmetry-based movement calculator; 4) FAU-based classifier.

Feature-point extractor extracts the visible non-occluded feature-points, calculates and time-stamps the line-segments as described in Table IV. The stream of time-stamped line-segments is passed to the second submodule.

Invariant ratio and Rotational Angle Calculator submodule derives the computable ratio $|n^B e_1^L| / |n^B e_1^R|$, $|n^T e_1^L| / |n^T e_1^R|$, $|n^T e_1^L| / |n^T n^B|$, $|n^B e_1^L| / |n^T n^B|$, $|n^T e_1^R| / |n^T n^B|$, or $|n^B e_1^R| / |n^T n^B|$, and looks up archived lookup tables to estimate the angle of rotation for each time-stamped feature-vector. This time-stamped angle is transmitted to: 1) FNN selector submodule within the CNN-classifier; 2) the third submodule of the geometric-classifier; 3) the classifier selector module. The submodule also computes the motion invariant ratio, as described in Section IV.

Symmetry-based movement calculator computes the normalized ratio conditions, as given in Table V for deriving the associated FAU movements. The input to the submodule is the motion invariant ratio and time-stamped angle derived in the submodule 2 of the classifier.

FAU-based classifier uses the derived motion of FAUs and their association with facial expressions (see Table II in Section II) to identify the facial expression [6], [16]. The input to the submodule is FAU motions derived in the third submodule of the geometric-classifier.

The CNN-classifier comprises three submodules: 1) three layers of cascaded convolution filters; 2) angle-based FNN distributor; 3) mutually exclusive FNNs that can recognize a facial-expression in one of the static orientations. FNNs are optimally trained for specific orientation in the lab conditions to improve the performance.

The choice of three layers in the convolution layer cascade and number of FNNs in the CNN-classifier is based upon size of the cropped images (56 × 56 pixels) after sampling and experimentation to reduce the computational overhead while maintaining sufficient accuracy.

Each FNN corresponds to a static orientation. Adjacent orientations are 15° apart instead of optimal 7-8° [41]. In our experiment, internal states change every 15° to reduce computational overhead. This choice slightly degrades (by 1-3%) the prediction accuracy for a tradeoff of reduced computational overhead.

Based upon each FNN corresponding to an orientation 15° apart, there are seven FNNs, one for each orientation (45° left rotation, 30° left rotation, 15° left rotation, frontal, 15° right rotation, 30° right rotation, 45° right rotation).

The angle-based distributor selects one of the FNN based upon the facial orientation. The angle is received from the second submodule of the geometric-classifier module.

The inputs to these FNNs are: output of CNN module's softmax layer distributed from the angle-based FNN distributor.

The *angle-based output selector module* selects the output from the CNN-classifier or geometric-classifier, depending upon the facial orientation. If the angle is less than the threshold value (45°), the output from the CNN-classifier is taken. For the head rotations greater than 45°, the output from the geometric-classifier is taken.

VI. IMPLEMENTATION AND EXPERIMENTATION

RaFD dataset was used for measuring the performance of the CNN-based model for various static alignments in different poses [35], [47]. Compared to other curated facial expression databases, RaFD gives comprehensive facial-expressions for 67 models (for all genders) with multiple camera angles and adjustment of lighting conditions [32]-[34]. We deployed 70% of the data for training, 15% validation, and 15% testing.

For the online video capturing, three frames per second were used for the facial expression analysis. The Epochs of 200 frames were used because the experimental data show that the accuracy of the facial expression recognition stabilizes around 200 frames. The stabilization of the 200 frame per second comes from the controlled lab conditions, and the hardware that we used for our experiments. The number will vary in noisy real environment.

A. CNN Classifier Implementation

The implemented CNN-based model is a cascade of three hidden layers: conv-32, conv-64 and conv-128, followed by a Softmax layer. Each *conv-m* layer contains *m* filters to extract different orientations. The conv-128 layer provides a subclassification of textures. After each convolution layer, there is a max-pooling layer for the subsampling of images. Each max-pool layer pools a 2 × 2 pixel macroblock.

After applying the Locality-Sensitive Hashing (LSH) and Gabor filter, the processed images are passed to the network of convolution layers through the input layer [17].

Each cropped image is scaled to 56 × 56 pixels. The data-size after the conv-32 layer is 56 × 56 × 32 pixels, and the output of first max-pooling layer after the conv-32 layer are 28 × 28 × 32 pixels. The output of the second max-pooling layer is 28 × 28 × 64 pixels. The output of the last hidden layer is 14 ×14 × 128. The output of the following max pooling layer is 7 × 7 × 128 pixels. Extracted features are concatenated by adding a fully connected layer at the end.

B. Result and Data Analysis

The hybrid model was executed in the wild. We use *recall* metrics to show accuracy of the proposed model because the recall gives the overall percentage of true positive. CNN model was also executed in wild for the frontal pose and compared against the results of RaFD dataset to derive the comparative deterioration of the *recall*.

In our statistical reporting of data, five occlusion states are used: 1) frontal face with no occlusion ($|\theta| < \epsilon$); 2) partial left-side or right-side occlusion (ϵ < rotation < 45°); 3) full left-

side or right-side occlusion (> 45°). Internal states map to one of the five states based upon interval inclusion.

C. Performance Evaluation and Discussion

The results are summarized in Tables VI, VII, and VIII. Tables VI and VIII show the recall values of CNN-model with RaFD dataset and the proposed hybrid model in wild, respectively. Table VII shows the confusion matrix for CNN model for a frontal pose in the wild.

Table VI illustrates CNN based prediction, even for a cured RaFD database, deteriorates quickly due to the unavailability of discriminatory feature-points on the occluded part of the face. The deterioration varies from 48% for sadness to 41% for happiness for complete occlusion.

Comparison of Tables VII and VIII illustrates that the accuracy of facial expression classification deteriorates in the wild even for the frontal pose: more for sadness (around 22%) and the least for disgust (around 6%). Even neutral face is labeled as sad for 10% of the time in the wild.

TABLE VI. RECALL IN THE CNN MODEL WITH RADB DATASET

	Right complete occl.	Right part occl.	Front no occl.	Left part occl.	Left complete occl.
sadness	49%	83%	97%	79%	48%
disgust	54%	81%	98%	88%	63%
anger	53%	81%	96%	87%	64%
fear	51%	86%	95%	81%	55%
surprise	57%	84%	98%	90%	53%
happiness	59%	85%	99%	92%	62%
neutral	54%	82%	95%	79%	51%

TABLE VII. CONFUSION MATRIX - CNN MODEL (FRONTAL) IN WILD

	sad. %	disg. %	ang. %	fear %	sur. %	happ. %	neutral
sadness	74.5	0.1	8.0	12.3	0.9	0.7	3.5
disgust	0.7	92.4	1.4	1.1	1.3	1.7	1.4
anger	6.4	2.3	79.3	2.5	1.6	2.4	5.5
fear	7.2	0.6	6.1	82.3	1.2	0.8	1.8
surprise	1.8	0.7	2.6	5.2	86.9	1.7	1.1
happiness	1.4	0.2	2.2	2.5	3.0	87.2	2.5
neutral	10.2	0.2	4.2	5.7	2.2	3.7	73.8

TABLE VIII. RECALL IN THE HYBRID MODEL IN WILD

	Right complete occlusion	Right part occlusion	Front no occlusion	Left part occlusion	Left complete occlusion
sadness	57%	68%	75%	69%	59%
disgust	70%	81%	92%	82%	70%
anger	73%	75%	79%	77%	76%
fear	66%	75%	82%	76%	67%
surprise	71%	74%	87%	76%	75%
happiness	75%	79%	87%	81%	77%

The reasons for this deterioration are: 1) mixing of facial muscles and feature-points for negative facial expressions, *sadness*, *fear* and *anger*, in real-time expressions; 2) variations in the intensity level of the expressed facial expressions in real-time; 3) continuous random head-motions during real-time facial-expressions causing noise; 4) uneven ambient lighting conditions with shadows obscuring feature-points; 5) randomly picking the video-frame may not correspond to the apex image corresponding to a facial-expression [46].

The facial expressions for the negative emotions: *sadness*, *fear*, and *anger* are often confused due to 1) the presence of common facial muscles; 2) the mixing of facial expressions in real-time; 3) improper temporal labeling during transition of a negative facial expression to another; 4) uncontrolled thought patterns affecting involuntary facial expressions in real-time. Another problem is that CNN is trained using fixed alignments, and a head-movement is approximated to one of the fixed poses.

Comparison of the occluded parts in Table VI and Table VIII shows that the hybrid model outperforms CNN-based prediction even for the curated RaFD dataset for beyond the partial occlusion. The improvement is 8% for sadness (minimum) to 21% for the happiness (maximum). In a multi-party interaction, where the change in the line-of-view may cause extreme occlusion, the hybrid model provides better accuracy and information.

VII. CONCLUSION AND FUTURE WORK

Head-motions during conversational gestures and multi-agent interactions cause extreme occlusion on one side of facial features. Automated feature-extracting and deep learning schemes are limited by the facial feature detections. Their performance degrades during extreme occlusion due to the nonavailability of discriminatory feature-points. Facial symmetry reconstructs the occluded discriminatory feature points. Combining CNN-based schemes with the proposed geometric modeling improves the performance in such a scenario by 8% – 21% beyond the partially occluded state.

The current scheme can be further improved by smoothening the derived facial-expression sequence and predicting the next facial-expression using Dynamic Bayesian Network (DBN), the knowledge of average duration of facial-expressions during emotional conversation, and sampling more video-frames for near-apex facial expressions [46].

In this paper, we have assumed that conversational agents are not suffering from any acute pain. As described in Section II, the facial-expressions for basic emotions and acute pain significantly overlap [7], [18]-[20], [48], [49]. Many times chronic pain is displaced and expressed as a combination of negative emotions, depression and anxiety [48], [49]. Many times, negative emotions and pain occur together and are inseparable [49]. Such cases can only be resolved by the knowledge of the situation, dialog understanding or scene analysis to build the needed context. For example, knowledge of cause of pain or video analysis of person's gesture or wound detection can provide sufficient context.

We are currently investigating the DBN on a sequence of facial-expressions to smoothen out the errors due to image frames missing the apex image for the corresponding facial expressions [46].

REFERENCES

[1] A. K. Bansal and M. Ghayoumi, "A Hybrid Model to Improve Occluded Facial Expressions Prediction in the Wild During Conversational Head Movements," The Tenth International Conference on Intelligent Systems and Applications (INTELLI 2021), July 2021, pp. 36-42, ISBN: 978-1-61208-882-2.

[2] M. I. Yenilmez, "Economic and social consequences of population aging the dilemmas and opportunities in the twenty-first century," Applied Research in Quality of Life, vol. 10, no. 4, pp. 735-752, Dec. 2015, doi: 10.1007/s11482-014-9334-2.

[3] D. H. García, P. G. Esteban, H. R. Lee, M. Romeo, E. Senft, and E. Billing, "Social Robots in Therapy and Care," 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019), March 2019, pp. 669-670, doi: 10.1109/HRI42470.2019.

[4] C. Peter and R. Beale (eds), "Affect and Emotion in Human-Computer Interaction: From Theory to Applications," LNCS 4868, Berlin / Heidelberg: Springer-Verlag, 2008, ISBN: 978-3-540-85098-4.

[5] R. Plutchik, "Emotion: A Psychoevolutionary Synthesis," New York, NY: Harper & Row, 1980.

[6] P. Ekman and W. V. Friesen, "Nonverbal Behavior," Communication and Social Interaction, P. F. Ostwald, (editor), New York, NY: Grune & Stratton, pp. 37- 46, 1977.

[7] K. M. Prakachin, "Assessing Pain by Facial Expression: Facial Expression as Nexus," Pain Research Management, vol. 14, no. 1, pp. 53-58, Jan./Feb. 2009, doi: 10.1155/2009/542964.

[8] F. Rothganger, "Computation in the Wild," [Online] Available from https://www.osti.gov/servlets/purl/1644432, accessed date; September 12, 2021.

[9] Y. Zong, W. Zeng, X. Huang, K. Yan, J. Yan, and T. Zhang, "Emotion recognition in the wild via sparse transductive transfer linear discriminant analysis, Journal of Multimodal Interfaces, vol. 10, pp. 163-172, June 2016, doi:10.1007/ s12193-015-0210-7.

[10] M. Ghayoumi, M. Thafar, and A. K. Bansal, "A Formal Approach for Multimodal Integration to Derive Emotions," Journal of Visual Languages and Sentient Systems, vol. 2, pp. 48-54, Oct. 2016, doi: 10.18293/DMS2016-030.

[11] C. M. Lee and S. Narayanan, "Towards Detecting Emotions in Spoken Dialogs," IEEE Transactions on Speech and Audio Processing, vol. 13, no.2, pp. 293-303, March 2005, doi: 10.1109/TSA.2004.838534.

[12] K. Han, D. Yu, and I. Tashev, "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine," 15th Annual Conference of the International Speech Communication Association, Sept. 2014, pp. 223-227, ISBN: 978-1-63439-435-2.

[13] A. Singh and A. K. Bansal, "Towards Synchronous Model of Non-Emotional Conversational Gesture Generation in Humanoids," Computing Conference, July 2021, Lecture Notes in Networks and Systems Series, vol. 283(1), K. Arai (editor), Cham, Switzerland: Springer, pp. 737-756, July 2021, ISBN 978-3-030-80118-2, doi: 10.1007/978-3-030-80119-9.

[14] J.-M. Fernandez-Dols, H. Wallbott, and F. Sanchez, "Emotion Category Accessibility and the Decoding of Emotion from Facial-expression and Context," Journal of Nonverbal Behavior, vol. 15, no. 2, pp. 107- 123, June 1991, doi: 10.1007/BF00998266.

[15] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial Expression Recognition: A Survey," Symmetry, vol. 11, no. 10, Article 1189, Sept. 2019, doi: 10.3390/sym11101189.

[16] M. Ghayoumi and A. K. Bansal, "An Integrated Approach for Efficient Analysis of Facial expressions," The 11th International Conference on Signal Processing and Multimedia Applications (SIGMAP 2014), Aug. 2014, pp. 211-219, ISBN: 978-989-758-046-8.

[17] M. Ghayoumi and A. K. Bansal, "Emotions in Robot using Convolutional Neural Network," International Conference on Social Robotics (ICSR 2016), Nov. 2016, Lecture Notes in Artificial Intelligence Series, Editor: A. Agah et el., , vol. LNAI 9979, Cham, Switzerland: Springer, pp. 285-295, ISBN: 978-3-319-47436-6, doi: 10.1007/978-3-319-47437-3_28.

[18] K. D. Craig, K. M. Prkachin, R. V. Grunau, "The facial expression of pain," In: D. C. Turk and R. Melzack, eds., Handbook of Pain Assessment, 3rd edition, New York: Guilford, pp. 117-133, 2011, . ISBN 978-1-60623-976-6.

[19] A. C. de C. Williams, "Facial Expression of Pain: an Evolutionary Account," Behavioral and brain sciences, vol. 25, no. 4, pp. 439–455, Aug. 2002, doi: 10.1017/S0140525X0200 0080.

[20] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, K. M. Prkachin, "Automatically Detecting Pain in Video Through Facial Action Units," IEEE Transactions on System, Man and Cybernetics, vol. 41, no. 3, pp. 664-674, June 2011, doi: 10.1109/TSMCB.2010.2082525

[21] A. Gračanin, E. Krahmer, M. Balsters, Dennis Küster, and A. J. J. M. Vingerhoets, "How Weeping Influences the Perception of Facial Expressions: The Signal Value of Tears," Journal of Nonverbal Behavior, vol. 45, pp. 83–105, March 2021, doi: 10.1007/s10919-020-00347-x.

[22] N. J. Emery, "The Eyes Have it: The Neuroethology, Function and Evolution of Social Gaze," Neuroscience and Behavioral Reviews, vol. 24, issue 6, pp. 581-604, Aug. 2000, doi: 10.1016/S0149-7634(00)00025-7.

[23] L. Zhang, B. Verma, D. Tjondronegoro, and V. Chandran, "Facial Expression Analysis Under Partial Occlusion: A Survey," ACM Computing Surveys, vol. 51, no. 2, Article No. 25, April 2018, doi: 10.1145/3158369.

[24] S. S. Liu, Y. Zhang, and K. P. Liu, "Facial Expression Recognition under Random Block Occlusion Based on Maximum Likelihood Estimation Sparse Representation," International Joint Conference on Neural Networks (IJCNN 2014), July 2014, pp. 1285-1290, doi: 10.1109/IJCNN32435. 2014.

[25] L. Shuaishi, Z. Yan, and L. Keping, "Facial-expression Recognition under Partial Occlusion Based on Weber Local Descriptor Histogram and Decision Fusion," The 33rd Chinese Control Conference (CCC 2014), July 2014, pp. 4064-4068, doi: 10.1109/CCC32826.2014.

[26] Q. Cheng, B. Jiang, and K. Jia, "A Deep Structure for Facial Expression Recognition under Partial Occlusion," The Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2014), Kitakyushu, Japan, Aug. 2014, pp. 211-214, ISBN: 978-1-47995-391-2.

[27] J. Y. R. Corenjo and H. Pedrini, "Recognition of Occluded Facial Expressions based on CENTRIST Features," IEEE International Conference on Acoustics, Speech and Signal Processing, March 2016, pp. 1298-1302, ISBN: 978-1-47999-988-0.

[28] R. Li, P. Liu, K. Jia, and Q. Wu, "Facial Expression Recognition under Partial Occlusion Based on Gabor Filter and Gray-level Co-occurrence Matrix," The International Conference on Computational Intelligence and Communication Networks (CICN 2015), Dec. 2015, pp. 347-351, doi: 10.1109/CICN.2015.6.

[29] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439-2450, May 2019, doi: 10.1109/TIP.2018. 2886767.

[30] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust LSTM-Autoencoders for Face De-Occlusion in the Wild," IEEE Transactions on Image Processing, vol. 27, no. 2, pp. 778-790, Feb. 2018, doi: 10.1109/TIP.2017.2771408.

[31] Y. Miyakoshi and S. Kato, "Facial Emotion Detection Considering Partial Occlusion of Face using Bayesian Network," IEEE Symposium on Computers & Informatics (ISCI 2011), March 2011, pp. 96-101, ISBN: 978-1-61284-689-7.

[32] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-specified Expression," International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2010), San Francisco, CA USA, June 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[33] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker, "Multi-PIE," In Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition, vol. 28, no. 5, May 2008, pp. 807-813, doi: 10.1016/j.imavis.2009.08.002.

[34] M. J. Lyons, M. Kamachi, and J. Gyoba, "Japanese Female Facial-expressions (JAFFE), " 1998, doi: 10.5281/zenodo.3451524.

[35] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and Validation of the Radboud Faces Database," Cognition & Emotion, vol. 24, no. 8, pp. 1377-1388, Nov. 2010, doi: 10.1080/ 02699930903485076.

[36] K. Seshadri and M. Savvides, "Towards a Unified Framework for Pose, Expression, and Occlusion Tolerant Automatic Facial Alignment," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 10, pp. 2110-2122, Oct. 2016, doi: 10.1109/TPAMI.2015.2505301.

[37] T-H S. Li, P-H Kuo, T-N Tsai, and P-C Luan, "CNN + LSTM Based Facial Expression Analysis Model for a Humanoid Robot," IEEE Access, vol. 7, pp. 93998-94011, July 2019, doi: 10.1109/ACCESS.2019.2928364.

[38] S. Derrode and F. Ghorbel, "Shape Analysis and Symmetry Detection in Gray-level Objects using the Analytical Fourier-Mellin Representation," Signal Processing, vol. 84, no. 1, pp. 25-39, Jan. 2004.

[39] S. Kondra, A. Petrosino, and S. Iodice, "Multi-scale Kernel Operators for Reflection and Rotation Symmetry: Further Achievements," IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2013), June 2013, pp. 217-222, doi: 10.1109/CVPRW.2013.4.

[40] M. Ghayoumi and A. K. Bansal, "Real Emotion Recognition by Detecting Symmetry Patterns with Dihedral Group," Third International Conference on Mathematics and Computers in Sciences and in Industry (MCSI 2016), Chania, Greece, Aug. 2016, pp. 178-184, doi: 10.1109/MCSI.2016.041.

[41] M. Amiri, G. Jull, and J. Bullock-Saxton, "Measuring Range of Active Cervical Rotation in a Position of Full Head Flexion using the 3D Fastrack Measurement System: an Intra-tester Reliability Study," Manual Therapy, vol. 8, no. 3, pp. 176-179, Aug. 2003, doi: 10.1016/s1356-689x(03)00009-2.

[42] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, et el., "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions" Journal of Big Data, vol. 8, Article 53, 2021, 74 pages, doi: 10.1186/s40537-021-00444-8.

[43] Y. Yu, X. Si, C. Hu, J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," Neural Computation, vol. 31, pp. 1235-1270, 2019, doi: 10.1162/ neco_a_01199.

[44] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," Transactions of Neural Networks and Learning Systems, vol. 28, issue 10, pp. 2222-2232, Oct. 2017, doi: 10.1109/TNNLS. 2016.2582924.

[45] S. J. Pan and Q. Yang, "A Survey of Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 10, pp. 1345 – 1359, Oct. 2010, doi: 10.1109/ TKDE.2009.19.

[46] A. Cruz, B. Bhanu, and N. S. Thakoor, "Vision and Attention Theory-Based Sampling for Continuous Facial Emotion Recognition," IEEE Transactions of Affective Computing, vol. 5, no. 4, pp. 418-431, Oct-Dec. 2014, doi: 10.1109/TAFFC.2014.2316151.

[47] A. R. Dores, F. Barbosa, C. Queirós, I. P. Carvalho, and M. D. Griffiths, "Recognizing Emotions Through Facial Expressions- A Large Scale Experimental Study," International Journal of Environmental Research and Public Health, vol. 17, Article 7420, Oct. 2020, doi: 10.3390/ ijerph17207420.

[48] Z. Chen, R. Ansari, and D. Wilkie, "Automated Pain Detection from Facial Expressions using FACS: A Review," [Online], arXiv:1811.07988v1, Available at https://arxiv.org/pdf/ 1811.07988.pdf .

[49] M. S. H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, et al., "The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal EmoPain Dataset," IEEE Transactions of Affective Computing, vol. 7, no. 4, pp. 435-451, Oct.-Dec. 2016, doi: 10.1109/TAFFC.2015.2462830

Linked Care – A digital Revolution in Mobile Care and Support

Development and Implementation of a Digital Documentation System for Interdisciplinary Information Transfer

Kathrin Mühlhauser
Department Health Sciences
FH Campus Wien
Vienna, Austria
e-mail: kathrin.muehlhauser@pmu.ac.at

Elisabeth Haslinger-Baumann
Competence Center for Applied Nursing Research
FH Campus Wien
Vienna, Austria
e-mail: elisabeth.haslinger-baumann@fh-campuswien.ac.at

Theresa Galanos
Competence Center for Applied Nursing Research
FH Campus Wien
Vienna, Austria
e-mail: theresa.galanos@fh-campuswien.ac.at

Franz Werner
Department Health Sciences
FH Campus Wien
Vienna, Austria
e-mail: franz.werner@fh-campuswien.ac.at

Doris Zeidler
Competence Center for Applied Nursing Research
FH Campus Wien
Vienna, Austria
e-mail: doris.zeidler@fh-campuswien.ac.at

Katharina Nopp
Department Health Sciences
FH Campus Wien
Vienna, Austria
e-mail: ka.nopp@protonmail.com

Abstract - **Mobile care professionals are facing a great workload and a lack of information exchange with other health professionals. Documentation work is often done off-time due to minimal time resources. Information and communication technology support can bring improvements for home care and all stakeholders, but its full potential is not exhausted. This project, namely Linked Care, develops information technology systems that address practice-oriented challenges, increase digital data availability, and reduce the work-related burdens of health professionals as well as clients and their support systems. Experiences with existing IT-systems, including the current existing Austrian Electronic Health Record System, serve as a starting point to enable a continuous information supply in mobile care within the project. The solution will provide interdisciplinary support of relevant participants in the care process, offering a new type of responsive user interface. Linked Care investigates the end-users' needs and specifies the processes in workshops with local representatives and regional administrators. Therefore, the solution allows for the exchange and evaluation of care data via standardized interfaces throughout Austria while taking the socio-economic, regional, and national environments into account. The project will create a software portal that can be linked to existing systems. The project consortium parties ensure the development of the solution as a business model. The generated Care Summary creates new possibilities for data exchange within the health care sector. In conclusion, Linked Care is the digital solution, which needs to be implemented in mobile care and support as it enables a precise and rapid communication transfer.**

Keywords - digital documentation; ICT; interdisciplinary communication; mobile health and care; digital healthcare.

I. INTRODUCTION

Nurses experience many burdens and challenges in their everyday work [1]. Documentation is often described as particularly stressful [2]. It is estimated that 30% of nurses' workload (working hours) is used for documentation. This time cannot be spent with clients, which in turn can lead to nurses finishing documentation work in their free time to have more available hours for direct client-based care [3]. The dramatically increasing number of people in need of care leads to an increase not only in the objectively measurable but also in the subjectively experienced care effort. As the complexity within the formal and informal care support network is increasing, the need for adequate client-related coordination between the involved parties is also growing. Continuous and detailed client documentation is the starting point of client related coordination and communication. However, the organization of client documentation in the care system is not only challenging in terms of effort, but it also shows deficiencies and complications regarding the information flow and exchange between the different formal and informal care giving/providing parties. Adequate documentation is indispensable for high-quality nursing and medical care, as well as interdisciplinary cooperation. Due to the dual financing system of medical care, nursing care and

homecare, the resulting responsibilities may be different in nature and "not necessarily conducive to the efficiency of the systems" [4]. Moreover, the lack of interdisciplinary exchange is closely related to financial and time constraints: only 60% of all mobile care service providers cooperate with professional groups outside their services [4]. On the one hand, the heterogeneity of the documentation and communication systems offered by different providers complicates the communication connection/link within the different mobile services/software solutions. On the other hand, the non-existence of an obligatory uniform solution forms a massive barrier to continuous information transfer. Sophisticated services, which are necessary to support recent technological innovations, have not been provided while creating data gaps in all areas of social welfare and health care. This causes discontinuance and translation problems within software solutions.

So far, the record validity was and is questionable due to existing language barriers between the carer's first language and the available application languages. Comprehensive documentation is particularly important. For example, specific (target) groups were identified, that are not familiar with professional care within their households (e.g., 'people with migration backgrounds', 'people with mental illnesses'). The reason is the lack of affordability in the presence of social disadvantage or the lack of information of particularly vulnerable groups, for whom care is mainly provided informally or within the family. This example points out the need for comprehensive documentation and communication that combines several information clusters. Therefore, it contributes to a subjectively facilitated situation and an objectively increased efficiency [4]. Beyond that, it also advocates the improvement of communication and interface management based on detailed surveys.

The aim of this project is to enable technology-based cooperation between clients as well as their support system and professional caregivers, nursing professionals, doctors, therapists, and pharmacies while using an efficient, secure and low-threshold digital tool, offering optimal information technology (IT) support.

This paper will provide insight into the Linked Care-project; starting with the 'State of the Art' in Section 2, followed by Section 3, 'Method', and finally describing the 'Results' as well as the 'Conclusion and Further Works' in Section 4 and 5.

II. State of The art

Care network services, such as Linked Care wants to provide, currently exist worldwide, but they neither include all necessary participants and stakeholders (e.g., 'medical care providers', 'pharmacies', 'care organizations') nor do they function end-to-end on a digital level. Once integrated into one organization, other necessary services such as visiting services, home services and nursing care have to be organized by the employees [5]. That is their greatest detriment, which the Linked Care-solution wants to eradicate.

Most health care professionals in Austria currently work with the digital documentation system Mobile Case and Care (Mocca) [6] and programs such as Gesellschaft für Software, Entwicklung und Datentechnick mbH (GSD Software®) [7] and Das Pflegeplanungs- und Dokumentationssystem mit ENP® (RECOM®-GriPS) [8], which are linked to one another via interfaces. Among the health care organizations involved in the Linked Care-project, documentation is still being practiced in analogue form while, in some cases, programs such as Microsoft Excel and Word are included. The use of dual documentation systems (simultaneous use of two systems by one organization) is currently also ubiquitous. This often results in double documentation, which leads to increased workload.

In the European Union (EU) and the United States of America (USA), there have been some digital developments in this field. Siemens eHealth Solutions [9] has created an electronic health network that brings together clients, care teams and medical doctors to exchange relevant data. This system was designed for the clinical setting and not for the homecare sector. Furthermore, it does not include clients involved relatives/support system or therapists. In the USA, the My HealtheVet portal [10] is available for Veterans by Veterans Health Administration (VHA) to renew prescriptions, organize doctor's appointments, contact health care teams, and retrieve information.

Online Care is a US-platform for clients and healthcare providers that is oriented towards GPS (or postcode) entries and enables corresponding connections. In Europe, cross-border healthcare data flow is already established, especially in the "eHealth Digital Service Infrastructure" [11]. Patient Summaries, ePrescriptions and eDispensation, are currently being rolled out. The Austrian Electronic Health Record System (ELGA, an acronym of Elektronische Gesundheitsakte) uses the same IT and technology standards, to create a feasible connection. In addition to numerous other initiatives, the EU is also striving for an "EU Health Data Space" to make available data more accessible for practical use within the healthcare system.

In consideration of all these innovative denouements the creative minds behind the project will pick out the best and combine them into the solution. At the end of development Linked Care will be a software portal that can be linked to existing systems (e.g., 'ELGA') with "one click".

III. Method

The research project started in April 2021 and will last until 2025. The project includes five end-user partners ('Akademie für Altersforschung am Haus der Barmherzigkeit', ‚Johanniter Österreich Ausbildung und Forschung gem. GmbH', ‚Wiener Rotes Kreuz- Rettungs-, Krankentransport-, Pflege- und Betreuungsgesellschaft m.b.H.', ‚Volkshilfe Gesundheits- und Soziale Dienste GmbH', and ‚Volkshilfe Wien gemeinnützige Betriebs-GmbH'), five technology partners (‚CareCenter Software GmbH', ‚Loidl-Consulting & IT Services GmbH', ‚Compugroup Medical CGM', ‚Österreichische Apotheker-Verlagsgesellschaft m.b.H', and ‚Steszgal Informationstechnologie GmbH'), as well as three scientific partners (‚Fachhochschule Technikum Wien', ‚Universität Wien', and , Fachhochschule Campus Wien - Verein zur

Förderung des Fachhochschul-, Entwicklungs- und Forschungszentrums im Süden Wiens', whereby one partner is accountable for the information privacy protection mechanisms.

All required technical functionalities of the product will be identified and described by the team of developers and the potential end-users, while implementing the user-centered design approach [12], which is already currently running. To achieve these goals, Linked Care will allow for the exchange and evaluation of care data (e.g., 'nursing', 'care', and 'therapy') via standardized interfaces, such as Integrating the Healthcare Enterprise (IHE), Health Level 7 (HL7), and Fast Healthcare Interoperability Resources (FHIR) based throughout Austria.

A mixed-method approach, supported by profound literature research, is used to identify, and approach the target groups in question as well as implement the appropriate methods for the various target groups and settings. Within the first five months of the project, focus group interviews, one-on-one interviews, research diaries, and documentary analysis are conducted. These methods have been chosen to get an encompassing insight into the needs of the end-user. The use of a participatory approach [13] enables the involvement of people and groups that are difficult to reach (e.g., '24-hour care but also very vulnerable people'). All research methods are applied low threshold and in a culturally sensitive manner. The generated data is processed using a qualitative content analysis [14].

Additional stakeholders from the regional and national environment are involved to find solutions for specific challenges. Over the course of the project, goal-oriented, cross-disciplinary, and cross-role networks of individuals and organizations will be created. Sustaining and continuing these networks after the end of the project, represents one of the long-term project goals. Experts define acceptance factors in all development steps, which are validated several times during the project in an iterative manner and therefore, influence the development process directly.

Ethical concerns take a very important part in the development of the digital communication system. Therefore, ethical reflections are being carried out in the areas of care, self-determination, security, justice, privacy, participation, and self-image. For example, in terms of justice, it is ensured that everyone has equal access to the digital documentation, and that especially vulnerable groups such as low educational classes are not excluded. Security is discussed particularly in terms of IT security, but also in terms of the privacy of data.

IV. RESULTS

In total 43 one-on-one interviews, as well as five focus group interviews including 30 participations, were carried out in the summer of 2021 with people/experts from the fields of health, business and politics who represent the healthcare providers. Additionally, various digital and paper forms of nursing documentation are currently being analyzed.

The results of the project Linked Care contribute to the expansion and further development of innovative methods in data collection and participatory product development. New knowledge tools are going to be developed, which will be available for further research issues in the field of care and support after the end of the project.

Currently, during the project, data gaps and new (data) information regarding the various stakeholders in the care sector are identified and collected. On the one hand, this data refers to the situation and needs of people in care as well as professional caregivers in the informal or formal sector. On the other hand, this data refers to the communicative interaction between the care and the medical sector. Thus, it provides a valuable basis for the development of new projects in both, the nursing, and the care sector. Furthermore, the newly developed data sets (indicators) will enable previously not used possibilities for data exchange in the care sector and for the first time allow for a standardized data exchange between general medicine practices and pharmacies as well as therapists and other stakeholders without digital data discontinuity. The comprehensive involvement of different service providers in the development of the product offers a solid foundation for the adoption of the product and ensures a high-quality exchange of relevant information. The job satisfaction of those working in the health sector may increase due to reduced documentation effort(s), duplications, and losses of information – leaving more time for the care of the clients themselves.

A. Healthcare Providers – Trends and Opinions

The results of the one-on-one interviews as well as the focus group interviews are currently being analyzed and evaluated, whereby first trends can be recorded.

1) Important interfaces: The clients' needs should be in focus and all health care professionals and providers engaged in the caring process of the client should intercommunicate, including doctors, nursing assistants, hospice, palliative care, etc. (depending on the intensity of care). Communication (also) allows for an exchange of information – the lack thereof can create insecurity and fear, which can result in further work-related errors. It is of utmost importance that health-related information is transparent and accessible to all parties involved. Therefore, it is essential that duplicate documentation is avoided to reduce time and cost resources and minimize the risk of errors. In order to guarantee a transparent information transfer and high-quality client medical treatment and care hospitals, regional health insurance funds, and pension insurance institutions (and many more) are important interfaces, and their staff members are hold on to share a client's documents and clinical documentation among each other

2) Opportunities and risks of digitalization in healthcare: All interviewed health professionals describe a very segmented documentation in the mobile nursing and care setting. None of the health professional groups mentioned (such as nurses, doctors, therapists) has

standardized access to the documentation of the other professional groups, even though they work with the same clients. This results in a large expenditure of time and resources for the collection of relevant information due to the non-access to digital documentation.

Another main finding of the survey is that clients are currently not involved in health documentation. This leads to dependencies and a great loss of information. A considerable challenge is therefore to involve the clients to strengthen their autonomy. Maintaining self-determination is an essential factor in health care, as the clients interviewed repeatedly emphasized. The documentation must be designed in such a way that the clients can influence it, for example by gaining access to the system and being able to grant permissions.

a) Firstly: Healthcare providers support the expansion of digitized information flows in the extramural area, but scepticism is omnipresent (regarding its success). Therefore, Austria (Linked Care) can learn from other countries, in terms of digital media/tools and documentation systems (e.g., 'teleconsultation', 'video documentation').

b) Secondly: A database needs to be connected to the Linked Care-solution, allowing clients to search for reliable information.

c) Thirdly: A digital tool should be added to ensure access to caring relatives and all parties involved. Also, a stored algorithm could be helpful – e.g., 'when a blood glucose meter is needed in client care, and complementary diet advice is required'.

d) Moreover: Enhanced multidisciplinary access to healthcare data can be very helpful in successful therapeutic decisions and planning, further scientific research. On top of this, data protection and confidentiality are important aspects when it comes to healthcare

3) Further aspects and possible solutions: All interviewed parties describe the relevance of a patient-related documentation, which has to be available for all technical devices, even in offline mode – documentation should not be specific to occupational groups. Community nurses are an important interface in home care. The interviewed parties clarify that this professional group should be responsible for the communication between various healthcare professionals. The analysis of the interviews also display that at this point in time there is hardly any systematic terminology stored, which doesn't allow for a presentation of nursing interventions comparably OR for effective and efficient comparison of nursing interventions.

Regarding the usability, the user interface should be clear, transparent, and not too complex in use, especially an intuitive usability of the digital system would be important. The language barrier of healthcare professionals poses a challenge – the Linked Care-solution should integrate a translation program and/or a digital dictionary.

B. Healthcare Professionals – Opinion Poll and Survey of Interests

In one-on-one and focus group interviews health professionals were asked about their opinion on the existing documentation system. They described prospects and their wishes regarding an optimal, interdisciplinary documentation system. The analysis is currently in progress. Early results show that healthcare professionals think that an interdisciplinary documentation system should include the following points:

- The digital system should integrate the planning and anamnesis, allow medical doctors to sign drug sheets and prescriptions digitally, enable physiotherapists and occupational therapists to document in the digital system, and all necessary parties involved to have access to this documentation.
- In addition, the digital documentation should be easy understandable and accessible for clients and caring relatives. Furthermore, health-related information should be automatically adopted when a client/patient is transferred from one to another organization. Above of this, risk surveys should be included in the digital program

C. Clients and their Relatives

The inclusion of affected clients and their relatives/support system, ensures the consideration of their needs and requirements in the (ongoing) product development process. This increases the quality of care considerably in two regards: Firstly, the improvement of care and nursing on a somatic level is addressed and secondly, the consistent participation of clients and their social environment also increases the quality of care according to the World Health Organization (WHO) comprehensive concept of health: Since physical as well as mental and social factors are equally included in the documentation process and clients themselves can play a decisive role during the process, they experience themselves as individuals with the power to act, who can influence their environment and living conditions. Having the power to act represents a central resource for subjective well-being according to the salutogenic approach [15].

The analysis of the one-on-one interviews with this target group is currently ongoing. At present, it can be said that each individual needs a different form of nursing, care and therapy. This should also be considered when it comes to healthcare and should be included in the nursing- and care process. Communication and a continuous exchange of information between the parties involved is particularly important – otherwise, the quality of nursing, care and therapy would suffer greatly.

D. Integrability into ELGA

The interoperability between the project solution and general practitioners' software, pharmacies, and mobile services, is novel. The link to ELGA is seminal because it is the mainly used electronic health record system in Austria. The elaboration and application of indicators in the form of a care summary is an essential part of the content-related work

and is only possible in a meaningful way by bringing different areas closer together. The increase in knowledge is possible due to the cooperation of mobile services, economy, scientific partners, and those affected.

V. Conclusion and Future Works

The Linked Care-project is innovative. For the first time, the broad use of information and communication technology (ICT) and the strong networking of different organizations and stakeholders will make the assessment of practical benefits regarding intended functionalities for the mobile care services and relevant support system possible. In turn, this will also allow for an assessment of a user-need-based design of user interfaces. The user-centered design approach, which is applied in the whole research process enables this plan.

The mentioned background literature as well as the focus group and one-on-one interviews show first significant results. These results provide a solid basis for the ongoing research within the next four years.

Acknowledgement

This research is sponsored by the Austrian research promotion agency (FFG) under grant number 884194. Project partners are: Akademie für Altersforschung am Haus der Barmherzigkeit, Johanniter Österreich Ausbildung und Forschung gem. GmbH, Wiener Rotes Kreuz- Rettungs-, Krankentransport-, Pflege- und Betreuungsgesellschaft m.b.H., Volkshilfe Gesundheits- und Soziale Dienste GmbH (GSD GmbH), Volkshilfe Wien gemeinnützige Betriebs-GmbH, CareCenter Software GmbH, Loidl-Consulting & IT Services GmbH, Compugroup Medical CGM, Österreichische Apotheker-Verlagsgesellschaft m.b.H, Steszgal Informationstechnologie GmbH, Fachhochschule Technikum Wien, Universität Wien.

References

[1] K. Muehlhauser, E. Haslinger-Baumann, T. Galanos, D. Zeidler, and F. Werner, “Linked Care - Information Transfer in Mobile Care and Support” [Conference presentation abstract]. eTELEMED 2021 The Thirteenth International Conference on eHealth, Telemedicine, and Social Medicine, Nice, France, July 2021. Available from: http://www.thinkmind.org/index.php?view=article&articleid=etelemed_2021_2_80_40109.

[2] E. Haslinger-Baumann, A. Lilgenau, C. Binder, and K. Gugenberger: “Quality characteristics in 24-hour care from the perspective of those affected. Results of a pilot survey study in Austria“. In Pflege Die wissenschaftliche Pflegezeitschrift für Pflegeberufe, 2019.

[3] B. Mittermann: “Long-term care: Where new challenges meet an old system“. Arbeit&Wirtschaft. Available from: https://www.arbeit-wirtschaft.at/langzeitpflege-herausforderungen/.

[4] M. Firgo and U. Famira-Mühlberger: “Expansion of inpatient care in the federal states. Quantitative and qualitative effects of the use of public funds compared to mobile care”. WIFO. Vienna. Available from: http://www.wifo.ac.at/wwa/pubid/47447, 2014.

[5] Care Network. Available from: https://www.carenetwork.org.uk/.

[6] Ilogs healthcare GmbH. Available from: https://www.ilogs.care/mocca/.

[7] GSD Gesellschaft für Software, Entwicklung und Datentechnik mbH. Available from: https://gsd-software.com.

[8] RECOM GmbH & Co. KG. Available from: https://docplayer.org/33810484-Leitfaden-recom-grips-v-online.html.

[9] Siemens Healthcare Diagnostics GmbH. Available from: https://www.siemens-healthineers.com/at/digital-health-solutions/digital-solutions-overview/patient-engagement-solutions/e-health-solutions.

[10] My HealtheVet. Available from: https://www.myhealth.va.gov/mhv-portal-web/home.

[11] CEF Digital Connectong Europe. Available from: https://ec.europa.eu/cefdigital/.

[12] L. Jung-Joo: “Against Method: The Portability of Method in Human-Centered Design”, Doctoral Dissertation, Aalto University, School of Arts, Design and Architecture, Department of Design, Helsinki, 2012.

[13] M. Wright, M. Block, and H. V. Unger: “Levels of Participation in Health Promotion: A Model for Assessing Participation“. In: Gesundheit Berlin (Hrsg.): Dokumentation 13. bundesweiter Kongress Armut und Gesundheit, Berlin, 2007.

[14] U. Kuckartz, “Qualitative Content Analysis. Methods, Practice, Computer Support” (4th edition). Weinheim: Beltz Juventa, 2018.

[15] A. Antonovsky, “Unraveling the mystery of health: how people manage stress and stay well”. San Francisco: Jossey-Bass, 1987.

An Effective Biometric Patient Identification System for Health Organizations

Eman Buhagiar, Clifford De Raffaele
Faculty of Science and Technology
Middlesex University
Malta
emails: buhagiareman@gmail.com, clifford.de.raffaele@gmail.com

Abstract— **Medical errors, such as patient misidentification, are the reason for around 2.6 million deaths per annum and around $42 billion in costs for health organizations in low to middle-income countries. While wristbands are the most common method for identifying patients, they can be easily misplaced and may contain missing or inaccurate information as this study shows. This may result in wrong medications and surgeries and in some instances, even preventable deaths along with liabilities for the health organizations. An in-depth literature review is conducted in this study and the current methods and process for identifying patients are also investigated, accompanied by a comparison of existing patient identification solutions, as well as issues and concerns about health data protection and privacy. Following this, the system requirements are determined through a qualitative analysis from a questionnaire distributed to different healthcare professionals. Subsequently, the effectiveness of biometric technology for patient identification through face recognition is examined. The paper finally proposes and evaluates a proof of concept with promising results for minimizing patient identification errors.**

Keywords-medical errors; health organizations; biometrics; patient identification; face recognition.

I. INTRODUCTION

Patient misidentification is a recognized worldwide problem faced by medical organizations of different types and sizes [1]–[3]. It is estimated that around 2.6 million people die each year, in just low to middle-income countries, due to medical errors [4], including errors of patient identification [5]. 9% of 7,600 (684) patient misidentification events captured in 181 different health organizations over the span of 32 months in the US led to patient harm, and in some cases, death [9]. Patient identification errors occur on different levels throughout the medical field. Various medical wards and units have been subject to such errors, including but not limited to, maternity wards, oncology centres, Intensive Care Units (ICU) and children's hospitals. In certain situations, such misidentification has led to severe consequences, one of which is the death of a patient [4][6]. According to the World Health Organization [4], between November 2003 and July 2005, the United Kingdom had 236 reported incidents related to missing wristbands or wristbands with incorrect information, the United States of America also had more than 100 similar cases reported from January 2000 to March 2003.

The National Patient Misidentification Report conducted by Ponemon Institute LLC in 2016 in the US [19] highlights the primary root causes of patient misidentification. The main three reasons include incorrect patient identification at the point of registration, time pressure when treating patients, and thirdly, lack of employee training and awareness. The report also outlines the health organization's financial impact, where the denial of claims costs the average healthcare organization $1.2 million a year. In a survey conducted by the same institute, seventy-six per cent (76%) of the respondents, who work in different types of organizations, such as large hospitals and small clinics, responded that biometrics at the patient registration point could reduce denied claims.

Patient misidentification may also lead to duplicate medical records that are time-consuming for organizations to manage and arrange [7]. An increase in insurance fraud for intentional misidentification may also be the cause of errors in patient identification. The National Health Care Anti-Fraud Association of the United States [8] estimates that the financial losses due to health care fraud are between 3-10% of the annual health care expenditure, which could lead to more than $300 billion a year. Moreover, according to the Medical Theft Alliance (MIFA), more than 2 million American citizens have been victims of medical identity theft, with cases rising each year [9].

This study analyses the effectiveness of using biometric technology for identifying patients by performing a literature review in Section II on the current problems caused by patient misidentification, followed by elucidating the process of identifying patients, current existing identification solutions, and the security and privacy issues and concerns regarding identifying patients. For the methodology in Section III, a list of system requirements is developed after distributing a questionnaire to a number of health professionals and analysing the responses. Once the requirements are documented, a system based on face recognition technology is proposed and designed in Section IV followed by its evaluation against a dataset in Section V. Finally, the results are analysed, and further improvements are suggested in Sections VI and VII.

II. LITERATURE REVIEW

Despite patient identification errors being preventable, many hospitals worldwide do not have patient identification systems implemented [13]. The first goal of The Joint Commission's National Patient Safety Goals (NPSG) for 2020 is to improve patient identification accuracy, both in hospitals and laboratories. Although the World Health Organization (WHO), the Emergency Care Research Institute (ECRI) [10], and other authors all promote the use of technology for

reducing errors in patient identification [4][10][11]. It was found by the same ECRI that technology itself was the actual cause in 15% of patient misidentification errors. One of the potential barriers to mitigating or reducing patient identification errors is the costs associated with implementing such solutions [3].

The National Patient Misidentification Report conducted by Ponemon Institute LLC in 2016 [19] determines the primary root causes of patient misidentification. The top three reasons included incorrect patient identification at the registration point of time, which may consist of placing an incorrect wristband, time pressure when treating patients, and lack of employee training and awareness. The report also outlines the health organisation's financial impact, mainly due to denial of claims, costing the average healthcare organization $1.2 million a year. Table I shows more detailed cost calculations. In a survey conducted by the same institute, seventy-six per cent (76%) of the respondents, who work in different types of organizations such as large hospitals and small clinics, responded that biometrics at the patient registration point could reduce denied claims. The report also outlined that sixty-nine per cent (69%) of the survey respondents spend more than thirty (30) minutes per shift contacting medical records or other departments to get critical information about their patients. Patient misidentification may also lead to duplicate medical records that are time-consuming for organizations to manage and fix [13]. It is therefore necessary to understand the process for patient identification in health organizations to determine any points where patient misidentification is likely to occur.

TABLE I. COST OF FAILED CLAIMS DUE TO PATIENT MISIDENTIFICATION

Cost of failed claims resulting from patient misidentification			
Step	Cost categories	Source	Average value*
A	Total billings (gross revenue) per year	AHA Financial Facts 2015	$164,300,000
B	Percentage of denied clams	Ponemon Institute survey	30.10%
C	Estimated value of denied claims per year	Calculation (A X B)	$49,454,300
D	Percentage of denied claims resulting from patient misidentification	Ponemon Institute survey	35.23%
E	Estimated value of denied claims resulting from patient misidentification	Calculation (C X D)	$17,422,750
F	Percentage of denials resulting from patient misidentification that are successfully appealed	Ponemon Institute benchmark	93.0%
G	Estimated value of denials resulting from patient misidentification that were not successfully appealed (or dropped)	Calculation (E X (1-F))	$1,219,592
*The amounts presented pertain to the average-sized registered hospital in the United States with 169 bed capacity.			

A. Patient Identification Process (PIP)

77% of wrong-patient incidents, identifying the patient was not described at all in the incident reports [14]. In a report conducted for the Australian Commission on Safety and Quality in Health Care [15], patient identification and profiling mark the beginning of a patient journey in a hospital (Figure 1). The report emphasises on the importance of providing the patient with a unique identifier that stays with the patient for the rest of the journey and other future journeys or visits. The Bay of Plenty District Health Board [16] recommends using at least three approved identifiers to identify a patient correctly. These include the patient's name, date of birth, and the National Health Index number, with the latter depending on the country's person identification system in place. A patient's bed or room number is not considered an approved identifier and should always be avoided [16], [17]. A patient's profile would include any information that can be used to confirm the patient's identity. The Australian Commission's same report [16] suggests that part of this information can be included in the wristband. It also recommends that the wristband (or a tag) should link to the patient's health record in the system.

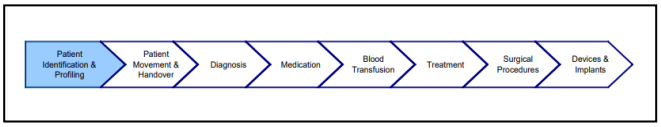


Figure 1. A patient journey in a hospital *[15]*

B. Identification Methods

While patient identification errors can be preventable [10], many hospitals worldwide do not have patient identification systems implemented [18]. The first goal of The Joint Commission's National Patient Safety Goals (NPSG) for 2020 is to improve patient identification accuracy, both in hospitals and laboratories [19]. Although many [10]–[12], [17], [18], [20] promote the use of technology for reducing errors in patient identification, it was found that technology itself was the actual cause in 15% of patient misidentification errors [10]. One of the potential barriers for mitigating or reducing patient identification errors is the costs associated with implementing such solutions [17].

1) Wristbands

Full implementation of a barcode-based Electronic Positive Patient and Specimen Identification (EPSID) system can result in a significant reduction in mislabeled specimens over three (3) years [21]. However, other studies identified wristbands as one of the leading causes of patient misidentification [22], [5]. The main issues are missing or wrong information and patients having more than one wristband. Implementing a simple wristbands system for patient identification is considered a low-cost practice for health organizations [23]. Since the simple wristbands with handwritten information on them are still prone to human error, the use of barcodes [23] or RFID [18] can reduce or mitigate patient identification errors.

Efficacy of a barcode wristband system on the prevention of medical errors indicated that the system can reduce some medical errors by an estimated 12.22% to 57.4% in different hospitals [24], [25], and medication error rate by 56% and by 47% in neonatal intensive care units [26], [27]. A barcode wristband system can help such organizations in saving

roughly $684,000 a year, from just denial of claims [28]. There are various standards and specifications for patient wristbands [15], [16], [29], [30]. The main specifications of a wristband include the size, colour, usability, method of identification, and information presentation while allowing for integrating technologies such as barcodes on the bands.

While wristbands are portable, relatively cheap, and generally easy to use, multiple problems can arise. One study concluded that 1 out of 84,000 barcode scans generated an incorrect patient identifier and as many as three (3) incorrect patient identifiers were outputted from a barcode [31]. Although this is a minimal number, these cases can still be fatal for a patient and costly for health organizations.

2) Palm Vein Pattern Recognition

Palm vein scanning is a widespread method of verifying and authenticating a user [32]. Given that each patient's palm vein pattern is unique and very stable over the person's lifetime, it makes this method the most commonly used successful technology for identifying people [33]. A palm vein scanner uses a near-infrared light wave to capture the user's vein pattern on the palm. In contrast with other recognition methods, palm veins have internal features making it almost impossible to reproduce with fake palms [34].

The stages of palm print authentication include acquiring the palm vein image, enhancement, extraction, matching, and authentication [32]. As for developing a palm vein pattern image, only specific blood flow patterns are considered for the sake of image clarity. The three methods for capturing vascular images are X-Rays, Ultrasonic Images (Ultrasound), and Infrared Imaging. The latter is the preferred method because of its non-invasive contactless, and nonharmful technique. While there are two types of Infrared technologies that can be used, Far-Infrared and Near-Infrared, the latter is used as it is less expensive to operate and is able to capture smaller veins, making it adequate for identification. However, Far-Infrared technology can capture thermal patterns that are unique even to identical twins [35].

While taking into consideration the accuracy of the palm vein scanning method, it is worth noting that this method is more costly when compared to the barcode wristband alternative. This is due to its unique software in addition to the installation and the implementation of the palm scanners. This form of method is also considered to be more intrusive for a patient as it may raise palm image storage security concerns. However, when compared to the fingerprint or face recognition methods, the palm vein scanners are favorable within this regard. Another issue worth considering is the matter of hygiene as when comparing methods, a noncontact method would be ideal, examples of this include barcode scanning and face recognition.

3) Ocular-based Identification

Two types of ocular-based identification technologies used to identify a person uniquely are iris and retinal scanning. The retina is the thin tissue located at the back of the eyeball, containing cells sensitive to light. It is composed of a complex structure of capillaries that supply the retina with blood and therefore, every person's retina is unique. Similar to palm vein pattern recognition, a retinal scan would map a person's retina's unique patterns. The iris is a thin circular structure behind the cornea of the eye, which is responsible for controlling the size of the pupils and, therefore, the amount of light reaching the retina. The complexity of the retina patterns makes it unique for every person. Unlike iris or palm vein scanning, retinal scanning uses camera technology with little infrared illumination to capture the retina's intricate structures' images.

Iris recognition method would be ideal in a health organization environment as it does not require proximity to a camera for a successful scan and uses safer low-energy infrared lighting. Moreover, retina scan accuracy may be affected by certain diseases [36] and iris scanning proved to be the most secure patient identification method in UCSD's Moore Cancer Center when implemented [37].

4) Face Recognition

Face recognition can be described as determining the identity of an individual based on the person's facial features. The challenge of facial recognition in its simplest form involves comparing two face images and deciphering if they are of the same person [38]. A more significant challenge arises when faces exhibit changes in appearance due to make-up, facial hair, and accessories, such as jewellery.

The process of identifying a face through a face recognition system is similar to that of iris recognition. The steps involved include acquiring the face image, the face detection, recognition, and identification [39]. During the face detection phase, an algorithm is used to do corrections, skin segmentation, and facial feature extraction from the digital face image. One of these algorithms is the Viola-Jones Algorithm, which is considered the first-ever real-time face detection system [40]. In the next stage of face recognition, the modified face image from the previous phase is classified to identify the person from a database. Different algorithms, which include FeedForward Neural Network (FFNN) [39], and Local Binary Pattern (LBP) [41] are used here as well.

One implementation of facial recognition with Microsoft Kinect v2 sensor for patient verification proved to be over 96% accurate [5]. However, each scan took around thirty (30) seconds to complete, a time frame that is unsatisfactory for a healthcare environment, but this can be classified as a limitation to the technology used, Microsoft Kinect v2, as other studies showed promising results in terms of performance, with time reduced to 100ms with the same level of accuracy [27][28].

C. Security and Privacy

The security and privacy areas in patient identification are habitually overlooked [29]. Privacy is also a significant concern for the patients themselves [46], and implementing a biometric system for improving patient identification accuracy is known to impose more privacy concerns for the patient [47].

1) Health and Data Breaches

According to a report issued by McAfee [48], a stolen health record would generally sell more than financial data on the black market. This is mainly because health data does not have that many established markets like financial data. Another study conducted by Infosec Institute (2015), shows that there was a 73% increase in cyberattacks between 2013 and 2014 targeted to healthcare organizations and that the average cost of a stolen health record amounted to $363 on the black market compared to $1 - $2 of the stolen credit card information. Health data breaches tripled in a year between 2017 and 2018 and there were over 15 million patient records breached in 2018 in the United States [49].

One of the most common causes of insider-related breaches is family member snooping [49], that is, healthcare workers spying on their family members. This cause amounted to around 67% of the breach cases, while the second most common type of breach was snooping on their co-workers, amounting to approximately 15% of the violations. Insiders, which are the healthcare workers, are also more likely to commit another breach after their first violation, as 51% of the offences are repeated.

2) Privacy

Storing and processing patients' personal and sensitive data calls for strict privacy protection measures to minimize patient privacy issues as much as possible. Biometrics privacy can be interlinked with personal privacy, given that our biometric information can uniquely identify us [50]. Various studies address different patient privacy concerns and implications [32][36]. In some cases where biometric technology is in place, patients refused to be subject to such technology due to privacy and confidentiality concerns [12]. Some biometric technologies proved to have a high acceptability rate, such as face recognition and voice recognition [5]. In contrast, others, such as iris and retina scanning [52], had a lower acceptability rate. Other studies however showed that biometric technologies are less or non-invasive than traditional methods of identification [5].

While there are no legislations covering the usage of biometric identification systems [53], and yet the right of privacy is considered a fundamental human right [54], safeguards must be set down for every step, from collection to retention of the data collected. Individuals must be given rights to access, correct and delete their data [50]. Furthermore, individuals should be assigned the ability to opt out, so biometric technologies should not be the only implementation for identification.

3. Security

Biometric technologies can help in identifying patients accurately and provide the right authorization and authentication or verification for accessing and amending medical records [38], [55]. The user asserts an identity for confirmation, and the biometric system confirms if the assertion is genuine. This process is generally used to prevent unauthorized access to a system or services. Verification can be explained formally using (1), where, given a claimed identity I and a query feature set x^A, the decision if (I, x^A) belongs to the 'genuine' or 'impostor' class needs to be taken. If x_I^E is the stored template that corresponds to the identity I, x^A is compared with x_I^E and a score s is matched, which measures the similarity between $x^A$and x^E, and η would be a predefined threshold.

$$(I, x^A) \in \begin{cases} genuine, \ if \ s \geq \eta, \\ impostor, if \ s < \eta, \end{cases} \quad (1)$$

However, biometric solutions can have their security flaws as well [41][42]. The biometric system's integrity is determined by its ability to guarantee non-repudiable authentication, that is, ensuring that a user who accesses a specific resource cannot later deny in using it. There are four major classes of security threats to biometric systems [6][26] and these are Denial of Service (DoS), Intrusion, Repudiation and Function Creep. Although it is much harder for an impostor to forge biometric traits than hacking traditional passwords, there are studies suggesting the use of multimodal biometric systems where multiple types of biometric features would be measured and compared, for example, fingerprints and face, for better accuracy [43][44].

The goal of this paper is to provide a proof of concept of the most favoured method of biometric patient identification determined through the methodology and evaluate its results against an already established dataset. Limitations and possible improvements are then suggested.

III. METHODOLOGY

A questionnaire was developed and distributed to professional healthcare participants using purposeful sampling, and its results are analysed. Consequently, requirements are determined, documented and validated using a House of Quality matrix and system designs are proposed.

A. Research Instrument

The questionnaire developed and conducted was sectioned into four (4) main sections:

- **Background Information** - gathering brief, non-personal information about the stakeholder, including their profession, roles, and practical experience.
- **The Problem** - capturing the stakeholder's awareness of patient misidentification and its consequences, globally and in the organization in which they practice in.
- **Their Process of Identifying a Patient in their Organization** – gathering information about the current process that health professionals use to identify patients. They were asked to explain the process briefly and what identifiers are used and at what point. They were also asked of awareness of any of the patient identification methods mentioned

earlier in this study and which of them are used in their organization, if any. Finally, they were asked to provide feedback on their current patient identification method, and if they think that it can be improved and on what aspects, such as accuracy, security and cost.

- **The Solution(s)** - participants were asked which patient identification methods they would implement in their organization, how would they prioritize them and why. They were also invited to prioritize the characteristics and the concerns of a biometric patient identification system in terms of security, accuracy, and efficiency.

B. Participants

Participants chosen that successfully answered all the questions which were provided to them amounted to nine (9), and these were staff nurses (2), an Accident and Emergency (A&E) nurse, a doctor, a general practitioner (GP), a urology surgeon, physiotherapists (2), and a speech-language pathologist. All the participants work in local health organizations in Malta. The GP owns, manages, and works in a small private healthcare clinic. The nurses, the doctor, and the urology surgeon work in a national hospital, while the physiotherapists and the speech-language pathologist work in smaller private healthcare organizations.

C. Results Analysis

All the respondents think that their current patient identification system works moderately well (67%) or very well (33%) (Fig. 2). Hence, participants were also asked to identify the vital positive characteristics of their current patient identification system, and these included the cost, where 58% rated it as very well, followed by ease of use and efficiency (17%), and patient's comfort in using it, where 30% of the participants classified it as very well, as shown in Figure 1. Security was the least rated, with 33% rating it as just slightly well. 25% of the respondents classified security as an aspect that needs to be improved in their current system, along with accuracy. Moreover, 89% of the participants said that, currently, it takes less than 15 seconds to identify a patient, with 33% of them stating that it even takes less than 5 seconds. Therefore, essential requirements that needed to remain there are the system's cost, ease of use, patient's comfort in using it, and efficiency (processing speed). On the other hand, other aspects that require improvements are security and accuracy.

D. Requirements

One of the most commonly used methods to achieve a standard view of the relationship between customer requirements and product design is Quality Function Deployment (QFD) [60]. QFD is a product development methodology that gives importance to the customer's opinions throughout the development process. QFD was used in this study to determine the list of important requirements for the proposed system. Customer importance ratings for system requirements were calculated based on the results obtained

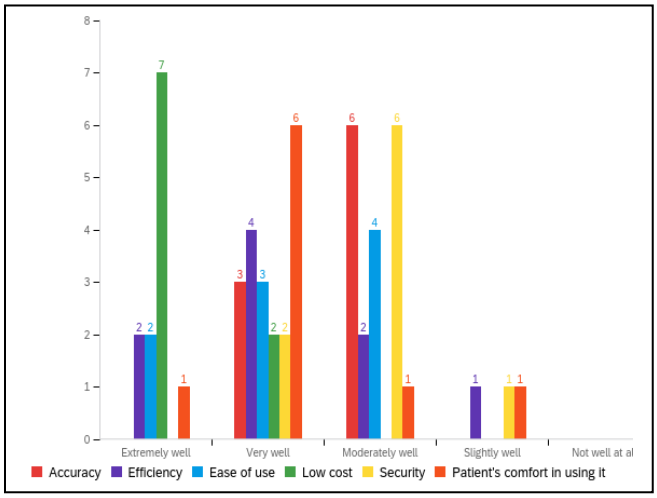


Figure 2. Question results for ranking aspects of the system

from the questionnaire distributed to health professionals in the order of Accuracy, Efficiency, Security, Ease of Use, Cost and Patient's Comfort. On the other hand, the requirements concluded from the previous section were listed on the other side of the matrix. The conditions that scored the highest importance ratings were found to be *Cost*, *Use of Secondary Identification Methods*, *Accuracy*, *Availability*, and *The Use of Alternative Non-Biometric Identification Methods*.

With the cost being the topmost essential requirement for the customer, any negative correlations related to this requirement should be addressed and ideally eliminated as soon as possible. Therefore, alternative non-biometric identification methods should be kept to a minimum and only used in cases where the patient refuses to use other biometric methods, for example. Patients will most probably opt for these alternative methods if they have trust concerns about the system, and hence the importance of *Transparency*. Each customer should be as transparent as possible to the patients about the biometric system, ensuring no physical harm will be done and securing their data safety while pointing out the benefits of such techniques for their own good. We must remember that using alternative non-biometric systems may negatively impact user training, the effort of operating the system, and identification accuracy.

IV. Design

High-level and low-level designs of the system and integration with the possible current systems are proposed. Furthermore, designs of the proposed mobile application are also portrayed together with data and process flows. Taking Systems Theory into perspective, the proposed system would have biometric information as an input and after biometric processing and communication with the Patient Medical Record System (MRS) or Database, outputs the patient information.

For patient identification, the app user needs to be authorised and authenticated. The app should display multiple authentication choices, including but not limited to Face

Recognition, Fingerprint Recognition, or a user account. Data collected at this stage is transferred securely to an internal API where it is processed. Through in-house or third-party APIs, if needed, roles and permissions are determined and set, and the user is then allowed to proceed and identify a patient on the app.

Provided the user is authenticated and authorized on the mobile application, the user can identify a patient, which has already been registered before, that is, the patient's biometric data required for identification has been securely stored on a database or service. After the patient has given consent and the biometric data is collected, such as a face photo, this data is sent through a secure and encrypted channel, such as HTTPS, to an internal API, which communicates to third party APIs, such as Microsoft Facial Recognition, and handles the identification and the fetching of patient information and medical records if needed to be sent back to the app so that vital patient information can be displayed.

The users can be provided with different options to authenticate themselves (Fig. 4a). Viable options include face or fingerprint recognition or a user account.

a) Face Recognition - A

A similar process used to identify patients through face biometrics can be used here to identify and authorize a user.

b) Fingerprint - B

Most of the modern smartphones are equipped with an inbuilt fingerprint sensor. This may be used to authenticate the user. However, this may require more development effort to implement.

c) User account - C

A user account is also another option, although less preferred since it is more time-consuming to develop and maintain. The integration of Microsoft Azure AD will help in improving performance if Microsoft Face Recognition is used for patient identification.

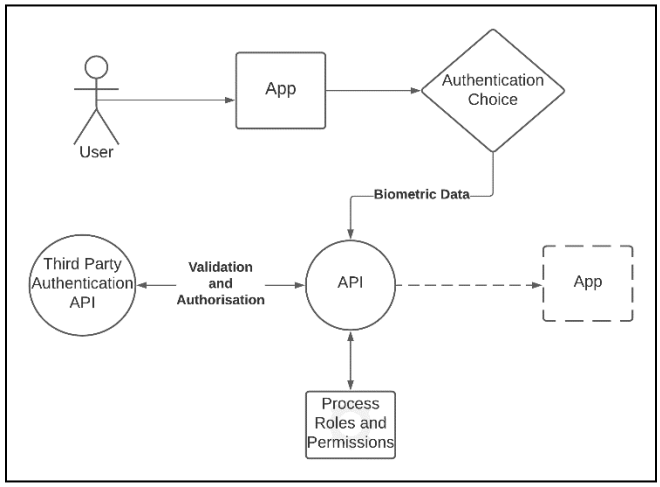


Figure 3. High-Level Authorisation Flow

Figure 3 shows the proposed flow of the mobile application used to identify patients by their face. Adopting two identifiers for identifying a patient, as suggested by the WHO [2], the first stage includes scanning the barcode or QR code printed on the wristband wearing the patient. The app should immediately display the camera preview after successful authentication, for the user to scan the barcode (Fig. 4b). The barcode should be recognised very quickly, and the Patient Identifier stored on the barcode or QR code is captured by the app. Once a barcode or QR code is successfully captured, the user should be prompted to capture the patient's biometric data. For this study, the method of face recognition using Microsoft Face Recognition is showcased. Therefore, the user is asked to take a photo of the patient's face, as straightforward as possible. The user should confirm the image taken for the identification process to initiate.

Upon identification completion, if succeeded, the user is prompted with a pop-up dialog asking to confirm the identification details, to ensure the identification and update biometrics or to scan again (Fig. 4c). Updating biometrics would send the last patient's face photo to Microsoft Face Recognition API and is added to the patient's list of faces for AI training.

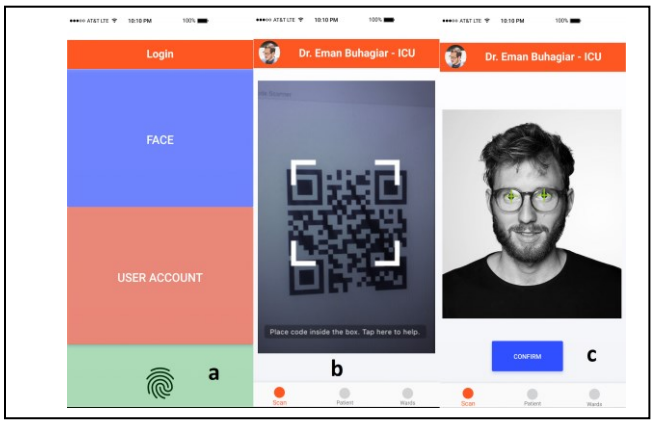


Figure 4. Proposed app system designs. a – Authentication, b – Barcode/QR code scanning, c – Identification confirmation

V. EVALUATION

The primary identification method of face recognition is implemented using Microsoft Cognitive Services and their Face API [61]. There are various reasons for opting for Microsoft Cognitive Services, and these reasons all cohere with the system requirements established earlier. Evaluation of the proposed system was divided into three stages:

A. Applicability

Microsoft Face API is a seamless, secure and an easy to integrate and operate API for face detection, emotion recognition, and identification. Microsoft Face API can be utilised in different scenarios, such as user authentication and counting people in a crowd.

1) Face Detection

Face Detection can detect up to 100 faces in an image along with different attributes such as age, position, smile, emotion, facial hair, makeup and occlusions, such as masks and bandanas when a photo or image URL is passed as a parameter. No images are stored, but only the landmarks are stored, which cannot be used on their own to identify a person.

2) Face Recognition

Face Identification compares face landmarks previously stored on the API from adding faces to a person in a person group or large person group to an input image face landmarks. The API returns a confidence level (1-10) for the user to decide if the prediction is up to the user's expectations or not. The API also accepts a confidence threshold as a parameter to filter out results based on the user's preference for confidence. For example, in identifying patients, a confidence level below 80% (0.8) might not be acceptable, and therefore, a confidence threshold of 0.8 or greater should be passed.

3) Security

One of the most critical concerns addressed by health professionals in the questionnaire conducted for this study is the system's security. Microsoft ensures security by firstly not storing any actual face images on their servers, and secondly by encrypting any data stored using FIPS 140-2 compliant 256-bit AES encryption. FIPS 140-2 is a U.S. Government computer security standard used to approve cryptographic modules [62].

4) Cost

As for the cost of usage, the standard version allows for up to 10 transactions per second, with €0.506 per 1,000 transactions for 5 to 10 million transactions and €0.338 per 1,000 transactions for transactions amounting to more than 100 million. As for storage, €0.009 per 1,000 faces per month is charged. A transaction constitutes an API call, apart from the training calls where a transaction counts for every 1,000 images trained. Table II shows a detailed pricing scheme for Microsoft Face API.

TABLE II. MICROSOFT FACE API PRICING

Version	Transactions	Price
Free	20 per **minute**	30,000 transactions free per month
Standard	10 per **second**	0-1M transactions - €0.844*
		1M-5M transactions - €0.675*
		5M-100M transactions - €0.506*
		100M+ transactions - €0.338*
		€0.009 per 1,000 faces stored per month

5) Limitations

Like all other face recognition methods, there are some limitations that may hinder the system's accuracy. Various face occlusions, such as masks and makeup, or face injuries and ageing, may prevent face recognition algorithms from detecting or identifying a face. While many face recognition technologies cater for occlusions measurement when detecting a face, face masks during the COVID-19 pandemic impacted face recognition algorithms' overall accuracy [63]. Such a situation requires the need for alternative identification methods, such as retina recognition, to be available in the system.

Figure 5. Face image under different lighting conditions

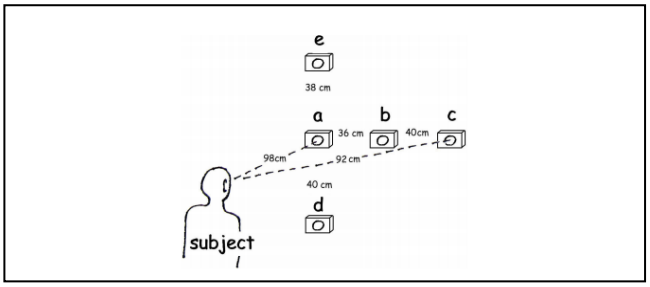


Figure 6. Angles of the camera from which the photos were taken [63]

B. Accuracy Evaluation

A dataset of multiple faces was used to evaluate a person's identification accuracy through Microsoft Face API. The dataset was introduced in an another study [64] to provide more diversity than the existing publicly available datasets regarding lighting, age, and ethnicity. The dataset consists of 3755 faces, totalling to 276 participants in all. Each participant has at least eight (8) face photos, each from a different angle or different lighting (Fig. 5, Fig. 6).

The image filenames have the form of i000qa-fn.jpg, where:

- i is the prefix of all files,
- 000 is the subject identification number,
- q is the lighting type, ranging from q to z,
- a is the camera angle, ranging from a to e,
- f for female or m for male,
- n for no glasses or g for with glasses.

This naming convention was used as wildcards in code during the evaluation process, as explained later on in this sub-chapter. This dataset was chosen to be used in this study as it fits nicely into healthcare scenarios for identifying different patients.

The following 'setup' process was adopted to evaluate the identification accuracy of Microsoft Face API:

1. A large person group with a name 'test' was created, and the returned *largePersonGroupId* was stored, to be used later to identify a face.
2. For each of the subjects in the dataset:

a. The subject was added to the large person group just created, and the identification number was used as the name (for example '000').

b. The face was added to the person using the first image file of the subject.

3. After all subjects are added, the large person group was trained by calling the train API endpoint.

Once all the subjects in the dataset were registered, the following identification process was conducted for each participant:

- The face was detected, and the *faceId* returned was stored.
- The face is identified, passing the *faceId* in the request body and the *largePersonGroupId* captured earlier when creating the group. If identified, a list of potential candidates should be returned, each with a *personId* and a confidence level.
- The person was identified and confirmed by getting the person in the large person group by the *personId* captured in the previous step. The person name and the file identification number were compared, and if these matched, identification was successful.

Figure 7 portraits the identification evaluation flow.

C. Performance Evaluation

The second stage of evaluation ensures that the second most crucial requirement established, efficiency is maintained throughout the identification process. For this, a simple mobile application was developed, simulating a patient's identification using two identifiers, a barcode and a face. Once this is done, the app prompts the user to take a photo of a face, and this is sent to Microsoft Face API upon confirmation for identification. Both the barcode key and the person identified from the API are compared, and if matched, a call to a database is made to fetch the records of the patient. The whole process was timed for efficiency evaluation.

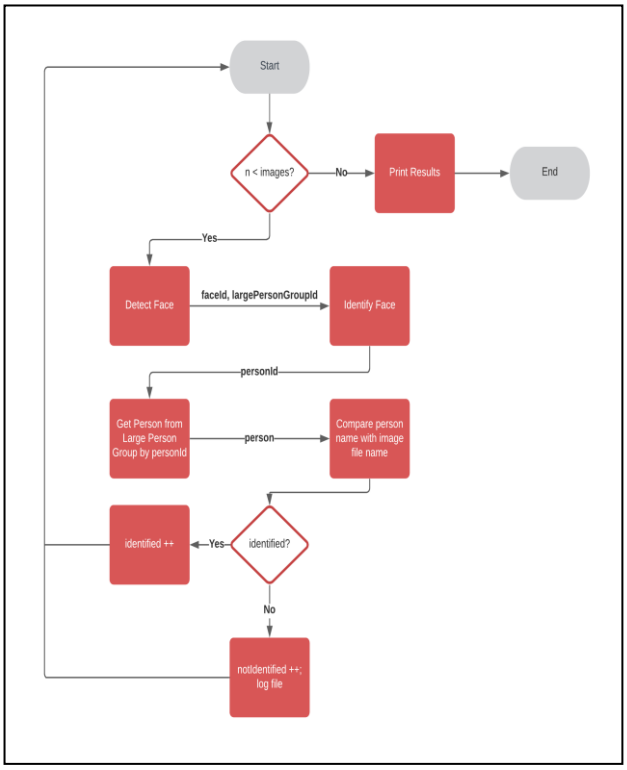


Figure 7. Identification Evaluation Flowchart

VI. Results and Discussion

The results of the previously explained evaluation processes are analysed and discussed in detail. There are specific scenarios where the system performed very well, but there are others in which accuracy was challenged, and possible improvements are suggested for these cases.

A. Accuracy of Face Recognition

There were eight (8) types of datasets that were used to assess the accuracy of the proposed system (Fig. 7). These will be referenced as *a*, *b*, *c*, *d*, *e* and *qa*, *rb*, *sb*. *a* – *e* represent the different angles of the camera from which the photo was taken, while *q*, *r* and *s* represent the amount of light exposed to the face, with $s > r > q$.

The thresholds used for evaluating the accuracy of the system ranged from 0.97 to 0.92. Anything above the threshold of 0.97 resulted in less than 80% accuracy, which is not considered safe enough for such a critical system. On the other hand, any threshold below 0.92 always resulted in 100% accuracy in all scenarios tested.

1) Case 1 – Camera Angles

In the first case, the first set of faces *qa*, that is, photos taken from in front of the person (*a*) and with lighting set *q*, was added to the API and trained. After that, all the other photos from the remaining angles were tested for identification against different thresholds.

As seen in Table III, the accuracy results obtained by training just *qa* were always above 88% for 276 photos. It can be noted that angles *b* and *c* resulted in less accuracy than angles *d* and *e*. Therefore, side angles seem to be less accurate than front angles. Angles *c* and *d* provided the same accuracy results, while angle *d* proved to be the best angle for obtaining accurate results.

TABLE III. CASE 1 RESULTS

Thr. / Sc.	0.97	0.96	0.95	0.94
b	88.04	98.55	99.28	100.00
c	88.04	98.55	99.28	100.00
d	99.28	100.00	100.00	100.00
e	93.48	97.83	98.91	100.00

n = 276

2) Case 2 – Lighting

In the second case, with trained set *qa*, all the other remaining photos with r and s lighting exposed to them were tested for identification with different thresholds.

In this case, accuracy suffered much more when different lighting was used on the person's face. As shown in Table IV, the accuracy went down to 43.96% and 39.56% from 88.04 and above 99% from the previous case. This indicates that lighting has a significant effect on identifying a person from their face, and a less threshold of 0.92 compared to 0.94 had to be used for achieving 100% accuracy on the 91 photos tested. Lighting set *r* performed better than set *s* significantly. Significant changes were also noted when the threshold was changed each time by 0.01, with accuracy changes of more than 30% in some cases.

Since this case resulted in low accuracy results in some scenarios, set *ra* was added to the API and trained, and set *sb* was tested again. The same angle of the previously trained set was used (*a*) for consistency. This was done to note the difference in accuracy and the effectiveness of training. Table IV shows the accuracy results of set *rb* when tested, while sets *qa* and *ra* are trained already.

Accuracy improved significantly for set *sb* when *ra* was added and then trained. With 0.97 as the threshold, accuracy improved by more than 24% and by more than 17% for the 0.96 threshold (Table V). This shows that dataset training provided by Microsoft Face API does improve identification accuracy.

B. Integration Efficiency Test

For this case, a simple mobile application was developed to showcase the use of the proposed system by the users. The app communicates with a custom developed API hosted on Microsoft Azure, which then communicates to Microsoft Face API and a database with records of patients, also hosted on Microsoft Azure. The process took between 5 to 7 seconds when timed in code, with full-bar Wi-Fi connectivity, to detect and identify the face through the API, and to get the patient's allergies and conditions list from a sample database. This result coheres with the efficiency requirement established earlier for identification to take not more than 15 seconds and ideally not more than 5 seconds.

TABLE IV. CASE 2 RESULTS

Thr. / Sc.	0.97	0.96	0.95	0.94	0.93	0.92
rb	43.96	73.63	90.11	96.70	100.00	100.00
sb	39.56	64.84	82.42	93.41	98.54	100.00

n = 91 (photos 000 - 090)

TABLE V. CASE 3 RESULTS

Thr. / Sc	0.97	0.96	0.95	0.94	0.93
sb	64.84	82.42	92.31	96.70	100.00

n = 91 (photos 000 - 090)

VII. CONCLUSION AND FUTURE WORK

We recognise that patient misidentification is a known global problem in various healthcare organizations, and it can lead to further complications to the patients and the organizations themselves. While biometric technology is applied in multiple sectors, such as authentication and security, payroll, and banking, there are fewer studies on the application of biometric technology for positive patient identification. This study conducted a questionnaire among different health professionals to determine the top concerns for implementing a biometric system in healthcare. These included security, accuracy, cost, and patient cooperation. While most of the participants were aware of some of the biometric methods for patient identification, none of them has ever made use of any of them but would consider in doing so, given a better accuracy rate and robust security. The quantitative analysis obtained from the questionnaire helped in determining and prioritising the proposed system requirements, although a further study can be conducted with a more extensive questionnaire and more participants.

For the proof of concept, this study evaluated the implementation of face recognition biometric technology for identifying patients, as this was the most preferred biometric

method chosen by the questionnaire participants. Microsoft Face API was used as the third-party provider for identifying faces, and the proposed system was evaluated against its accuracy and efficiency, among other requirements determined. While results were promising with over 80% accuracy in most cases, this technology seemed to lack in identifying faces with occlusions, such as different lighting. When more than one face photo from different angles and different lighting are registered and trained, accuracy was improved significantly.

As for future works, the system needs to be evaluated against a larger dataset with a larger variety of face occlusions to mimic real-case scenarios in health organizations. Further studies on the security aspects of the system are also important to be conducted to minimise the risks of malicious attacks on the system and gain more confidence from the system users.

REFERENCES

[1] E. Buhagiar and C. De Raffaele, "An Effective Biometric Patient Identification System for Health Organizations," in eTELEMED 2021, The Thirteenth International Conference on eHealth, Telemedicine, and Social Medicine, 2021, pp. 62–70, doi: 978-1-61208-872-3.

[2] E. Buhagiar, "Implementing an Effective Biometric Patient Identification System in a Health Organization," Middlesex University Malta, 2021.

[3] S. Kelly, "The patient misidentification crisis," Health Manag. Technol., pp. 12–14, 2016.

[4] WHO, "WHO calls for urgent action to reduce patient harm in healthcare," Saudi Med. J., vol. 40, no. 10, pp. 1075–1076, 2019, Accessed: Apr. 18, 2021. [Online]. Available: https://www.who.int/news/item/13-09-2019-who-calls-for-urgent-action-to-reduce-patient-harm-in-healthcare.

[5] G. Lippi, L. Chiozza, C. Mattiuzzi, and M. Plebani, "Patient and Sample Identification. out of the Maze?," J. Med. Biochem., vol. 36, no. 2, pp. 107–112, 2017, doi: 10.1515/jomb-2017-0003.

[6] M. Jonas, S. Solangasenathirajan, and D. Hett, "Patient Identification, A Review of the Use of Biometrics in the ICU," 2014, pp. 679–688.

[7] T. H. Tase, E. R. S. Quadrado, and D. M. R. Tronchin, "Evaluation of the risk of misidentification of women in a public maternity hospital," Rev. Bras. Enferm., vol. 71, no. 1, pp. 120–125, 2018, doi: 10.1590/0034-7167-2017-0134.

[8] NHCAA, "The Challenge of Health Care Fraud - The NHCAA," 2015. https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud.aspx (accessed Aug. 04, 2020).

[9] Ponemon Institute, "Fifth Annual Study on Medical Identity Theft," no. February, 2015, [Online]. Available: http://medidfraud.org/wp-content/uploads/2015/02/2014_Medical_ID_Theft_Study1.pdf.

[10] ECRI Institute, "Patient identification errors," 2016. doi: 10.1016/j.enfcli.2011.07.006.

[11] M. A. Yahiro, "Biometrics Solutions in e-Health Security," Spine (Phila. Pa. 1976)., vol. 19, no. Supplement, pp. 2274S-2278S, 2015, doi: 10.1097/00007632-199410151-00004.

[12] N. Anne et al., "Feasibility and acceptability of an iris biometric system for unique patient identification in routine HIV services in Kenya," Int. J. Med. Inform., vol. 133, no. September 2019, p. 104006, 2020, doi: 10.1016/j.ijmedinf.2019.104006.

[13] M. A. McClellan, "Duplicate medical records: a survey of Twin Cities healthcare organizations.," AMIA Annu. Symp. Proc., vol. 2009, pp. 421–425, 2009.

[14] M. Härkänen, M. Tiainen, and K. Haatainen, "Wrong-patient incidents during medication administrations," Journal of Clinical Nursing, vol. 27, no. 3–4. pp. 715–724, 2018, doi: 10.1111/jocn.14021.

[15] S. Allworth, P. Lapsley, J. Kelly, D. Martin, and V. J. Kelly, "Technology Solutions to Patient Misidentification Report of Review Final," no. October, 2008.

[16] Bay of Plenty District Health Board, "Patient Identifiction Standards," p. 50061, 2016.

[17] World Health Organization, "Patient Identification," J. Nurs. Care Qual., vol. 1, no. 2, 2007, doi: 10.1097/00001786-200301000-00010.

[18] A. Aguilar, W. Van Der Putten, and G. Maguire, "Positive Patient Identification using RFID and Wireless Networks," Undefined, 2006.

[19] The Joint Comission, "National Patient Safety Goals Effective July 2020 for the Hospital Program Goal," Jt. Com., no. July, p. 14, 2020, [Online]. Available: https://www.jointcommission.org/-/media/tjc/documents/standards/national-patient-safety-goals/2020/npsg_chapter_hap_jul2020.pdf.

[20] S. Kelly, "The patient misidentifi cation crisis," Health Manag. Technol., pp. 12–14, 2016.

[21] P. D. Hain et al., "An intervention to decrease patient identification band errors in a Children's Hospital," Qual. Saf. Heal. Care, vol. 19, no. 3, pp. 244–247, 2010, doi: 10.1136/qshc.2008.030288.

[22] S. W. Renner, "Wristband Errors in Small Hospitals," vol. 28, no. 3, 1997.

[23] L. V. Hoffmeister and G. M. S. S. De Moura, "Use of identification wristbands among patients receiving inpatient treatment in a teaching hospital," Rev. Lat. Am. Enfermagem, vol. 23, no. 1, pp. 36–43, 2015, doi: 10.1590/0104-1169.0144.2522.

[24] M. Khammarnia, A. Kassani, and M. Eslahi, "The efficacy of patients' wristband bar-code on prevention of medical errors: A meta-analysis study," Appl. Clin. Inform., vol. 6, no. 4, pp. 716–727, 2015, doi: 10.4338/ACI-2015-06-R-0077.

[25] B. Wegerbauer, “Can barcoded wristbands improve patient safety ?,” no. January, pp. 1–5, 2007.

[26] N. Dwibedi et al., “Effect of bar-code-assisted medication administration on nurses’ activities in an intensive care unit: A time-motion study,” Am. J. Heal. Pharm., vol. 68, no. 11, pp. 1026–1031, 2011, doi: 10.2146/ajhp100382.

[27] F. H. Morriss et al., “Effectiveness of a Barcode Medication Administration System in Reducing Preventable Adverse Drug Events in a Neonatal Intensive Care Unit: A Prospective Cohort Study,” J. Pediatr., vol. 154, no. 3, 2009, doi: 10.1016/j.jpeds.2008.08.025.

[28] Ponemon Institute LLC, “2016 National Patient Misidentification Report Independently conducted by Ponemon Institute LLC Sponsored by Imprivata,” no. December, 2016.

[29] Australian Commission for Safety and Quality in Health Care, “Specifications for a standard patient identification band,” pp. 3–4, 2007, [Online]. Available: http://www.safetyandquality.gov.au/wp-content/uploads/2012/02/FactSheet-PatID-Band.pdf.

[30] The Joint Commission, “Patient identification policy,” Patient Saf. Solut., vol. 1, no. May, pp. 1–26, 2011.

[31] M. L. Snyder, A. Carter, K. Jenkins, and C. R. Fantz, “Patient misidentifications caused by errors in standard bar code technology,” Clin. Chem., vol. 56, no. 10, pp. 1554–1560, 2010, doi: 10.1373/clinchem.2010.150094.

[32] O. Akinsowon, B. Alese, and O. Adewale, “Infrared Capture of Palm-Vein Blood Vessel Patterns for Human Authentication,” J. Internet Technol. Secur. Trans., vol. 3, no. 1, pp. 203–209, 2014, doi: 10.20533/jitst.2046.3723.2014.0027.

[33] Imprivata, “Improving Patient Care with Positive Patient Identification - White Papers - HealthITAnalytics,” 2015.

[34] H. Setiawan and E. M. Yuniarno, “Biometric Recognition Based on Palm Vein Image Using Learning Vector Quantization,” Proc. 2017 5th Int. Conf. Instrumentation, Commun. Inf. Technol. Biomed. Eng. ICICI-BME 2017, no. June, pp. 95–99, 2018, doi: 10.1109/ICICI-BME.2017.8537770.

[35] D. C.Lakshmi, A.Kandaswamy, and C.Vimal, “Protection of Patient Identity and Privacy Using Vascular Biometrics,” Int. J. Secur., vol. 4, 2010.

[36] M. D. Abràmoff, M. K. Garvin, and M. Sonka, “Retinal Imaging and Image Analysis,” IEEE Rev Biomed Eng, pp. 169–208, 2010, doi: 10.1109/RBME.2010.2084567.Retinal.

[37] B. N. Haile, “The Eyes Have It : Iris Biometrics Safely Identify UCSD Patients for Radiation Oncology Treatment,” 2010.

[38] A. K. Jain, A. A. Ross, and K. Nandakumar, Introduction to Biometrics. Boston, MA: Springer US, 2011.

[39] A. Gurel, C., & Erden, “Design of a Face Recognition System,” 15th Int. Conf. Mach. Des. Prod., vol. 1, no. 2012, 2012.

[40] Y.-Q. Wang, “An Analysis of the Viola-Jones Face Detection Algorithm,” Image Process. Line, vol. 4, pp. 128–148, 2014, doi: 10.5201/ipol.2014.104.

[41] S. M. Bah and F. Ming, “An improved face recognition algorithm and its application in attendance management system,” Array, vol. 5, no. November 2019, p. 100014, 2020, doi: 10.1016/j.array.2019.100014.

[42] R. Ranjan et al., “A Fast and Accurate System for Face Detection, Identification, and Verification,” IEEE Trans. Biometrics, Behav. Identity Sci., vol. 1, no. 2, pp. 82–96, 2019, doi: 10.1109/tbiom.2019.2908436.

[43] S. Anila and N. Devarajan, “Simple and Fast Face Detection System Based on Edges,” Researchgate.Net, no. May, 2010, [Online]. Available: https://www.researchgate.net/profile/Anila_Satish/publication/225292501_Simple_and_Fast_Face_Detection_System_Based_on_Edges/links/09e414fd75a23d2c1b000000.pdf.

[44] Hembroff, G. C., Wang, X., and Muftic, S., “Providing an Additional Factor for Patient Identification Based on Digital Fingerprint,” *2nd {USENIX} Workshop on Health Security and Privacy (HealthSec 11)*, 2011.

[45] E. B. Heinlein, “Medical records security,” Comput. Secur., vol. 15, no. 2, pp. 100–102, 1996, doi: 10.1016/0167-4048(96)89322-9.

[46] P. Houston, “Research Responses to Patient Privacy Concerns,” 2017.

[47] C. L. Parks and K. L. Monson, “Automated Facial Recognition of Computed Tomography-Derived Facial Images: Patient Privacy Implications,” J. Digit. Imaging, vol. 30, no. 2, pp. 204–214, 2017, doi: 10.1007/s10278-016-9932-7.

[48] C. Beek, C. McFarland, and R. Samani, “Health Warning Report,” 2018.

[49] Protenus Inc., “Protenus 2019 Breach Barometer,” pp. 1–21, 2019, [Online]. Available: https://email.protenus.com/hubfs/Breach_Barometer/2018/2019 Breach Barometer Annual Report.pdf.

[50] C. U. Ebelogu, O. Adelaiye, and F. Silas, “Privacy Concerns in Biometrics,” no. July, 2019.

[51] D. Birnbaum, K. Gretsinger, M. G. Antonio, E. Loewen, and P. Lacroix, “Revisiting public health informatics: patient privacy concerns,” Int. J. Heal. Gov., vol. 23, no. 2, pp. 149–159, 2018, doi: 10.1108/IJHG-11-2017-0058.

[52] N. Dahiya and C. Kant, “Biometrics security concerns,” Proc. - 2012 2nd Int. Conf. Adv. Comput. Commun. Technol. ACCT 2012, pp. 297–302, 2012, doi: 10.1109/ACCT.2012.36.

[53] R. Gellman, “Privacy and Biometric ID Systems : An Approach Using Fair Information CGD Policy Paper 028

August 2013," Cent. Glob. Dev., no. August, 2013.

[54] A. Puri, "Privacy is a fundamental human right," pp. 1–3, 2013.

[55] A. Dandashi and W. Karam, "Biometrics security and experiments on face recognition algorithms," 2012 IEEE Symp. Comput. Intell. Secur. Def. Appl. CISDA 2012, 2012, doi: 10.1109/CISDA.2012.6291532.

[56] S. O. Olatinwo, O. Shoewu, and O. O. Omitola, "Iris Recognition Technology : Implementation , Application , and Security," Pacific J. Sci. Technol., vol. 14, no. 2, pp. 228–233, 2013.

[57] J. C. Mazura, K. Juluru, J. J. Chen, T. A. Morgan, M. John, and E. L. Siegel, "Facial recognition software success rates for the identification of 3D surface reconstructed facial images: Implications for patient privacy and security," J. Digit. Imaging, vol. 25, no. 3, pp. 347–351, 2012, doi: 10.1007/s10278-011-9429-3.

[58] K. Sasidhar, V. L. Kakulapati, K. Ramakrishna, and K. KailasaRao, "Multimodal Biometric Systems - Study to Improve Accuracy and Performance," Int. J. Comput. Sci. Eng. Surv., vol. 1, no. 2, pp. 54–61, 2010, doi: 10.5121/ijcses.2010.1205.

[59] W. Dahea, "Multimodal biometric system : A review Multimodal biometric system : A review," no. November, pp. 25–31, 2018, doi: 10.13140/RG.2.2.34056.65287.

[60] R. J. Hauser, A. Griffin, L. R. Klein, M. G. Katz, and P. S. Gaskin, "Quality function deployment," 2010, doi: 10.4271/870272.

[61] Microsoft, "Facial Recognition | Microsoft Azure," 2020. https://azure.microsoft.com/en-us/services/cognitive-services/face/ (accessed Dec. 08, 2020).

[62] FIPS 140-2, "FIPS 140-2 Change Notices (12-03-2002) Federal Information Processing Standards Publication (Supercedes FIPS PUB 140-1," FIPS PUB 140-2, 2001.

[63] M. Ngan, P. Grother, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) - Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms," 2020.

[64] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," Proc. Pattern Recognit. Assoc. South Africa, pp. 32–34, 2010.

Regulating Interoception Through Low Frequency Mechanical Dermal Stimulation to Improve Sleep

Gina Sensale, Sahithi Garikapati, Angelina Distefano, Jean Toher, Hanna Villa, Sean Hagberg
Feelmore Labs, Inc.
Brooklyn, NY, USA
e-mail: gina@feelmorelabs.com, sahithi@feelmorelabs.com, gina.distefano@feelmorelabs.com, jmtoher@gmail.com, hannakvr@gmail.com, sean@feelmorelabs.com

Abstract— **More than 50 million adults living in the United States suffer from disordered sleep. Yet few safe, effective, drug-free interventions are available. In this 30-day open-label home study, participants (n=25) reporting poor sleep were recruited to test a novel wearable mechanical stimulation device. The device is designed to modulate the interoceptive network by producing gentle, slow mechanical stimulation. After using the device each night before bed for 30 days, significant improvements in sleep quality were reported. Additionally, participants reported improvements across multiple dimensions of interoception as measured by the Multidimensional Assessment for Interoceptive Awareness. On average, participants (n=22) reported a 43% improvement in the overall quality of their sleep, measured by the Pittsburgh Sleep Quality Index. Participants (n=15) contacted 3 to 7 months after completing the study, maintained improvements in sleep quality and interoceptive regulation. These findings indicate that mechanical stimulation may offer an effective, safe, non-drug alternative to improving sleep via interoceptive regulation and suggest a novel approach to treatment.**

Keywords-sleep; neurostimulation; interoception; affect; c-tactile afferents.

I. INTRODUCTION

The term 'interoception,' coined in 1906, initially described the total afferent input of the viscera to the brain [2]. The concept was mostly dormant until decades later when it evolved and expanded to include all afferent signaling to the central nervous system [3]. Most responses to interoceptive signals are understood to be autonomic, rarely rising to the level of consciousness (e.g., retracting a hand back when touching something hot), while others only indirectly reach awareness (e.g., filtration action of the kidneys only requires a response to the urge to void) [4]. Interoceptive signals such as those associated with emotion, can be mistakenly regarded as instinctual or autonomic; however, they are generally learned habits that remain malleable [5][6].

Methods to assess interoceptive capacity in individuals (e.g., heart rate detection tasks) have been developed, as well as self-report measures to assess various aspects of interoception [7][8][9]. Lower interoceptive awareness (i.e., lack of response to interoceptive signals) and dysregulated interoception (i.e., misinterpretation of the interoceptive signals) are associated with impaired decision making, poor sleep quality, increased psychiatric disorders, and reduced empathy [10] [11]. With that considered, dysregulated interoception is often recognized as a common feature of many affective and somatic disorder and their symptoms, prompting researchers to investigate neural bases of interoception [12]. Studies have shown that areas in the brain such as the anterior insula (IA) and anterior cingulate cortex (ACC), which are essential for processing emotion, affect, and behavior, are also proving to be fundamental in interoceptive processing [17][18]. Multiple studies suggest that modulating activity in these areas through practices such as mindfulness and meditation, appear to improve interoception [19][20][21]. Similarly, the same practices reduce symptomology among many affective and somatic disorders [22][23][24].

Disciplines of the mind and/or body, such as meditation, or yoga, or physical exercise, can improve interoceptive regulation because these activities stimulate the same areas of the brain as those dedicated to specific inputs from the periphery [12][13][15][22]. That specific interoceptive network is comprised of C-tactile afferent (CTAs) nerves [25]. CTAs are dermal mechanoreceptors, found only in hairy skin of mammals. They respond to a narrow range of slow, light touch stimuli, such as the kind of touch seen between a parent and child, the comforting caress of a partner, the grooming of baboons, and similar social behaviors [26]. This type of touch, called affective touch, also modulates areas of the brain involved in interoception, emotional regulation, and related functioning [27][28]. In turn, affective touch is associated with reduced stress, increased relaxation, a sense of belonging, and increased empathy [29][30][31].

1) Interoception and Sleep

Interoceptive signals such as those relating to temperature or pain, have been found to significantly impact sleep. For example, there is strong evidence in support of the relationship between internal body temperature and sleep regulation, where studies have shown that environmental temperature can alter sleep regulation [11] [32]. Similarly, as interoception includes processing nociceptive information, the present literature reveals a strong relationship between pain and sleep [11] [33]. Evidence of the relationship between pain and sleep is often observed in clinical populations such as chronic

pain patients, where one study found that 53% of chronic pain patients reported also suffering from insomnia as well, compared to only 3% of the healthy population [34]. As discussed earlier, dysregulated interoceptive processing can produce a wide range of symptoms, often overlapping in different clinical populations, which may explain the relationship seen among disorders of sleep and chronic pain, as well as with disorders of affect [5][10][35][36][37].

Neuroscience research, both electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) research, has provided evidence that cortical regions play a role in particular aspects of sleep, help regulate sleep-wake patterns, and are similar to those prominent in regulating the interoceptive system of interest here [38]. The ACC, a brain region found to stimulate wakefulness, AI, and orbitofrontal cortex have all been studied for their roles in insomnia disorder, a highly prevalent sleep disorder associated with impaired functioning during waking hours [38][39][40][41]. For instance, functional neuroimaging studies have found that those with insomnia have an increased insula coactivation with cortical networks associated with salience and arousal, such as the ACC, when compared to healthy controls [38][42]. Interestingly, these same brain regions are also recognized for their involvement in interoceptive processing and affect [10]. Insomnia disorder has also been associated with heightened interoceptive awareness, see [11] for review. When interoception is dysregulated, heightened interoceptive awareness may lead to a hyperawareness of physiological states that becomes disruptive to sleep, especially when falling asleep or staying asleep, as discussed earlier.

The relationship between interoception and sleep quality has been observed in a study investigating the relationship between components of subjective sleep quality, measured by the Pittsburgh Sleep Quality Index (PSQI), and dimensions of interoception, measured by the Multidimensional Assessment for Interoceptive Awareness (MAIA-II) [43]. In that sample of young adults (n=545), ages 18 to 25 years old, 90% were considerably poor sleepers, measured by Global PSQI scores., Poor sleep quality, was found to be significantly negatively correlated with dimensions of interoception including non-distracting and trusting, with the strong effect sizes [43]. These results suggest that those who experience better quality sleep may be more adept at recognizing sensations such as pain and discomfort, and more trusting of the internal bodily information regarding various physiological states, which is consistent with previous work investigating sleep and pain tolerance, as well as subjective hunger [43]. Overall, these findings provide evidence that there is a relationship between perceivably improved sleep and improved interoception. Therefore, considering the overlap of brain regions and cortical networks involved in both interoceptive processing and sleep, as well as the relationship between interoception, disordered sleep, and subjective sleep quality, one may argue that interoceptive processes are salient to at least some sleep disturbances.

2) Impact of Disordered Sleep

Disordered sleep, including insomnia, obstructive sleep apnea, and hypersomnia, impacts between 50 and 70 million Americans each year and an estimated 83.6 million U.S. adults get below the recommended 7 hours of sleep per night [44][45]. Roughly 15% of the U.S. population report insomnia related symptoms, and a quarter of the population are dissatisfied with their sleep. One study found that one in five adult females in Australia experience chronic insomnia, and one in seven experience obstructive sleep apnea [46].

The prevalence of disordered sleep can have a profound impact on both societal and individual levels. For example, insufficient and low-quality sleep decreases workplace productivity, increases the risk of motor vehicle accidents, and elevates the number of risk-factors for health outcomes. Poor-quality sleep and sleep deprivation have also been linked to metabolic dysfunction, cognitive impairment, alterations in neuroendocrine function and affective dysregulation [47][48]. Additionally, disordered sleep has been linked to a reduced quality of life and increases in anxiety related disorders [49].

In the primary care setting, patients complaining of sleep dysfunction have become a common occurrence and many people around the country have turned to available treatments. The most common agents of treatment include prescription drugs, such as benzodiazepines and antidepressants, and over the counter options, such as antihistamines and melatonin [50]. Between 1998 and 2006 the number of Americans reliant on sleep aid prescriptions tripled and roughly 4% of the adult population reported use of a prescription sleep aid [44]. A review of 35 studies relating to the effectiveness of melatonin found that most of the studies were inconclusive and melatonin was rarely superior to placebo [51]. Several of the reviewed studies gave weak recommendations for the use of melatonin among healthy volunteers or those with a history of insomnia due to melatonin's lack of effectiveness in initiating sleep or improving the quality of sleep [51]. Prescription drug such as. benzodiazepines leave users with a "hungover" feeling and often result in daytime drowsiness, impaired memory, difficulty concentrating, lack of coordination, amnesia, dizziness, and many other side effects that impact day-to-day activities [52]. Other prescription options, such as antidepressants, can lead to increases in suicidal ideation and mania [49]. The lack of reliable treatment options suggests a need for an alternative approach to a significant public health issue.

Considering the apparent relationship between sleep and interoception, as well as the research showing how practices that improve interoception (e.g., mindfulness,

yoga, and extreme sports/intense exercise), can also improve affective and somatic symptomology, we hypothesized that improving interoception will also improve sleep [22][23][24][53]. To do so, we developed a mechanical stimulation technology targeting CTAs (low intensity and low frequency (~5-15Hz) stimulation) [54]. This technology offers a non-invasive means of activating the affective touch pathway by targeting specific mechanoreceptors in the skin, and in turn, modulating interoceptive processing [1]. The objective of this research was to assess the effects, if any, of a device designed to improve interoceptive regulation via CTAs on symptoms of sleep disorder measured by the Pittsburgh Sleep Quality Index (PSQI) [55].

The rest of this paper is organized as follows. Section II details the nature of our novel approach to improve sleep through regulating interoception. Section III outlines the study method, including a description of the participant sample and eligibility criteria, the device technology, and measures used to assess sleep quality and interoceptive characteristics. Section IV details all study procedures, including a statement of ethics, participant eligibility screening, participant responsibilities, and visit procedures. Section V contains the detailed results of the present study and Section VI closes out the paper with the conclusion and a discussion of future work.

II. State of the Art

Most sleep interventions, aside from known drug or comorbid causes, treat sleep as a primary disorder. Our interest, as in other work, is to approach sleep disturbances as a manifestation of dysregulated interoception and others have done preliminary work looking at that pathway [32] [43] [56]. While some interventions for sleep, such as mindfulness, are also known to improve interoceptive regulation [53], few interventions appear to make the explicit connection, with at least one exception [57]. More specifically, no study, to our knowledge, explicitly approaches sleep as a function of dysregulated interoception and further, none use mechanical stimulation of the affective touch pathway to improve interoceptive regulation.

III. Study Method

This section details the overall method for the 30-day study, including the recruitment of qualifying participants and exclusion criteria, description of our novel mechanical stimulation device, and the measures used to assess sleep quality and characteristics of interoception.

A. Participants

Participants for this trial included healthy adults (mean age=34) with self-reported poor sleep. The subject population included a distribution of female (n=13) and male (n=9) participants. Poor sleep was measured using the PSQI (Global Score >10) [55].

Exclusion criteria included the use of sleep medication, current psychiatric disorder, a skin condition that may be exacerbated with device use, a metal implant in the head/neck region and finally if participants were pregnant or breastfeeding. These exclusion criteria aimed to control for any comorbid psychopathology and for any variance arising from interactions with other drugs.

B. Mechanical Stimulation Device

A simple headband with small piezoelectric actuators at the distal ends, seen in Figure 1 (a), was developed to deliver short bursts of very low intensity, low frequency mechanical stimulation, targeting CTAs associated with the affective touch pathway. The actuators were positioned behind the ears, on the mastoid, for convenience. The specific wave form was derived from a combination of empirical study (changes in alpha power pre/post stimulation in in-lab studies over 2 years) and known response characteristics of the CTA mechanoreceptors (low intensity, low frequency ~10 Hz).

C. Measures

The subsections below describe the measures used to assess sleep, including both behavioral and subjective measures, as well a subjective measure to assess various aspects of interoception.

1) Behavioral Sleep

A daily sleep diary with a 1- item sleep rating scale was used to track sleep quality throughout subject's participation in the study.

A Garmin Vivosmart 4 (https://buy.garmin.com/en-US/US/p/605739) watch was used to measure sleep duration during the study. The Garmin watch uses a wrist-based Pulse Ox sensor to measure sleep activity. Although the watch reports detailed sleep stages as well, that was excluded from this study due to the lack of reliability in those assessments [58][unpublished internal testing].

2) Subjective Sleep

Changes in sleep quality were assessed via PSQI and a 1-Item Sleep Quality Rating Scale. The Pittsburgh Sleep Quality Index (PSQI) is a self-report measure assessing sleep quality and disturbances over a 1-month timeframe. It is an 18-item questionnaire with seven component scores (range 0–3) that result in a global sleep quality score (range 0–21). Scores of five or greater indicate a probable sleep disorder [55]. The seven component measures include subjective sleep quality (PSQI_C1), sleep latency (PSQI_C2), sleep duration (PSQI_C3), habitual sleep efficiency (PSQI_C4), sleep disturbances (PSQI_C5), use of sleep medication (PSQI_C6) and daytime drowsiness (PSQI_C7).

The 1-Item Sleep Quality Rating Scale is a self-report measure utilizing a scale of 1 to 5, where 1 represents little to no sleep at all, and 5 represents great sleep (no problems falling or staying asleep).

An exit interview was conducted by a study coordinator at the end of the study to assess the usability of the device.

3) Affective/Interoception Characteristics

The Multidimensional Assessment of Interoceptive Accuracy (MAIA) is a 32-item self-reported measure of interoception (scale of 0 Never – 5 Always) that tracks changes in interoceptive awareness across eight dimensions [7]. The eight dimensions are Noticing (MAIA_MN), Not-Distracting (MAIA_MND), Not-Worrying (MAIA_MNW), Attention Regulation (MAIA_MAR), Emotional Awareness (MAIA_MER), Self-Regulation (MAIA_MSR), Body Listening (MAIA_MBL), and Trust (MAIA_MT).

IV. Study Procedure

Subjects were enrolled in the sleep protocol for 30 days and instructed to use the device at home at least once a day, within one hour of their regular bedtime. Written informed consent was obtained from all participants. During enrollment, subjects were provided with device instructions, device calibration and their first stimulation session. A subset of participants (n=13) wore a wrist device to track sleep (Garmin VivoSmart 4). Subjects were compensated up to $300 at the end of the study in the form of a gift card. Refer to Figure 1 (b) below for a visualization of the 30-day study protocol. All study procedures were reviewed and approved by Solutions IRB (#: FWA00021831) [59].

A. Screening

Participants were recruited for the study via digital ads on Facebook, as well as flyers posted around the community. Interested subjects were then screened for the use of sleep medication, and participants that did not use any sleep medication were then invited to fill out an application via Google Forms. All applications were reviewed and those subjects meeting the inclusion criteria (n=39) were invited to schedule their enrollment appointments, of whom 25 participants (14 females, 11 males) were enrolled into the study, displayed in Figure 2.

B. Study Visits

Eligible participants were enrolled via a written informed consent in the lab. They completed a set of baseline questionnaires, which included demographic information, subjective sleep measures and interoception measures. Upon completion of these assessments, all participants were instructed on device use and went through calibration to find the lowest level at which they could perceive the stimulation. Participants then had their first 20-minute stimulation session in the lab to ensure proper training and use of the device. Participants in the cohort with the Garmin sleep tracker were given additional instruction.

Participants returned to the lab after 30 days of home use for the final assessment of sleep and interoception measures and an exit interview was conducted with a study coordinator to discuss usability of the device.

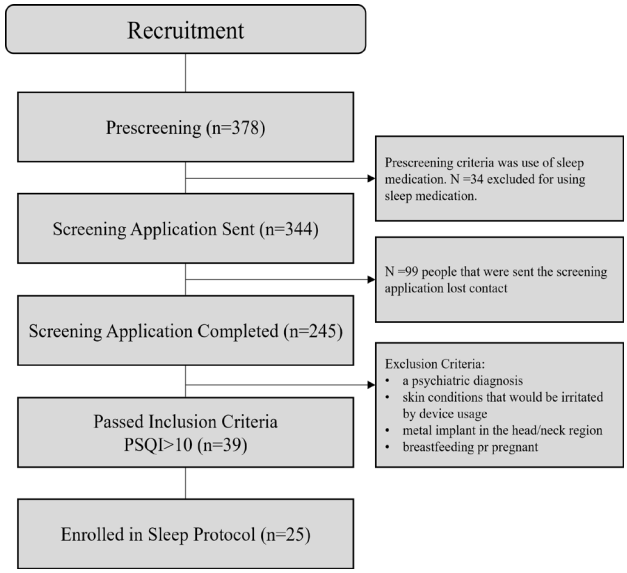


Figure 1: Participant Recruitment and Screening Procedure.

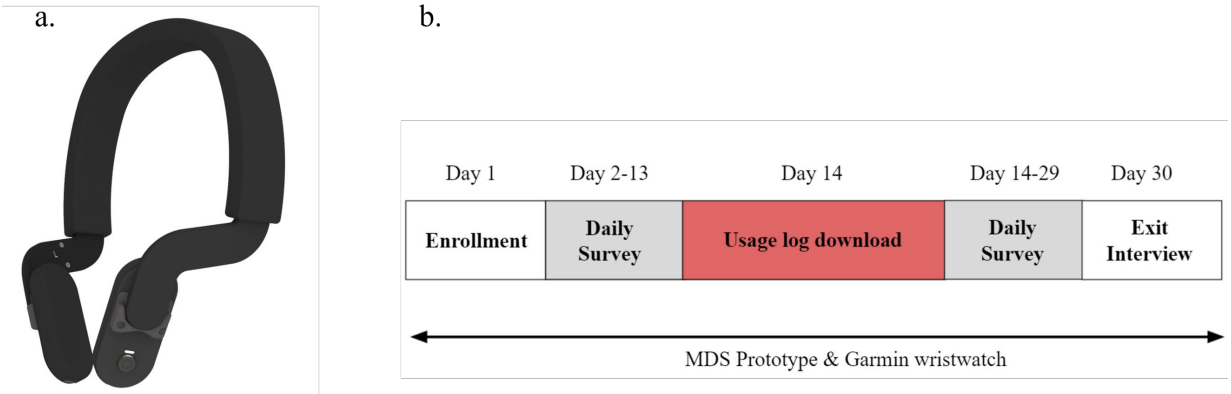


Figure 2: (a) Mechanical Dermal Stimulation Device Prototype (b) 30-Day Study Protocol.

C. Protocol

All participants were instructed to use the stimulation device for 30 days at home within one hour of their regular bedtime. Participants received a text message each morning with a link to fill in their daily sleep diary via Google Forms. Participants in the Garmin tracker cohort were asked to wear the Garmin watch each day and night during the 30-day study, and data from the watch was collected at the end of the study.

D. Analysis

Data analysis was performed on MATLAB and Microsoft Excel. Due to the small sample size of the study, t tests were conducted, and p values (with a 95% confidence interval) are reported as outcome measures in the following section.

V. Results

This section outlines the detailed results of the 30-day study, including descriptive characteristics of the participant sample, observed changes in sleep quality via self-report measures and a wearable sleep tracking device, as well as changes in aspects of interoception.

A. Descriptive Characteristics of Participants

In the sample of 25 participants, 3 were excluded from the primary analysis due to lack of compliance with the 30-day study protocol (i.e., device usage, completing the study). More specifically, those that used the device less than 33% of the time (based on usage 1x/day for 30 days) were considered noncompliant and removed from analysis. The results discussed in this section will be reported as mean ± standard deviation. The characteristics of the sample of compliant participants (n=22) are displayed in Table 1. In short, the sample (n=22) included 13 females and 9 males with ages ranging from 24 to 60 years old (34.27 ± 9.64).

B. Sleep Characteristics

After using the MDS device before bed for 30 days, 86% of participants (n=22) reported an overall improvement in sleep, measured by Global PSQI scores seen in Figure 3 (a), where lower scores indicate improvement in symptoms. Global PSQI scores improved significantly by 43% on average (Pre: 9.77 ± 3.04, Post: 5.18 ± 2.58, p=0.00), displayed in Table 2. In terms of individual characteristics of sleep measured by the PSQI (PSQI component scores), we observed significant improvements in Subjective Sleep Quality (p=0.00), Sleep Latency (p=0.00), Sleep Disturbances (p=0.04), and Daytime Drowsiness (p=0.00), shown in Table 2. On average, Subjective Sleep Quality improved by 67% (Pre: 2.0 ± 0.53, Post: 0.64 ± 0.58), Sleep Latency improved by 51% (Pre: 1.32 ± 0.48, Post: 1.05 ± 0.38), Sleep Disturbances improved by 14% (Pre: 1.86 ± 0.89, Post: 0.86 ± 0.64) and Daytime Drowsiness improved by 45% (Pre: 1.86 ± 0.89, Post: 0.86 ± 0.64), seen in Figure 3 (a). Similarly, we saw significant improvements in sleep quality ratings after 30 days of device use (p=0.00), displayed in Table 2. On average, sleep quality ratings increased by 45% (Pre: 3.09 ± 0.53, Post: 0.64 ± 0.58), shown in Figure 3 (b) (c), where higher scores indicate better sleep quality.

Sleep hours were assessed via a commercial wrist Photoplethysmography (PPG) device (Garmin VivoSmart 4) for a subset of compliant participants (n=13). This PPG device was chosen as it was most reliable for sleep time based on earlier studies, where we found that no devices to our knowledge appeared to accurately measure sleep stages, nor is PPG a suitable substitute for a hypnogram. After 30 days of using the MDS device, sleep hours increased by 65 minutes on average (Pre: 6.60 ± 0.29, Post: 7.45 ± 0.45), shown in Figure 3 (d).

Table 1: Compliant Study Sample Demographics

Demographics	N = 22
Age	34.27 ± 9.64
Gender	
Female	13
Male	9
Ethnicity	
Hispanic	2
White / Caucasian	7
Black or African American	3
Prefer not to say	1
Asian / Pacific Islander	7
Hispanic and Middle-Eastern American	1
Mixed Race	1

Table 2: Compliant Study Sample Results

	Pre	**Post**		
	Mean ± SD	*Mean ± SD*	*pvalue*	*tstat*
Sleep Rating	3.09 ± 0.75	4.27 ± 0.63	**0.000001**	-5.654
PSQI_C1	2 ± 0.53	0.64 ± 0.58	**0.000000**	8.101
PSQI_C2	2.05 ± 0.72	0.95 ± 0.9	**0.000064**	4.439
PSQI_C3	1.32 ± 1.09	0.82 ± 0.96	0.112858	1.619
PSQI_C4	0.82 ± 0.91	0.64 ± 0.9	0.508631	0.667
PSQI_C5	1.32 ± 0.48	1.05 ± 0.38	**0.040966**	2.109
PSQI_C6	0.41 ± 0.73	0.23 ± 0.53	0.351170	0.943
PSQI_C7	1.86 ± 0.89	0.86 ± 0.64	**0.000105**	4.283
PSQI_Global	9.77 ± 3.04	5.18 ± 2.58	**0.000003**	5.406
MAIA_MN	3.34 ± 0.79	3.8 ± 0.84	0.070784	-1.854
MAIA_MND	2.18 ± 1.04	2.05 ± 1.15	0.681956	0.413
MAIA_MNW	2.61 ± 1.45	2.62 ± 1.16	0.969703	-0.038
MAIA_MAR	2.69 ± 0.91	2.98 ± 0.96	0.307309	-1.033
MAIA_MEA	3.69 ± 1.14	3.8 ± 1.17	0.756368	-0.312
MAIA_MSR	2.76 ± 1.01	2.91 ± 1.15	0.653277	-0.452
MAIA_MBL	2.47 ± 1.07	3.02 ± 1.16	0.111566	-1.625
MAIA_MT	3 ± 1.42	3.23 ± 1.04	0.547755	-0.606

C. *Affective/Interoception Characteristics*

After 30 days of using the MDS device, on average participants (n=22) reported an overall improvement in interoception, measured by the MAIA, displayed in Table 2. Most notably, participants improved in the following dimensions of interoception: Noticing (Pre: 3.34 ± 0.79, Post: 3.8 ± 0.84), Not Worrying (Pre: 2.61 ± 1.45, Post: 2.62 ± 1.16), Attention Regulation (Pre: 2.69 ± 0.91, Post: 2.98 ± 0.96), Body Listening (Pre: 2.47 ± 1.07, Post: 3.02 ± 1.16), and Trusting (Pre: 3 ± 1.42, Post: 3.23 ± 1.04), shown in Figure 3 (e).

D. *Follow Up Results*

Approximately 3 to 7 months following the 30-day study, a subgroup of compliant participants (n=15) completed follow-up assessments. The follow-up sample (n=15) included 8 females and 7 males, with ages ranging from 24 to 44 years old (31.73 ± 5.71), shown in Table 3.

Despite discontinued use of the MDS device for 3 to 7 months, participants (n=15) maintained some improvements in sleep from their 30-day study participation. When compared to baseline scores, participants reported significant improvements in overall sleep quality (p=0.03), represented by Global PSQI scores displayed in Table 4. On average, Global PSQI scores at follow up improved by 26% when compared to baseline (Pre: 9.93 ± 3.15, Post: 5.4 ± 2.8, Follow Up: 7.07 ± 3.75), where lower scores represent improved sleep, displayed in Figure 4(a). Apart from Global PSQI scores, significant improvements were also found in Subjective Sleep Quality (p=0.04), and Sleep Duration (p=0.00) when comparing follow-up PSQI component scores to baseline, displayed in Table 4. On average, Subjective Sleep Quality improved by 22% (Pre: 2.07 ± 0.59, Post: 0.67 ± 0.62, Follow Up: 1.53 ± 0.74) and Sleep Duration improved by 53% (Pre: 1.27 ± 1.16, Post: 0.93 ± 1.03, Follow Up: 0.2± 0.41), and although not significant, Daytime Drowsiness improved by 22% (Pre: 2.07 ± 0.88, Post: 0.87 ± 0.64, Follow Up: 1.47± 1.13), shown in Figure 4(a). Similarly, on a scale of 1 to 10, where higher scores indicate better sleep, sleep quality ratings did not return to baseline, shown in Table 4. On average, participants (n=15) reported an improvement in sleep by 12% at follow up when compared to baseline (Pre: 3.07 ± 0.8, Post: 4.33 ± 0.62, Follow Up: 3.27± 0.96), displayed in Figure 4(b).

Overall, the improvements reported in interoception following 30 days of MDS device use were maintained following cessation of the MDS device for 3 to 7 months, shown in Table 4. Most notably, participants (n=15) maintained an improvement in not distracting, attentional regulation, self-regulation, body listening, and trusting dimensions of interoception, measured by the MAIA. On average, participants reported a 28% improvement in Not Distracting (Pre: 2.36 ± 1.16, Post: 2.13 ± 1.19, Follow Up: 2.41 ± 1.05), 20% improvement in Attentional Regulation (Pre: 2.7 ± 0.86, Post: 3.22 ± 0.99, Follow Up: 2.96 ± 1.03), a 28% improvement in Self-Regulation (Pre: 2.83 ± 1.18, Post: 3.03 ± 1.28, Follow Up: 3.03 ± 1.23), a 38% improvement in Body Listening (Pre: 2.53 ± 1.15, Post: 2.98 ± 1.16, Follow Up: 3.08 ± 0.57), and 32% improvement in Trusting (Pre: 2.89 ± 1.45, Post: 3.09 ± 1.07, Follow Up: 3.17 ± 1.23), displayed in Figure 4(c).

Table 3: Follow Up Sample Demographics

Demographics	N = 15
Age	31.73 ± 5.71
Gender	
Female	8
Male	7
Ethnicity	
Hispanic	2
Caucasian	4
African American	2
Prefer not to say	1
Asian / Pacific Islander	4
Mixed Race	2

Table 4: Follow Up Results

	Pre	**Post**	**Follow-up**	**Pre vs Follow Up**
	Mean ± SD	*Mean ± SD*	*Mean ± SD*	*pvalue*
Sleep Rating	3.07 ± 0.8	4.33 ± 0.62	3.27 ± 0.96	0.540
PSQI_C1	**2.07 ± 0.59**	**0.67 ± 0.62**	**1.53 ± 0.74**	**0.039**
PSQI_C2	2.07 ± 0.8	0.93 ± 0.88	1.6 ± 0.83	0.127
PSQI_C3	**1.27 ± 1.16**	**0.93 ± 1.03**	**0.2 ± 0.41**	**0.002**
PSQI_C4	0.8 ± 0.94	0.8 ± 1.01	1.07 ± 1.28	0.521
PSQI_C5	1.33 ± 0.49	1.07 ± 0.26	1.07 ± 0.59	0.190
PSQI_C6	0.33 ± 0.72	0.2 ± 0.56	0.33 ± 0.82	1.000
PSQI_C7	2.07 ± 0.88	0.87 ± 0.64	1.47 ± 1.13	0.116
PSQI_Global	**9.93 ± 3.15**	**5.4 ± 2.8**	**7.07 ± 3.75**	**0.031**
MAIA_MN	3.3 ± 0.76	3.87 ± 0.94	3.33 ± 0.88	0.930
MAIA_MND	2.36 ± 1.16	2.13 ± 1.19	2.41 ± 1.05	0.889
MAIA_MNW	2.91 ± 1.23	2.82 ± 0.95	2.48 ± 1.08	0.320
MAIA_MAR	2.7 ± 0.86	3.22 ± 0.99	2.95 ± 1.03	0.473
MAIA_MEA	3.6 ± 1.21	3.61 ± 1.31	3.63 ± 1.09	0.950
MAIA_MSR	2.83 ± 1.18	3.03 ± 1.28	3.03 ± 1.23	0.653
MAIA_MBL	2.53 ± 1.15	2.98 ± 1.16	3.08 ± 0.57	0.109
MAIA_MT	2.89 ± 1.45	3.09 ± 1.07	3.17 ± 1.23	0.567

VI. Conclusion and Future Work

To our knowledge, this is the first human study to evaluate mechanical stimulation of the affective touch pathway in a sleep-disordered population. Although the trial is small, open-label, and used early prototypes, the results were significant, and participants benefited from the use of the MDS device [1]. Overall sleep quality, represented by Global PSQI scores, significantly improved after using the MDS device for 30 days. Participants reported falling asleep faster, experiencing fewer sleep disturbances, and experiencing less daytime drowsiness. Participants also reported an overall improvement in

aspects of interoception. These findings suggest that mechanical stimulation of CTAs can lead to improvements in sleep. There is a need for further research to better understand the relationship between sleep, sleep disorders and interoception, and more robust clinical trials to assess the effects of mechanical stimulation on sleep and other symptoms of dysregulated interoception.

A confirmatory Randomized Control Trial (RCT) is underway to address limitations such as sample size and inclusion of a control group, which will be completed in late 2021.

An abundance of sleep research has provided evidence suggesting that impaired sleep has potentially serious implications on a wide variety of health conditions involving cardiovascular, immune, endocrine, and nervous systems [43]. Even in healthy populations, insufficient and disordered sleep can still have profound impacts on health. For example, in a sample of healthy adolescents, sleep deprivation was found to induce mood deficits [60]. These findings are supported by how the same brain structures recognized for their roles in interoception and regulating mood and emotion, such as the insula, have also shown evidence to be sensitive to sleep-dependent modulation [2][61] [62]. Clinically diagnosed sleep disorders, especially Insomnia Disorder, are commonly comorbid to other conditions like chronic pain, as well as affective disorders, all of which have been associated with dysregulated interoception [11]. The relationship between chronic pain and subjective sleep quality is well-established, where those reporting poor sleep quality also report an increased level of pain, and vice-versa [11] [43] [63]. In a sample of patients with affective disorders, including anxiety and major depression, those who reported poor sleep quality via PSQI, performed worse on a measure of interoception (heart-rate discrimination task), representing a reduction in interoceptive accuracy when compared to healthy controls [56]. With consideration of the above research findings regarding the relationship between sleep and various health conditions, disorders, and physiological processing, it is worth investigating what may be the underlying connection: dysregulated interoception.

In this pilot trial of sleep-disordered subjects, regular use of the mechanical dermal stimulation device appears to have produced significant improvements in sleep and related areas. What is of potential interest to others in the nascent field of affective touch therapeutics is that the proposed underlying mechanism of action is not particularly salient to sleep. The intervention is not designed to specifically and differentially affect sleep. The intent is to improve interoceptive regulation. In the course of our work, we have used other markers of interoception, like stress, fear, depression, and anxiety, as outcome measures. In general, we find that improving interoceptive regulation in some domain is correlated with improvements in the specific feature of interest.

It should be the case that improving interoceptive regulation would lead to improvements in function in multiple domains as interoception is a key driver of emotional responses. The current, dominant psychiatric and even lay nosology around distress has a variety of classifications based on symptoms and complaints. A nosology based on mechanisms of action, such as dysregulated interoception, collapse many of the older 'symptom-based' nosology into that singular category. Although much of the therapeutic 'infrastructure' in psychiatry and psychology is rooted in the distinctiveness of disorders, practitioners recognize that, in the clinic, there is vast overlap among disorders and many comorbidities [10]. As we continue to work in developing therapeutic interventions around interoceptive regulation, we will continue to report using the existing symptom-based nosology, but always tie that back into the larger goal of improving interoceptive regulation. Future work that can help tie many currently disparate disorders (anxiety, eating disorders, sleep disturbances, etc.) to a common mechanism will help forge more common treatment methodologies and measures of improvement.

Conflicts of Interest

The authors declare the following financial interests which may be considered as potential competing interests: Gina Sensale, Sahithi Garikapati, Angelina Distefano and Sean Hagberg are employees of the Company. Jean-Marie Toher and Hanna Villa were employees of the Company. All authors hold stock ownership in the Company. Sean Hagberg has patent No. 16/241,227 pending to Affect Neuro, and a patent 10,786,666 issued to Affect Neuro. Sahithi Garikapati has patent No. 16/241,227 pending to Affect Neuro.

References

[1] G. Sensale, S. Garikapati, J. Toher, H. Villa, and S. Hagberg, "Mechanical Dermal Stimulation to Modulate the Interoceptive Network in Sleep-disordered Populations," The Sixth International Conference on Neuroscience and Cognitive Brain Information (BRAININFO 2021) IARIA, Jul. 2021, pp. 23–25, ISSN: 2519-8653, ISBN: 978-1-61208-885-3.

[2] E. F. Pace-Schott, "Chapter 48 - The Neurobiology of Dreaming," Principles and Practice of Sleep Medicine (Fifth Edition), pp. 563–575, Oct. 2011, doi: 10.1016/B978-1-4160-6645-3.00048-7.

[3] S. C. Sherrington, The Integrative Action of the Nervous System. New Haven, CT: Yale University Press, 1906.

[4] G. G. Berntson and S. S. Khalsa, "Neural Circuits of Interoception," Trends in Neurosciences, vol. 44, no. 1, pp. 17–28, Jan. 2021, doi: 10.1016/J.TINS.2020.09.011.

[5] P. Duquette, M. Tamietto, S. Gajraj Tewari, and L. Sun, "Increasing Our Insular World View: Interoception and Psychopathology for Psychotherapists," Frontiers in Neuroscience, vol. 11, p. 135, 2017, doi: 10.3389/fnins.2017.00135.

[6] L. F. Barrett and W. K. Simmons, "Interoceptive predictions in the brain," Nature Reviews Neuroscience, vol. 16, no. 7, pp. 419–429, 2015, doi: 10.1038/nrn3950.

[7] W. E. Mehling, C. Price, J. J. Daubenmier, M. Acree, E. Bartmess, and A. Stewart, "The Multidimensional Assessment of Interoceptive Awareness (MAIA)," PLOS ONE, vol. 7, no. 11, p. 48230, Nov. 2012, doi: 10.1371/JOURNAL.PONE.0048230.

[8] W. E. Whitehead, V. M. Drescher, P. Heiman, and B. Blackwell, "Relation of Heart Rate Control to Heartbeat Perception," Biofeedback and Self-Regulation, vol. 2, no. 4, Dec. 1977, doi: 10.1007/BF00998623.

[9] S. N. Garfinkel, M. F. Manassei, G. Hamilton-Fletcher, Y. In den Bosch, H. D. Critchley, and M. Engels, "Interoceptive dimensions across cardiac and respiratory axes," Philosophical transactions of the Royal Society of London, Series B, Biological Sciences, vol. 371, no. 1708, Nov. 2016, doi: 10.1098/RSTB.2016.0014.

[10] S. S. Khalsa et al., "Interoception and Mental Health: A Roadmap," Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, vol. 3, no. 6, pp. 501–513, Jun. 2018, doi: 10.1016/J.BPSC.2017.12.004.

[11] Y. Wei and E. J. van Someren, "Interoception relates to sleep and sleep disorders," Current Opinion in Behavioral Sciences, vol. 33, pp. 1–7, Jun. 2020, doi: 10.1016/J.COBEHA.2019.11.008.

[12] H. Y. Weng, J. L. Feldman, L. Leggio, V. Napadow, J. Park, and C. J. Price, "Interventions and Manipulations of Interoception," Trends in Neurosciences, vol. 44, no. 1, pp. 52–62, Jan. 2021, doi: 10.1016/J.TINS.2020.09.010.

[13] H. Schwarzmeier, N. I. Kleint, H. U. Wittchen, A. Ströhle, A. O. Hamm, and U. Lueken, "Characterizing the nature of emotional-associative learning deficits in panic disorder: An fMRI study on fear conditioning, extinction training and recall," European Neuropsychopharmacology, vol. 29, no. 2, pp. 306–318, Feb. 2019, doi: 10.1016/J.EURONEURO.2018.11.1108.

[14] C. Lamm and T. Singer, "The role of anterior insular cortex in social emotions," Brain Structure and Function, vol. 214, no. 5, pp. 579–591, Apr. 2010, doi: 10.1007/S00429-010-0251-3.

[15] J. Gibson, "Mindfulness, Interoception, and the Body: A Contemporary Perspective," Frontiers in Psychology, vol. 0, p. 2012, Sep. 2019, doi: 10.3389/FPSYG.2019.02012.

[16] E. Kuehn, K. Mueller, G. Lohmann, and S. Schuetz-Bosbach, "Interoceptive awareness changes the posterior insula functional connectivity profile," Brain Structure and Function, vol. 221, no. 3, pp. 1555–1571, Jan. 2015, doi: 10.1007/S00429-015-0989-8.

[17] K. S. Young et al., "The impact of mindfulness-based interventions on brain activity: A systematic review of functional magnetic resonance imaging studies," Neuroscience & Biobehavioral Reviews, vol. 84, pp. 424–433, Jan. 2018, doi: 10.1016/J.NEUBIOREV.2017.08.003.

[18] S. Guendelman, S. Medeiros, and H. Rampes, "Mindfulness and Emotion Regulation: Insights from Neurobiological, Psychological, and Clinical Studies," Frontiers in Psychology, vol. 0, p. 220, Mar. 2017, doi: 10.3389/FPSYG.2017.00220.

[19] N. M. Khoury, J. Lutz, and Z. Schuman-Olivier, "Interoception in Psychiatric Disorders: A Review of Randomized Controlled Trials with Interoception-based Interventions," Harvard Review of Psychiatry, vol. 26, no. 5, p. 250, Sep. 2018, doi: 10.1097/HRP.0000000000000170.

[20] N. Neukirch, S. Reid, and A. Shires, "Yoga for PTSD and the role of interoceptive awareness: A preliminary mixed-methods case series study," European Journal of Trauma & Dissociation, vol. 3, no. 1, pp. 7–15, Jan. 2019, doi: 10.1016/J.EJTD.2018.10.003.

[21] B. Bornemann, P. Kovacs, and T. Singer, "Voluntary upregulation of heart rate variability through biofeedback is improved by mental contemplative training," Scientific Reports, vol. 9, no. 1, pp. 1–13, May 2019, doi: 10.1038/s41598-019-44201-7.

[22] A. Wallman-Jones, P. Perakakis, M. Tsakiris, and M. Schmidt, "Physical activity and interoceptive processing: Theoretical considerations for future research," International Journal of Psychophysiology, vol. 166, pp. 38–49, Aug. 2021, doi: 10.1016/J.IJPSYCHO.2021.05.002.

[23] H. Olausson, J. Wessberg, I. Morrison, F. McGlone, and Å. Vallbo, "The neurophysiology of unmyelinated tactile afferents," Neuroscience & Biobehavioral Reviews, vol. 34, no. 2, pp. 185–191, Feb. 2010, doi: 10.1016/J.NEUBIOREV.2008.09.011.

[24] L. S. Löken, J. Wessberg, I. Morrison, F. McGlone, and H. Olausson, "Coding of pleasant touch by unmyelinated afferents in humans," Nature Neuroscience, vol. 12, no. 5, pp. 547–548, Apr. 2009, doi: 10.1038/nn.2312.

[25] P. Wei and R. Bao, "The Role of Insula-Associated Brain Network in Touch," BioMed Research International, vol. 2013, p. 11, 2013, doi: 10.1155/2013/734326.

[26] M. Björnsdotter, I. Morrison, and H. Olausson, "Feeling good: On the role of C fiber mediated touch in interoception," Experimental Brain Research, vol. 207, no. 3–4, pp. 149–155, Dec. 2010, doi: 10.1007/S00221-010-2408-Y.

[27] J. S. Feinstein et al., "Examining the short-term anxiolytic and antidepressant effect of Floatation-REST," PLoS ONE, vol. 13, no. 2, Feb. 2018, doi: 10.1371/JOURNAL.PONE.0190292.

[28] A. J. Arnold, P. Winkielman, and K. Dobkins, "Interoception and Social Connection," Frontiers in Psychology, vol. 0, p. 2589, Nov. 2019, doi: 10.3389/FPSYG.2019.02589.

[29] J. Ernst, G. Northoff, H. Böker, E. Seifritz, and S. Grimm, "Interoceptive awareness enhances neural activity during empathy," Human Brain Mapping, vol. 34, no. 7, pp. 1615–1624, Jul. 2013, doi: 10.1002/HBM.22014.

[30] Y. Wei et al., "I Keep a Close Watch on This Heart of Mine: Increased Interoception in Insomnia," Sleep, vol. 39, no. 12, p. 2113, Dec. 2016, doi: 10.5665/SLEEP.6308.

[31] D. Whibley et al., "Sleep and Pain: A Systematic Review of Studies of Mediation," The Clinical Journal of Pain,

vol. 35, no. 6, p. 544, Jun. 2019, doi: 10.1097/AJP.0000000000000697.

[32] N. K. Y. Tang, K. J. Wright, and P. M. Salkovskis, "Prevalence and correlates of clinical insomnia co-occurring with chronic back pain," Journal of sleep research, vol. 16, no. 1, pp. 85–95, Mar. 2007, doi: 10.1111/J.1365-2869.2007.00571.X.

[33] C. A. Palmer and C. A. Alfano, "Sleep and emotion regulation: An organizing, integrative review," Sleep Medicine Reviews, vol. 31, pp. 6–16, Feb. 2017, doi: 10.1016/J.SMRV.2015.12.006.

[34] M. E. Rumble, K. H. White, and R. M. Benca, "Sleep Disturbances in Mood Disorders," Psychiatric Clinics of North America, vol. 38, no. 4, pp. 743–759, Dec. 2015, doi: 10.1016/J.PSC.2015.07.006.

[35] V. Lo Martire, D. Caruso, L. Palagini, G. Zoccoli, and S. Bastianini, "Stress & sleep: A relationship lasting a lifetime," Neuroscience & Biobehavioral Reviews, vol. 117, pp. 65–77, Oct. 2020, doi: 10.1016/J.NEUBIOREV.2019.08.024.

[36] M. C. Chen et al., "Anterior Insula Regulates Multiscale Temporal Organization of Sleep and Wake Activity," Journal of Biological Rhythms, vol. 31, no. 2, pp. 182–193, Jan. 2016, doi: 10.1177/0748730415627035.

[37] D. Stoffers et al., "Orbitofrontal Gray Matter Relates to Early Morning Awakening: A Neural Correlate of Insomnia Complaints?," Frontiers in Neurology, vol. 3, p. 105, Jun. 2012, doi: 10.3389/FNEUR.2012.00105.

[38] H. S. Gompf et al., "Locus Ceruleus and Anterior Cingulate Cortex Sustain Wakefulness in a Novel Environment," The Journal of Neuroscience, vol. 30, no. 43, p. 14543, Oct. 2010, doi: 10.1523/JNEUROSCI.3037-10.2010.

[39] Y. Wei, T. F. Blanken, and E. J. W. Van Someren, "Insomnia Really Hurts: Effect of a Bad Night's Sleep on Pain Increases With Insomnia Severity," Frontiers in Psychiatry, vol. 0, p. 377, Aug. 2018, doi: 10.3389/FPSYT.2018.00377.

[40] M. C. Chen, C. Chang, G. H. Glover, and I. H. Gotlib, "Increased insula coactivation with salience networks in insomnia," Biological Psychology, vol. 97, pp. 1–8, Mar. 2014, doi: 10.1016/J.BIOPSYCHO.2013.12.016.

[41] T. Arora, M. Barbato, S. Al Hemeiri, O. M. Omar, and M. A. AlJassmi, "A mysterious sensation about sleep and health: the role of interoception," BMC Public Health, vol. 21, p. 1584, Dec. 2021, doi: 10.1186/s12889-021-11603-0.

[42] Y. Chong, C. D. Fryar, and Q. Gu, "Prescription sleep aid use among adults: United States, 2005-2010," Hyattsville, MD, Aug. 2013. [Online]. Available: https://www.cdc.gov/nchs/data/databriefs/db127.pdf. [Accessed: 2021.12.10].

[43] Y. Liu, "Prevalence of Healthy Sleep Duration among Adults — United States, 2014," MMWR. Morbidity and Mortality Weekly Report, vol. 65, no. 6, pp. 137–141, Feb. 2019, doi: 10.15585/MMWR.MM6506A1.

[44] N. McArdle et al., "The prevalence of common sleep disorders in young adults: a descriptive population-based study," Sleep, vol. 43, no. 10, p. zsaa072, Oct. 2020, doi: 10.1093/SLEEP/ZSAA072.

[45] J. Labad et al., "The Role of Sleep Quality, Trait Anxiety and Hypothalamic-Pituitary-Adrenal Axis Measures in Cognitive Abilities of Healthy Individuals," International Journal of Environmental Research and Public Health, vol. 17, no. 20, p. 7600, Oct. 2020, doi: 10.3390/IJERPH17207600.

[46] N. Darchia et al., "Relationship between Sleep Disorders and Health Related Quality of Life—Results from the Georgia SOMNUS Study," International Journal of Environmental Research and Public Health, vol. 15, no. 8, p. 1588, Jul. 2018, doi: 10.3390/IJERPH15081588.

[47] S. Wilson and S. Argyropoulos, "Antidepressants and sleep: a qualitative review of the literature," Drugs, vol. 65, no. 7, pp. 927-947, May 2005, doi: 10.2165/00003495-200565070-00003.

[48] T. Fitzgerald and J. Vietri, "Residual Effects of Sleep Medications Are Commonly Reported and Associated with Impaired Patient-Reported Outcomes among Insomnia Patients in the United States," Sleep Disorders, vol. 2015, pp. 1–9, 2015, doi: 10.1155/2015/607148.

[49] R. B. Costello et al., "The effectiveness of melatonin for promoting healthy sleep: a rapid evidence assessment of the literature," Nutrition Journal, vol. 13, no. 1, pp. 1–17, Nov. 2014, doi: 10.1186/1475-2891-13-106.

[50] J. Glass, K. L. Lanctôt, N. Herrmann, B. A. Sproule, and U. E. Busto, "Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits," BMJ, vol. 331, pp. 1169–1173, Nov. 2005, doi: 10.1136/BMJ.38623.768588.47.

[51] H. L. Rusch et al., "The effect of mindfulness meditation on sleep quality: a systematic review and meta-analysis of randomized controlled trials," Annals of the New York Academy of Sciences, vol. 1445, no. 1, pp. 5–16, Jun. 2019, doi: 10.1111/NYAS.13996.

[52] R. Ackerley, I. Carlsson, H. Wester, H. Olausson, and H. Backlund Wasling, "Touch perceptions across skin sites: Differences between sensitivity, direction discrimination and pleasantness," Frontiers in Behavioral Neuroscience, vol. 8, p. 54, Feb. 2014, doi: 10.3389/FNBEH.2014.00054/BIBTEX.

[53] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research," Psychiatry Research, vol. 28, no. 2, pp. 193–213, May 1989, doi: 10.1016/0165-1781(89)90047-4.

[54] D. L. Ewing, M. Manassei, C. G. van Praag, A. O. Philippides, H. D. Critchley, and S. N. Garfinkel, "Sleep and the heart: Interoceptive differences linked to poor experiential sleep quality in anxiety and depression," Biological Psychology, vol. 127, p. 163, Jul. 2017, doi: 10.1016/J.BIOPSYCHO.2017.05.011.

[55] J. A. Brewer, A. Roy, A. Deluty, T. Liu, and E. A. Hoge, "Can Mindfulness Mechanistically Target Worry to Improve Sleep Disturbances? Theory and Study Protocol for App-Based Anxiety Program," Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, vol. 39, no. 9, p. 776, Sep. 2020, doi: 10.1037/HEA0000874.

[56] D. Fuller et al., "Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review," JMIR Mhealth and Uhealth, vol. 8, no. 9, Sep. 2020, doi: 10.2196/18694.

[57] "Solutions IRB." [Online]. Available: https://www.solutionsirb.com/. [Accessed: 2021.12.10].

[58] M. A. Short and M. Louca, "Sleep deprivation leads to mood deficits in healthy adolescents," Sleep Medicine, vol. 16, no. 8, pp. 987–993, Aug. 2015, doi: 10.1016/J.SLEEP.2015.03.007.

[59] E. F. Pace-Schott et al., "Physiological feelings," Neuroscience & Biobehavioral Reviews, vol. 103, pp. 267–304, Aug. 2019, doi: 10.1016/J.NEUBIOREV.2019.05.002.

[60] E. A. Nofzinger et al., "Increased Activation of Anterior Paralimbic and Executive Cortex From Waking to Rapid Eye Movement Sleep in Depression," Archives of General Psychiatry, vol. 61, no. 7, pp. 695–702, Jul. 2004, doi: 10.1001/ARCHPSYC.61.7.695.

[61] A. G. Doufas, "Chapter 133 - Pain and sleep," Principles and Practice of Sleep Medicine (Sixth Edition), Elsevier, pp. 1313-1322, 2017, doi: 10.1016/B978-0-323-24288-2.00133-1.

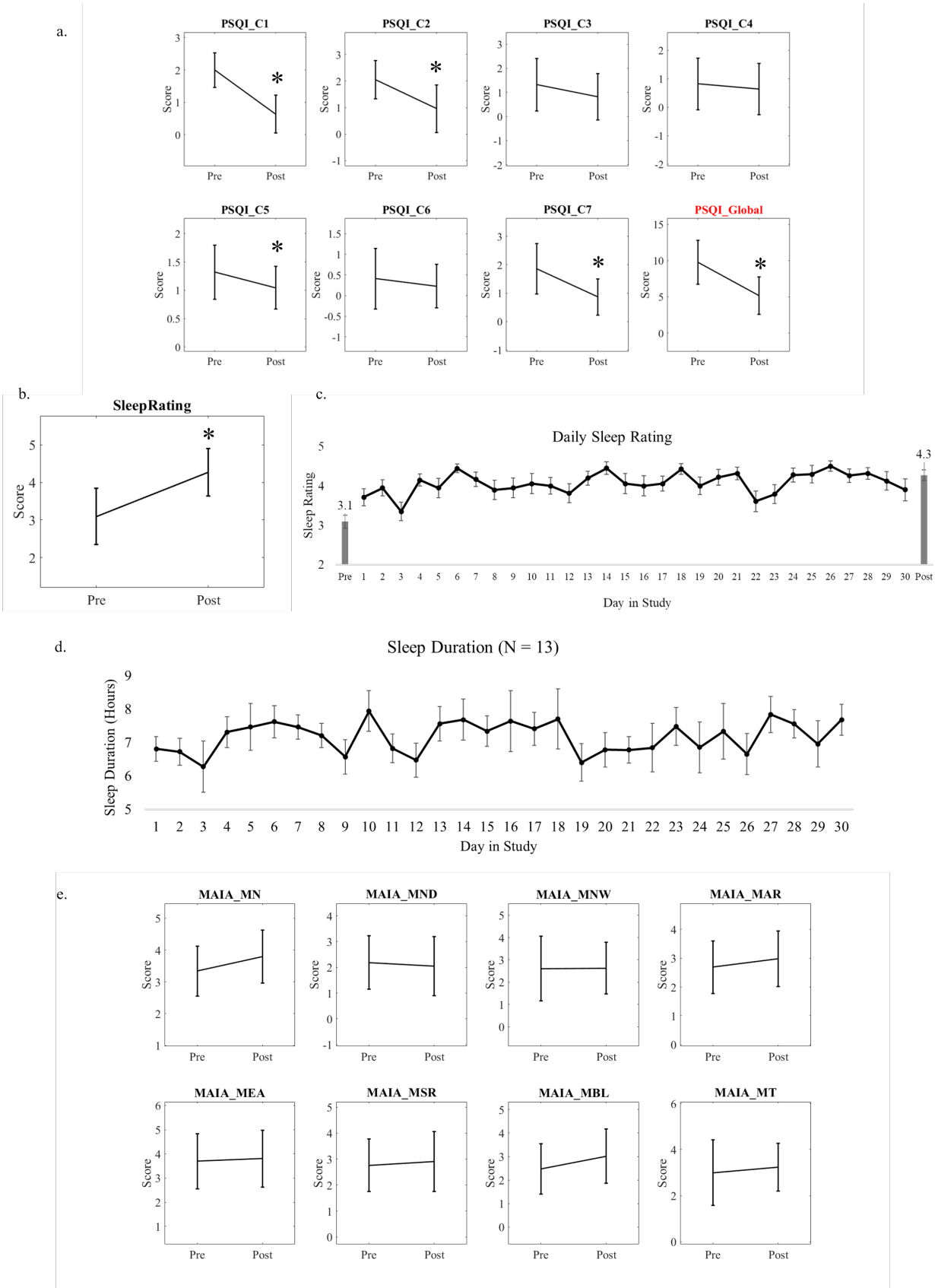


Figure 3: (a) Average PSQI Scores (b) Average Sleep Quality Rating Scores (c) Average Daily Sleep Quality Ratings (d) Average Sleep Duration measured via Garmin (n=13) (e) Average MAIA Scores.

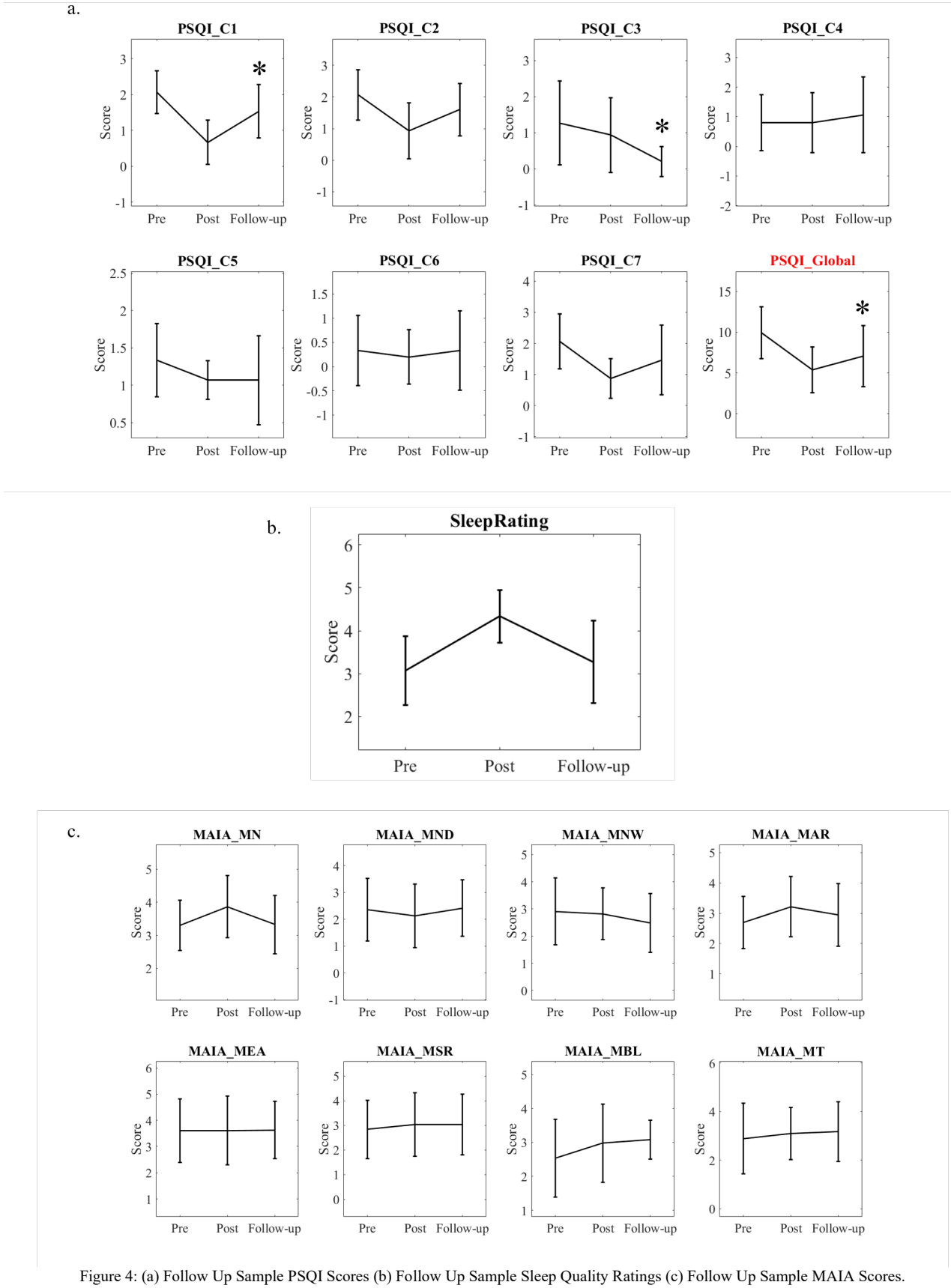


Figure 4: (a) Follow Up Sample PSQI Scores (b) Follow Up Sample Sleep Quality Ratings (c) Follow Up Sample MAIA Scores.

3D Gaze on Stationary and Moving Visual Targets in Mixed Reality Environments

Kenta Kato, Oky Dicky Ardiansyah Prima
Graduate School of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: g236s002@s.iwate-pu.ac.jp, prima@iwate-pu.ac.jp

Abstract— **With the spread of Head-Mounted Displays (HMDs), various simulations have been conducted using Mixed Reality (MR) environments that merge virtual objects in a physical space. Our Three-Dimensional (3D) perception may change as opportunities to have more virtual 3D experiences in such an environment increase. The relationship between differences in depth perception and changes in 3D gaze behavior would be of interest, but such detailed analysis is yet to be conducted. In this study, we developed an Optical See-Through Head-Mounted Display (OST-HMD) and experimentally evaluated the effect of an MR environment on 3D gaze measurement. Our experiments showed that the relative size of the 3D visual targets and the surrounding depth cues had no effect on the accuracy of the 3D gaze. However, it is important to consider the building of a polynomial for projecting the 3D gaze, and there is room for improvement in the polynomial. We have successfully measured the scanpath of 3D gaze on a visual target approaching from 200 cm. This result suggests that our OST-HMD can measure 3D gaze in the personal space, defined by Cutting (1997) as a space up to 1.5 m from the viewer.**

Keywords-3D gaze estimation; Optical See-Through Head-Mounted Display; depth perception; Mixed Reality.

I. INTRODUCTION

Three-Dimensional (3D) visualization of medical images of the human body and their interactive manipulation have been expected to be utilized in the medical field for safe surgery through preoperative scenario planning. In the past decade, the hardware for processing 3D information has improved dramatically, enabling more accurate tracking of the perspective and input of the user. In addition, Augmented Reality (AR) and Mixed Reality (MR) environments, which integrate the virtual world with the real world, have made it possible to manipulate 3D contents naturally. However, 3D contents in these environments may cause operation errors if the contents are perceived incorrectly. Studies on accurate projection of these contents in real environments are still in progress.

Problems related to the position, scale, and posture of 3D contents in virtual and real environments require not only verification based on subjective depth perception, but also objective analysis based on the 3D gaze information of viewers. We studied the relationship between each environment and 3D gaze information by measuring vergence eye movement-based 3D gaze to 3D visual targets in both environments [1]. However, there was no significant difference in the measured 3D gaze between the environments with and without depth cues.

There are nine sources of information for perceiving depth: binocular disparity, convergence, occlusion, relative size, height in the visual field, relative density, aerial perspective, accommodation, and motion parallax [2]. The first two pieces of information are binocular cues, which perceive depth from the disparity caused by the disparity between the eyes, and the remaining are monocular cues caused by the relative position of the observer and the object. Cutting (1997) argued that the effect of depth cues on distance estimation depends on the perceptual areas regarding the observer's position: personal space (~1.5m), action space (1.5m to 30m), and vista space (from 30m) [3]. Many studies on the relationship between vergence and perceived distance have experimented with visual targets of up to 2m [4]-[6].

Binocular disparity-based visual targets are necessary to measure 3D gaze information. Head-Mounted Displays (HMDs) are often used to present these visual targets. HMDs can be classified into two categories: Video See-Through Head-Mounted Display (VST-HMD) and Optical See-Through Head-Mounted Display (OST-HMD). The former projects a 3D object superimposed on a video image of the real world, while the latter projects a 3D object projected on a half-mirror superimposed on the real world. The viewable area in VST-HMD depends on the Field of View (FoV) of the scene camera and the display. In contrast, OST-HMD transparently displays 3D objects on the viewer's FoV, making it suitable for manipulating 3D virtual objects over real objects simultaneously. OST-HMDs are also relatively less susceptible to the VR sickness that occurs with devices that cover the field of view such as the VST-HMD [7][8]. Hence, the adoption of OST-HMD is under positive consideration in the medical field [9].

3D gaze can be measured by analyzing the relationship between the position of an object and the eye's vergence angle when looking at that object [10]-[13]. Öney et al. (2020) used binocular eye tracking of the Microsoft HoloLens to measure 3D gaze in visual search tasks for 3D objects located within 1.25m to 5m from the subject in an MR environment. Their experiments, however, showed a significantly larger measurement error of more than 1 m when the focused object was only 3.5 m away from the viewer [14]. Kato et al. (2018) confirmed the effect of adding patterns to visual targets used in gaze calibration to improve the accuracy of 3D gaze measurement [15].

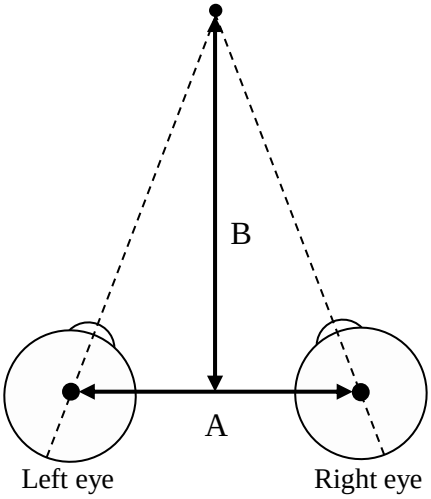


Figure 1. The basic binocular geometry

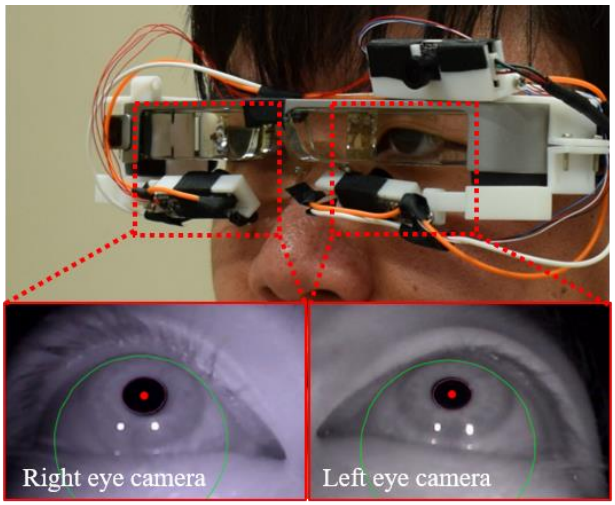


Figure 3. Our 3D eye tracker used in this study.

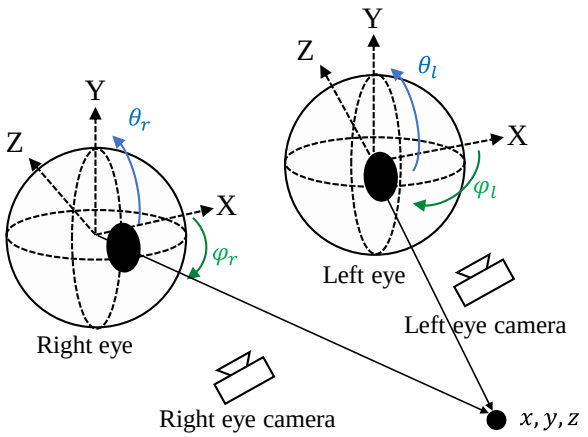


Figure 2. The relationship between the binocular cameras and the estimated the line-of-sight.

In this study, we measure 3D gaze in an MR environment to evaluate the effects of the relative size of the 3D visual target and the surrounding physical environment as a cue for 3D perception on the accuracy of the 3D gaze, and to characterize the scan path of 3D gaze for stationary and moving visual targets. For this study, a 3D eye tracker using OST-HMD intended for MR was developed.

The rest of this paper is organized as follows. Section II describes the calculation of 3D gaze based on eye vergence. Section III describes the 3D eye tracker based on OST-HMD developed for this study. Section IV explains our experiments. Section V describes the results. Finally, Section VI presents our conclusion.

II. 3D Gaze Measurement Based on Eye Vergence

The calculation of the 3D gaze depends on the relative positions of the two eyes to a given visual target. Vergence angle is the angle between the line-of-sight of the two eyes. The change in vergence angle occurs when the target is moved at a certain distance from the observer. Figure 1 shows the basic binocular geometry. A and B are the disparity distance between eye centers and the distance from the visual target to the observer, respectively. A 3D gaze point (x_g, y_g, z_g) can be calculated from each line-of-sight, ($Pitch_L$, Yaw_L) and ($Pitch_R$, Yaw_R). The pitch and yaw angles describe angles between the X axis and Y axis, respectively.

In a binocular eye tracker, the eye cameras are used to determine the lines of sight of the two eyes, but in many cases, the cameras are placed below the eyes to avoid interfering with the field of view. Due to the placement of these cameras, the calculated pitch and yaw angles of the eyes differ from the angles as viewed from the front. Therefore, to obtain the 3D gaze from the eye tracker, the position of the visual target (x, y, z) must be calibrated using the line-of-sight of both eyes ($Pitch_L$, Yaw_L, $Pitch_R$, Yaw_R). Polynomials are generally used for gaze calibration. Figure 2 shows the relationship between the binocular cameras and the estimated pitch and yaw angles. These relationships can be solved by the following nonlinear multiple regression equations.

$$x = f(Pitch_L, Yaw_L, Pitch_R, Yaw_R) + \alpha \quad (1)$$

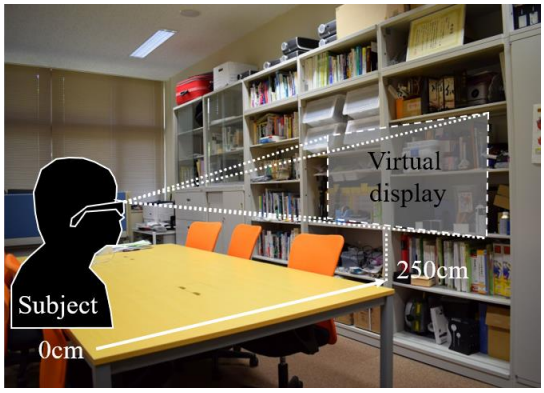
$$y = g(Pitch_L, Yaw_L, Pitch_R, Yaw_R) + \beta \quad (2)$$

$$z = h(Pitch_L, Yaw_L, Pitch_R, Yaw_R) + \gamma \quad (3)$$

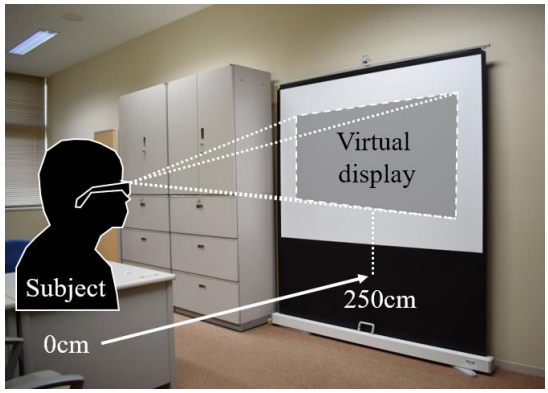
Here, f, g, h are composite functions of line-of-sight. α, β, γ are the residuals of each equation.

III. Our 3D Eye Tracker Based on OST-HMD

A variety of OST-HMDs are available on the market. In this study, we adopted the Moverio (BT-30E, EPSON) to build our OST-HMD 3D eye tracker in MR environment. This device can display images from a PC, making it easy to provide users with 3D visual stimulation. Three simultaneous USB camera modules (KYT-U030-3NF, KAYETON) were used to capture both eyes and the viewer's scene. These

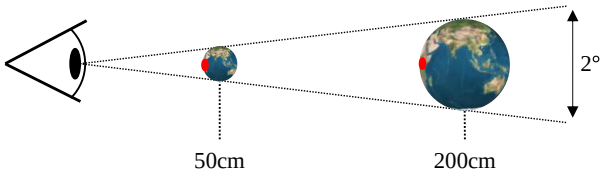


(a) A room with depth cues.

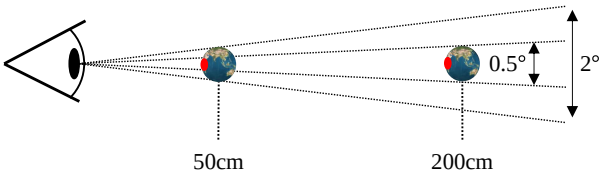


(b) A room without depth cues.

Figure 4. The physical environment for the experiments in this study.



(a) Visual target with a 2° viewing angle in size regardless of distance.



(b) A fixed-size visual target.

Figure 5. Two types of visual targets used in this study.

cameras operate at 60 Hz allowing tracking of eye movements at a rate comparable to low- to mid-end eye trackers, such as the Tobii Pro Nano. For tracking the pupil, we installed a 940nm high-pass filter (FUJIFILM IR-94) in the lens of the camera that captures the eyeball to block visible light while allowing infrared light to pass through. Figure 3 shows our OST-HMD 3D eye tracker. Each camera mount is made with a 3D printer. openFrameworks v0.11.0, an open-source C++ toolkit, was used to visualize the virtual targets and record the measurement data. We used Pupil Capture (Pupil Labs), an open-source eye tracking platform, to calculate the line-of-sight of each eye from the estimated eyeball center [16]. The line-of-sight data measured by the Pupil Capture is sent to openFrameworks using ZeroMQ, an asynchronous messaging library, and 3D gaze is calculated based on the relationship between the visual targets and the line-of-sight data.

The visual targets used for the measurements were created by generating binocular disparity to guide the vergence. The interpupillary distance was fixed at 6.3cm, which is the average interpupillary distance for Japanese [17].

A. MR Environments

In this study, two MR environments were established as shown in Figure 4. An environment with depth cues is a room with a table and chairs and shelves with various objects within the observer's FoV. On the other hand, an environment without depth cues is a room where a projector screen is placed in the observer's FoV. The viewer wearing OST-HMD is seated 250cm away from the shelf and screen projector to perceive a virtual screen equivalent to 40 inches. The lighting in the room is fluorescent with a color temperature of 5500K. The brightness of the room is approximately 150lx.

B. 3D Gaze Estimation

The polynomials in equations (1)-(3) are calculated using the positions of multiple visual targets and the line-of-sight of both eyes when gazing at those targets. Let (θ_l, θ_r) represent pitch angles and (φ_l, φ_r) represent yaw angles of the line-of-sight, respectively. Two types of polynomials were constructed for this study as follows.

1. Type I: polynomial used by Kato and Prima (2021) [1]

This polynomial achieves an average 3D gaze accuracy of less than 25cm, about four times better than the results presented by Öney et al. (2020) [14].

$$G_x = a_1\theta_r^2 + a_2\varphi_r^2 + a_3\theta_l^2 + a_4\varphi_l^2 + a_5\theta_r\varphi_r + a_6\theta_l\varphi_l + a_7\theta_r\theta_l + a_8\theta_r\varphi_l + a_9\theta_l\varphi_r + a_{10}\varphi_r\varphi_l + a_{11}\theta_r + a_{12}\varphi_r + a_{13}\theta_l + a_{14}\varphi_l + a_{15} \quad (4)$$

$$G_y = b_1\theta_r^2 + b_2\varphi_r^2 + b_3\theta_l^2 + b_4\varphi_l^2 + b_5\theta_r\varphi_r + b_6\theta_l\varphi_l + b_7\theta_r\theta_l + b_8\theta_r\varphi_l + b_9\theta_l\varphi_r + b_{10}\varphi_r\varphi_l + b_{11}\theta_r + b_{12}\varphi_r + b_{13}\theta_l + b_{14}\varphi_l + b_{15} \quad (5)$$

$$G_z = c_1\theta_r^2 + c_2\varphi_r^2 + c_3\theta_l^2 + c_4\varphi_l^2 + c_5\theta_r\varphi_r + c_6\theta_l\varphi_l + c_7\theta_r\theta_l + c_8\theta_r\varphi_l + c_9\theta_l\varphi_r + c_{10}\varphi_r\varphi_l + c_{11}\theta_r + c_{12}\varphi_r + c_{13}\theta_l + c_{14}\varphi_l + c_{15}. \quad (6)$$

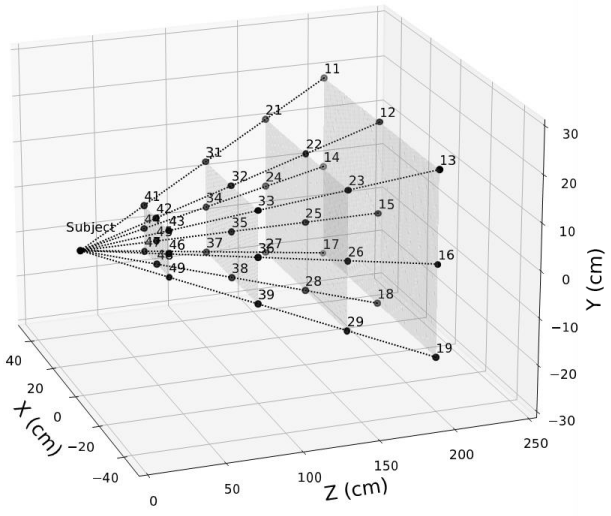


Figure 6. Locations of the visual targets for eye calibration.

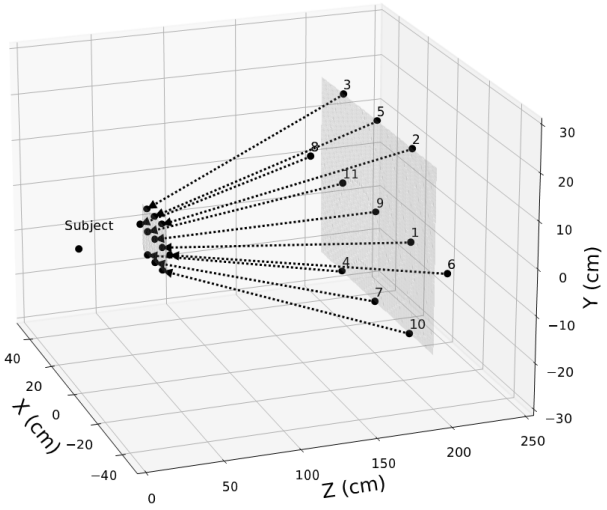


Figure 8. Path of the moving visual targets.

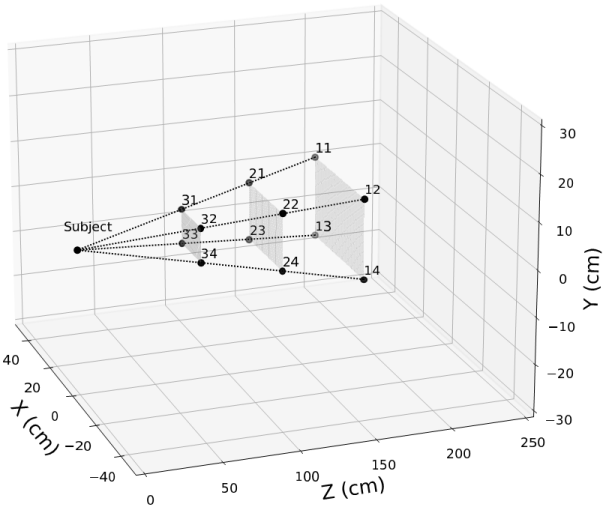


Figure 7. Locations of the stationary visual targets.

Coefficients $(a_1 \sim a_{15}, b_1 \sim b_{15}, c_1 \sim c_{15})$ are calculated by the least-squares method based on the correspondence between the pitch and yaw angles $(\theta_l, \theta_r, \varphi_l, \varphi_r)$ of each eye and the position of the gazing target.

2. Type II: polynomial with constrained pitch angles.

The range of vertical eye movements is shorter than the range of horizontal eye movements. Therefore, we consider it necessary to minimize the effect of vertical eye movements in the x- and y-axes of the 3D gaze. For this purpose, we introduce θ_{rl} as the average of the pitch angles of both eyes.

$$x = a_1\varphi_r^2 + a_2\varphi_l^2 + a_3\varphi_r\varphi_l + a_4\varphi_r + a_5\varphi_l + a_6 \quad (7)$$

$$y = b_1\theta_{rl}^2 + b_2\varphi_r^2 + b_3\varphi_l^2 + b_4\theta_{rl}\varphi_r + b_5\theta_{rl}\varphi_l + b_6\varphi_r\varphi_l + b_7\theta_{rl} + b_8\varphi_r + b_9\varphi_l + b_{10} \quad (8)$$

$$z = c_1\varphi_r^2 + c_2\varphi_l^2 + c_3\varphi_r\varphi_l + c_4\varphi_r + c_5\varphi_l + c_6 \quad (9)$$

Coefficients $(a_1 \sim a_6, b_1 \sim b_{10}, c_1 \sim c_6)$ are calculated by the least-squares method.

IV. EXPERIMENT

The following steps are taken to characterize the scan path of 3D gaze for stationary and moving visual targets as well as to evaluate the impact of the relative size of the 3D visual target and the surrounding physical environment on the accuracy of 3D gaze. The subject wears the OST-HMD and is seated 250cm away from a shelf, an environment with depth cues, and a screen projector, which is without depth cues.

Step 1. Adjust parameters such as blob thresholds and pre-defined pupil sizes to improve pupil detection in the Pupil Capture software program.

Step 2. Present the subject with 36 stationary 3D visual targets placed in a virtual space and measure the line of sight. Each target is presented continuously for 3 seconds.

Step 3. Perform a Two-Dimensional (2D) eye calibration before constructing the polynomial for 3D gaze estimation. After a successful 2D calibration within 2°, display the other 12 stationary visual targets to the subject. If the 2D calibration fails, return to step 1.

Step 4. Ask the subject to track a single visual target approaching from 200cm to 50cm in virtual space. Here, the 200cm was determined in consideration

TABLE I. THE ACCURACIES OF 3D GAZE IN ENVIRONMENTS WITH AND WITHOUT DEPTH CUES, AND WITH AND WITHOUT ADJUSTMENTS OF VISUAL TARGETS (cm)

Subject	With depth cues				Without depth cues			
	Fixed-size		Adaptively-adjusted size		Fixed-size		Adaptively-adjusted size	
	Type I	*Type II*	*Type I*	*Type II*	*Type I*	*Type II*	*Type I*	*Type II*
1	18.65	28.50	21.49	25.79	11.06	25.78	11.25	19.88
2	19.23	24.84	20.27	30.36	21.98	28.71	26.80	32.31
3	13.22	20.29	7.52	10.28	20.22	22.53	12.79	16.61
4	19.00	25.46	22.22	26.03	11.90	13.64	9.91	15.86
5	16.15	21.00	18.36	30.77	25.38	35.41	12.85	18.74
Mean	17.25	24.02	17.97	24.65	18.11	25.21	14.72	20.68
Std. Dev.	2.295	3.026	5.385	7.480	5.669	7.181	6.137	5.992

that the effective range of vergence is within 2m [3]. A total of 11 moving visual targets coming from different directions are displayed. Each target will take 5 seconds to travel. To give subjects a time to prepare for gazing, each target was paused for one second at the beginning and end of its movement.

During Step 2 and Step 3, line-of-sight of both eyes are recorded for 1 second while gazing at the target without blinking. A red dot blinks in the center of the target during the recording. In Step 5, all line-of-sight of both eyes are recorded during the appearance of the target. Post-processing is performed to remove the data when a blink occurs.

Figure 5 shows two types of visual targets in this study. The first type adjusts its size to appear consistently 2° regardless of distance, while the second type keeps the size constant. As a result, the second type of visual target is observed to change its size from 2° in diameter at 50cm to 0.5° at 200cm.

A. Visual Targets for the Eye Calibration

Vertical virtual planes are generated at 50cm intervals from 50cm to 200cm where nine visual targets are placed on each plane. These visual targets were placed at 5° horizontally and 2.5° vertically. Figure 6 shows the locations of the visual targets in each plane. The vector connecting the *n*-th target in each plane intersects the position of the viewpoint.

B. Stationary Visual Targets

Like the visual target for the eye calibration, three stationary vertical virtual planes are generated at 50cm intervals from 75cm to 175cm. Four visual targets are placed on each side, for a total of 12 targets. These visual targets were placed at 2.5° horizontally and 1.25° vertically. Figure 7 shows the locations of the visual targets in each plane.

C. Moving Visual Targets

11 moving visual targets are designed to start moving from a 200cm virtual plane and stop at a 50cm virtual plane. Figure 8 shows the locations of the start and stop locations of the visual targets. These visual targets were placed at 3° horizontally and 2.5° vertically. The dotted lines represent the trajectories of the moving targets.

V. RESULTS

Five subjects (male, mean age 23.6) participated in the experiment. They were tested for visual acuity using a Landolt ring to confirm that their vision achieved 1.0 or better.

TABLE II. ERRORS OF THE 3D GAZE TO THE STATIONARY VISUAL TARGETS PROJECTED BY TYPE I AND II POLYNOMIALS (cm)

	Type 1	Type 2
1	29.65	12.35
2	28.96	30.77
3	52.87	47.27
4	22.84	21.62
5	26.23	22.41
6	35.81	22.40
7	24.53	18.62
8	12.11	10.49
9	34.56	27.50
10	44.84	41.59
11	36.94	25.03
12	27.47	18.41
13	22.73	19.45
14	16.31	11.62
15	71.34	26.74
16	21.96	22.77
17	27.45	20.87
18	23.41	26.82
19	28.14	31.42
20	37.24	41.26
Mean	32.11	26.88
Std. Dev.	9.846	11.744

They were also asked to fill out a questionnaire to confirm that they had no health concerns. However, one of the subjects was not able to track the moving visual targets correctly, so there were no results of gazing at targets only for that subject.

A. Pre-processing

The data with a pupil detection confidence of less than 60% was excluded to avoid the inclusion of various noises in the gaze data. This is done using the features of the Pupil Capture software.

B. Measurement Accuracy

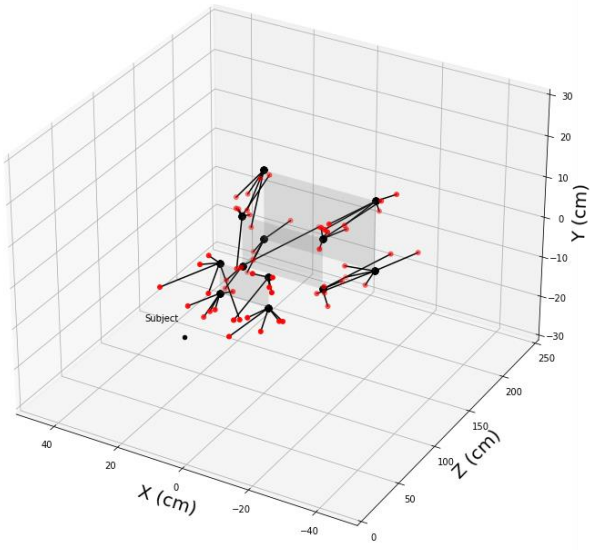
Accuracies for the 3D gaze measurement is measured by

$$Acc = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{xi} - G_{xi})^2 + (T_{yi} - G_{yi})^2 + (T_{zi} - G_{zi})^2}. \quad (10)$$

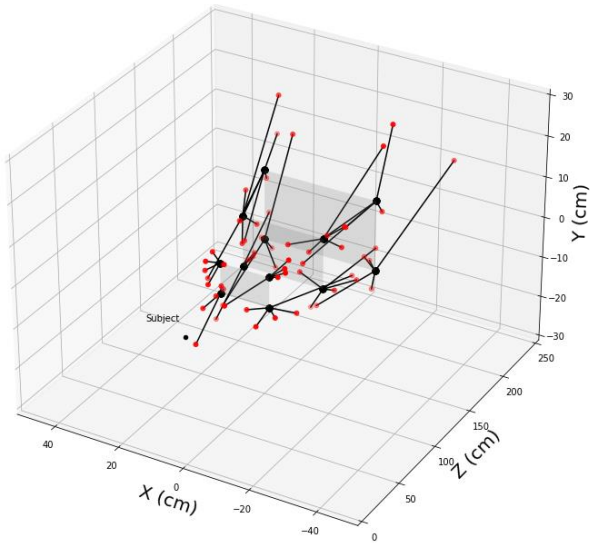
Here, n is the number of targets used for the measurement, T_{xi}, T_{yi}, T_{zi} and G_{xi}, G_{yi}, G_{zi} are the coordinates of the i-th target and the associated eye-gaze points projected by polynomials (Type I or Type II) described in Section III.

TABLE III. ACCURACIES OF THE 3D GAZE TO THE MOVING VISUAL TARGETS (*cm*)

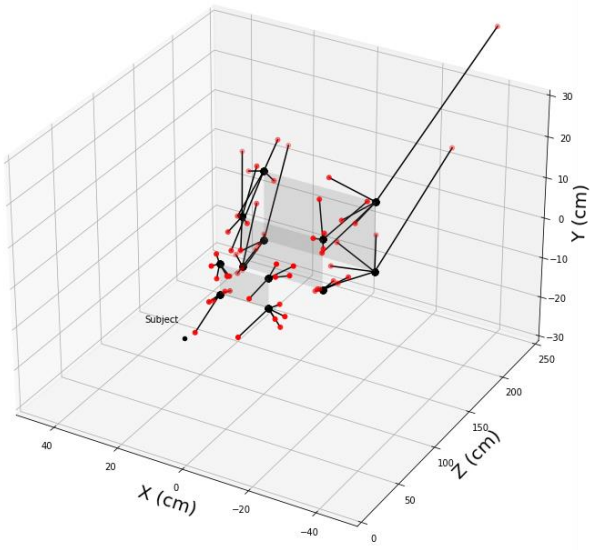
Subject	With depth cues				Without depth cues			
	Fixed-size		Adaptively-adjusted size		Fixed-size		Adaptively-adjusted size	
	Type I	*Type II*	*Type I*	*Type II*	*Type I*	*Type II*	*Type I*	*Type II*
1	33.49	18.31	89.35	17.77	25.87	16.51	43.50	19.34
3	43.13	29.55	15.42	16.33	16.26	14.57	124.74	39.83
4	21.37	28.30	73.70	40.16	24.32	26.48	49.82	56.04
5	27.91	24.36	33.07	26.98	365.92	25.96	99.56	25.68
Mean	31.48	25.13	52.89	25.31	108.09	20.88	79.41	35.22
Std. Dev.	7.980	4.376	29.830	9.500	148.900	5.388	34.008	14.123



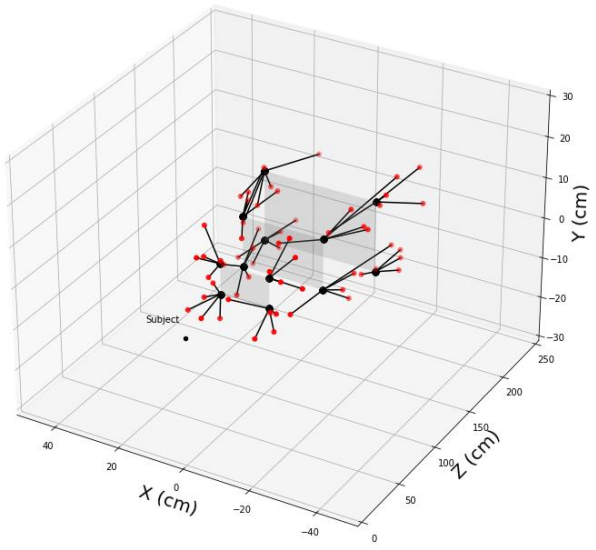
(a) 3D gaze in environments with depth cues and with adjustments of visual targets.



(b) 3D gaze in environments with depth cues and without adjustments of visual targets.

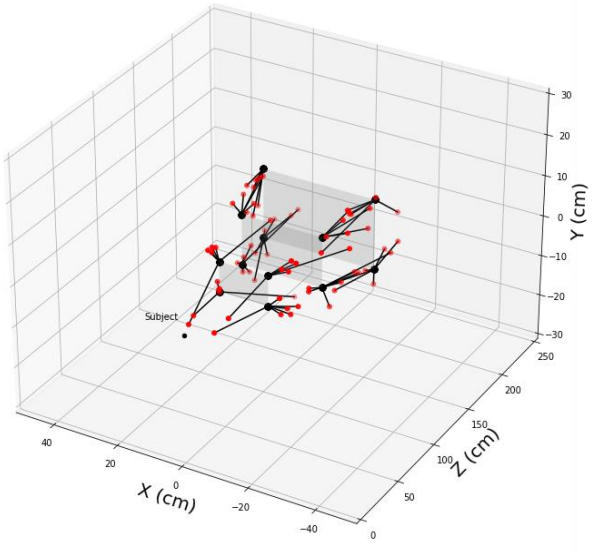


(c) 3D gaze in environments without depth cues and with adjustments of visual targets.

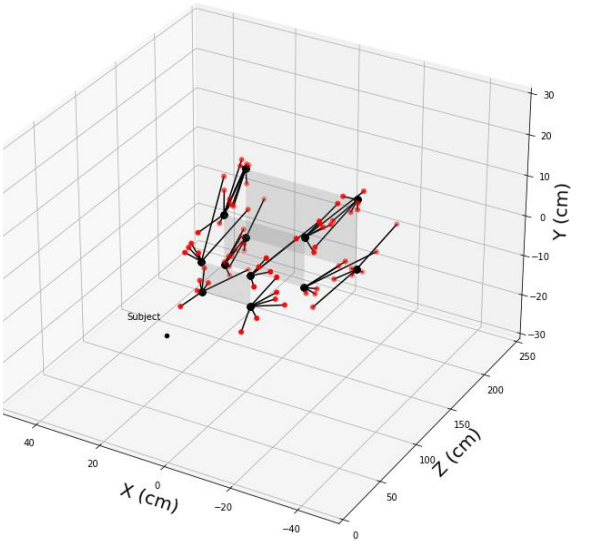


(d) 3D gaze in environments without depth cues and without adjustments of visual targets.

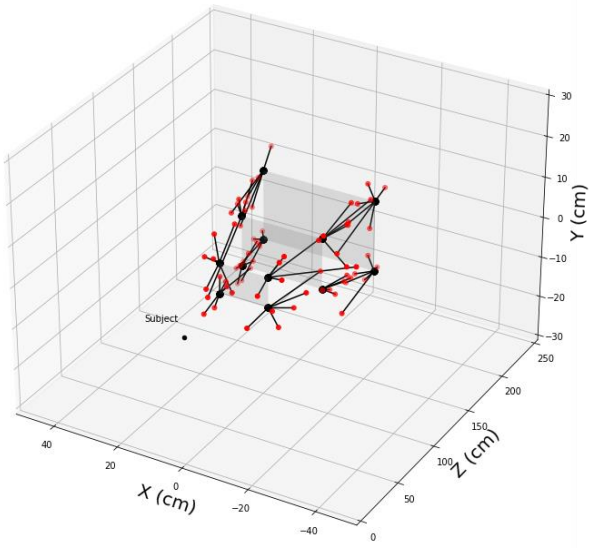
Figure 9. Distribution of 3D gaze to stationary visual targets projected by Type I



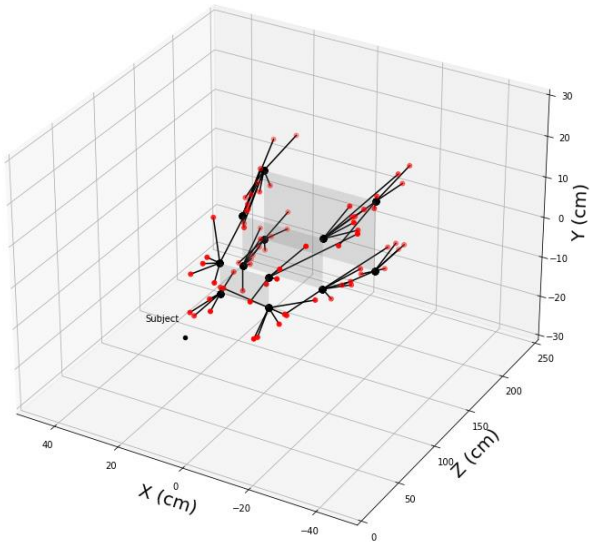
(a) 3D gaze in environments with depth cues and with adjustments of visual targets.



(b) 3D gaze in environments with depth cues and without adjustments of visual targets.



(c) 3D gaze in environments without depth cues and with adjustments of visual targets.

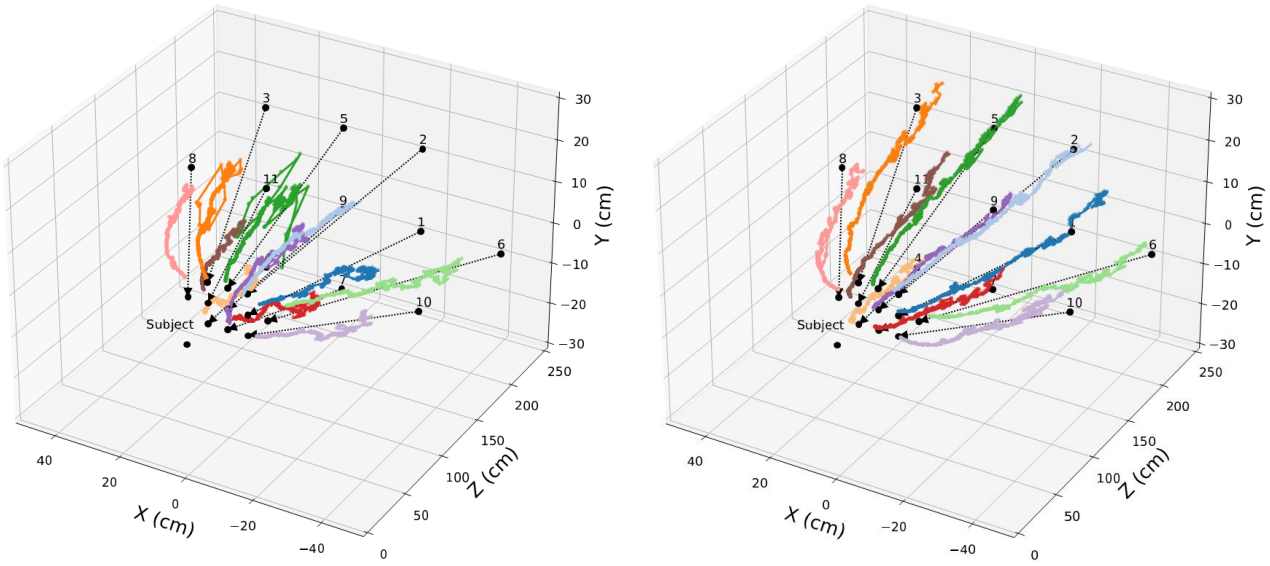


(d) 3D gaze in environments without depth cues and without adjustments of visual targets.

Figure 10. Distribution of 3D gaze to stationary visual targets projected by Type II

Table I shows the accuracies of gaze estimation in environments with and without depth cues, and with and without adjustments of the visual targets. Overall, the accuracy of the 3D gaze calculated by the Type I polynomial is superior regardless of the environments or the visual targets. To confirm what affects the accuracy of 3D gaze, we performed a 2 × 2 × 2, depth cue × target adjustment × polynomial type, three-way analysis of variance on these results. The main effect was found for polynomial type (F (1, 32) = 10.949, p = 0.002), supporting the result that Type I polynomials are better at computing 3D gaze.

Using the polynomial equations of Type I and Type II obtained from the eye calibration, we visualized the distribution of the 3D gaze to stationary visual targets. Figures 9 and 10 show the distribution of the 3D gaze projected by Type I and Type II, respectively. The red lines indicate the relationships between the positions of the projected 3D gaze (red dots) of all subjects and the visual targets (black dots). A longer line segment means a larger error of the 3D gaze. In contrast to the eye calibration results, overall, the errors in the 3D gaze projected by Type II are smaller, regardless of the environment or visual target. Table II shows errors of the 3D



(a) Scanpath of the 3D gaze of users #1.

(b) Scanpath of the 3D gaze of users #3.

Figure 11. Scanpath of the 3D gaze of users #1 and #3 to moving visual targets (without adjustments of visual target) in an environment with depth cues.

gaze projected by Type I and II polynomials. Significant differences were found when the accuracy of the 3D gaze obtained by each polynomial was analyzed by t-test, indicating that Type II can project the 3D gaze with higher accuracy ($t(19) = 2.640$, $p = 0.016$). Type I polynomials have a total of 45 coefficients, while Type II polynomials have only 22. Generally, the higher the number of coefficients produced the better the fitting. This applies to the 3D gaze to the visual targets for the gaze calibration. However, this is not inherently the accuracy of the 3D gaze as fitting of the polynomial to other visual targets is not guaranteed. Therefore, better results for 3D Gaze against the stationary visual targets projected using Type II polynomials are acceptable.

Table III shows the accuracies of the 3D gaze to the moving visual targets. Subject #2 was not able to correctly track the moving visual target hence the results for this subject were excluded. Two-way analysis of variance on these results revealed that no effects were found in both depth cue (F (1, 12) = 0.285, p = .603) and visual target adjustment (F (1, 12) =1.873, p = .196).

Figure 11 shows the scanpath of the 3D gaze of users #1 and #3 to moving visual targets (without adjustment) in an environment with depth cues. The scanpath of subject #1 extends to 1.5m in each direction of the visual target, while the scanpath of subject #3 extends to 2m. This result implies that our OST-HMD can measure 3D gaze in the personal space defined by Cutting (1997) [3]. Figures 12 and 13 show the scan path in Figure 11, grouped by height, and displayed from the top viewpoint. Observation of these scan paths shows that they are curved at locations close to the viewers.

Our experiments showed that the relative size of the 3D visual targets and the surrounding depth cues had no effect on the accuracy of the 3D gaze. However, it is important to consider the building of a polynomial for projecting the 3D gaze, and there is room for improvement in the polynomial.

VI. Conclusion

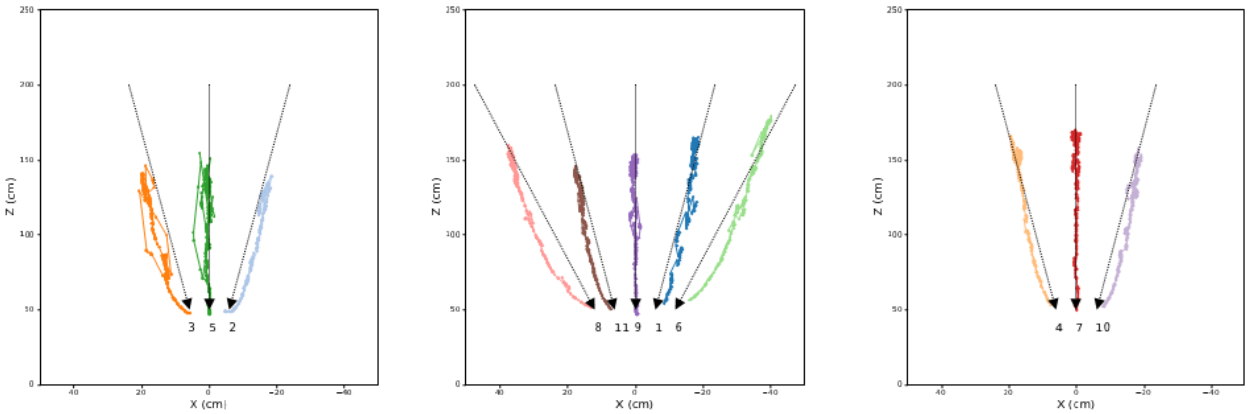
In this study, we conducted 3D gaze measurements in MR environments and confirmed that the relative size of the 3D visual targets and the surrounding physical environments did not affect the accuracy of the 3D gaze. Two types of polynomials were used to estimate the 3D gaze: the polynomial used by Kato and Prima (2021) and a newly constructed polynomial that constrains the pitch angle of the line-of-sight. The former is composed of 45 coefficients, while the latter has only 22 coefficients. Our experiments showed that we achieved higher accuracy in estimating the 3D gaze by using the polynomial consisting of these 22 coefficients. Using higher order polynomials may provide better eye calibration accuracy, but the accuracy achieved by this may not be inherently accurate.

We were concerned about the effect on the user's 3D perception in 3D visualization and interactive manipulation of medical images in the MR environment, however, we were unable to find these concerns in our current experimental results.

References

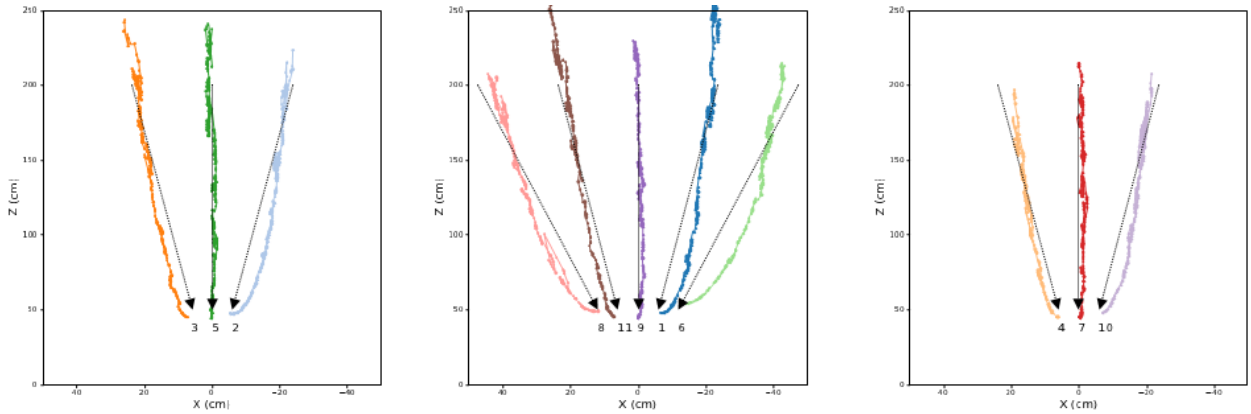
[1] K. Kato and O. D. A. Prima, “3D Gaze Characteristics in Mixed-Reality Environment,” eTELEMED 2021 : The Thirteenth International Conference on eHealth, Telemedicine, and Social Medicine, IARIA, pp.11-15, 2021.

[2] J. E. Cutting and P. M. Vishton, “Perceiving Layout and Knowing Distances : The Integration, Relative Potency, and Contextual Use of Different Information about Depth,” Percept. Sp. Motion, 22(5), pp. 69–117, 1995.



(a) Moving target #2, #3, #5 (b) Moving target #1, #6, #8, #9, #11 (c) Moving target #4, #7, #10

Figure 12. Top view of scanpath of the 3D gaze of users #1 to moving visual targets (without adjustments of visual target) in an environment with depth cues.



(a) Moving target #2, #3, #5 (b) Moving target #1, #6, #8, #9, #11 (c) Moving target #4, #7, #10

Figure 13. Top view of scanpath of the 3D gaze of users #3 to moving visual targets (without adjustments of visual target) in an environment with depth cues.

[3] J. E. Cutting, “How the Eye Measures Reality and Virtual Reality,” Behavior Research Methods, Instruments, & Computers, 29(1), pp. 27–36, 1997.

[4] A. Viguier, G. Clément, and Y. Trotter, “Distance Perception within Near Visual Space,” Perception, 30(1), pp. 115–124, 2001.

[5] S. Lee, X. Hu, and H. Hua, "Effects of Optical Combiner and IPD Change for Convergence on Near-Field Depth Perception in an Optical See-Through HMD," in IEEE Transactions on Visualization & Computer Graphics, 22(5), pp. 1540-1554, 2016.

[6] G. Singh, S. R. Ellis, and J. E. Swan, “The Effect of Focal Distance, Age, and Brightness on Near-Field Augmented Reality Depth Matching,” IEEE Transactions on Visualization Computing Graphics, 26(2), pp. 1385–1398, 2018.

[7] G. Ballestin, F. Solari, and M. Chessa, "Perception and Action in Peripersonal Space: A Comparison Between Video and Optical See-Through Augmented Reality Devices," 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 184-189, 2018.

[8] A. Vovk, F. Wild, W. Guest, and T. Kuula, “Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens,” In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Association for Computing Machinery, 209, pp. 1–9, 2018.

[9] J. P. Rolland and H. Fuchs, “Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization,” Presence Teleoperators Virtual Environment, 9(3), pp. 287–309, 2000.

[10] T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth, “Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments,” JVRB - Journal of Virtual Reality and Broadcasting, 5(16), 2008.

[11] E. G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci, “3D Gaze Estimation using Eye Vergence,” BIOSTEC, pp. 125-131, 2016.

[12] S. Weber, R. Schubert, S. Vogt, B. M. Velichkovsky, and S. Pannasch, “Gaze3DFix: Detecting 3D fixations with an ellipsoidal bounding volume,” Behavior Research Methods, vol. 50, no. 5, pp. 2004-2015, 2018.

[13] C. Elmadjian, P. Shukla, A. D. Tula, and C. H. Morimoto“3D gaze estimation in the scene volume with a head-mounted eye tracker,” Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18), No. 3, pp. 1-9, 2018.

[14] S. Z. Öney et al., "Evaluation of Gaze Depth Estimation from Eye Tracking in Augmented Reality," ACM Symposium on Eye Tracking Research and Applications, pp. 1-5, 2020.

[15] K. Kato, O. D. A. Prima, and H. Ito, "Development of Practical-used 3D Eye Tracker", IPSJ – Research Conference, 80th, 1, pp.133-134, 2018.

[16] M. Kassner, W. Patera, and A. Bulling. "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction," Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1151-1160, 2014.

[17] M. Kouchi and M. Mochimaru, "2008: Anthropometric Database of Japanese Head 2001," National Institute of Advanced Industrial Science and Technology, H16PRO-212.

Evaluations and Applications of Partial Body Joint Model in 3D Human Pose Estimation from Single Image

Yuta Ono, Oky Dicky Ardiansyah Prima, Kazuki Hosogoe
Graduate School of Software and Information Science, Iwate Prefectural University
152-52 Sugo Takizawa, Japan
e-mail: g236s001@s.iwate-pu.ac.jp, prima@iwate-pu.ac.jp, g231r026@s.iwate-pu.ac.jp

***Abstract*—Human pose estimation has been used to perform human motion analysis in widespread applications. Three-dimensional (3D) human pose estimation from single image has attracted much attention because of its ease of measurement. Methods of this approach have become more accurate with the introduction of deep neural networks. Most of these methods are trained to estimate the body joints of the whole human body. However, when a part of the body joints is obscured by the presence of other objects or the camera position and angle, the estimation accuracy of the overall body joints may be degraded. In this study, we attempt to experimentally construct a 3D human pose estimation model for partial body joints to accurately estimate the pose of a partially human body that can be visibly measured. To evaluate the performance of the proposed model, we construct a neural network model that estimates only the 3D position of the visible upper body joints, assuming that only those joints are visible. Our evaluations showed that the partial body joint model was more accurate in estimating the posture from frontal human images. However, there was no significant difference in the accuracy between our model and previous pose estimation model when the posture was estimated from images of people taken from extreme angles. Finally, we attempt to extend our model to a system for detecting deterioration in sitting postures to verify the effectiveness of the model.**

Keywords-3D human pose estimation; partial body joint; RGB-D camera; computer vision; sitting posture.

I. INTRODUCTION

Human pose estimation is a task that uses computer vision technology to estimate the location of body joints of a human body from a given image. Over the past five years, research in human pose estimation has shifted from 2D to 3D. 3D human pose estimation has attracted a significant interest from the scientific community. The technique has made significant progress by introducing methods based on deep learning. However, it still has some problems, such as depth ambiguities and lack of in-the-wild datasets. This study extends our previous work on the influence of occlusion on 3D human pose estimation [1].

3D human pose estimation has been applied in a wide range of fields including human-computer interaction, games, sports performance analyses, and other motion analyses [2][3]. Liu et al. (2013) extracted skeletal information of seven basic human actions using Microsoft Kinect v1 and classified them using k-means clustering and Hidden Markov Models (HMMs) [4]. Ono et al. (2021) used Microsoft Azure Kinect

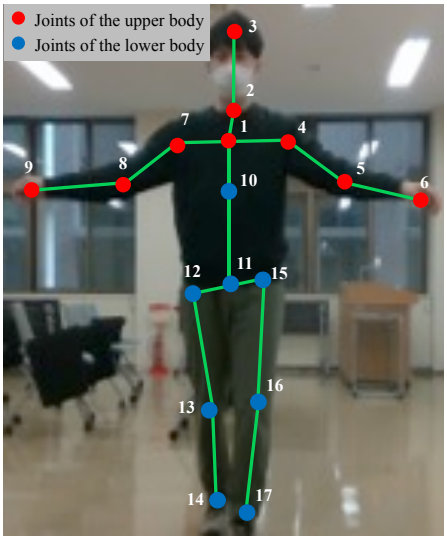


Figure 1. Body joints used in this study. 9 joints were used for the upper-body model, and 17 joints for the whole-body model.

[5] to measure the hand movements of pharmacists to monitor drug picking operations in pharmacies [6]. Their method uses finger landmark detection with MediaPipe [7] for more accurate detection of the picking tasks. 3D human pose estimation has also been used for a motion analysis to evaluate the effectiveness of rehabilitation. Prima et al. (2019) demonstrated the usability of 3D human pose estimation using a vision camera for measuring the range of motion of joints to promote self-rehabilitation by patients [8]. Their experiments show that the resulting 3D human pose estimated from a single image is more advantageous for estimating semi-occluded body joint locations than those estimated by a depth sensor.

3D human pose estimation methods can be broadly classified into two categories: a method using multiple cameras and a method using a single camera. Methods using multiple cameras are advantageous for depth measurement and occlusion avoidance. Ziegler et al. (2006) proposed a method to track an articulated upper body using four stereo cameras [9]. A point cloud of the synthesized body model was fitted to the measured 3D data using an iterative closest point (ICP) registration algorithm. Nakano et al. (2020) estimated the 2D human pose from images by OpenPose library [10] and estimated the 3D human pose based on triangulation of these 2D body joints [11]. However, these methods suffered from the difficulty of camera calibration. In contrast, 3D human pose estimation from a single camera is based on an estimation

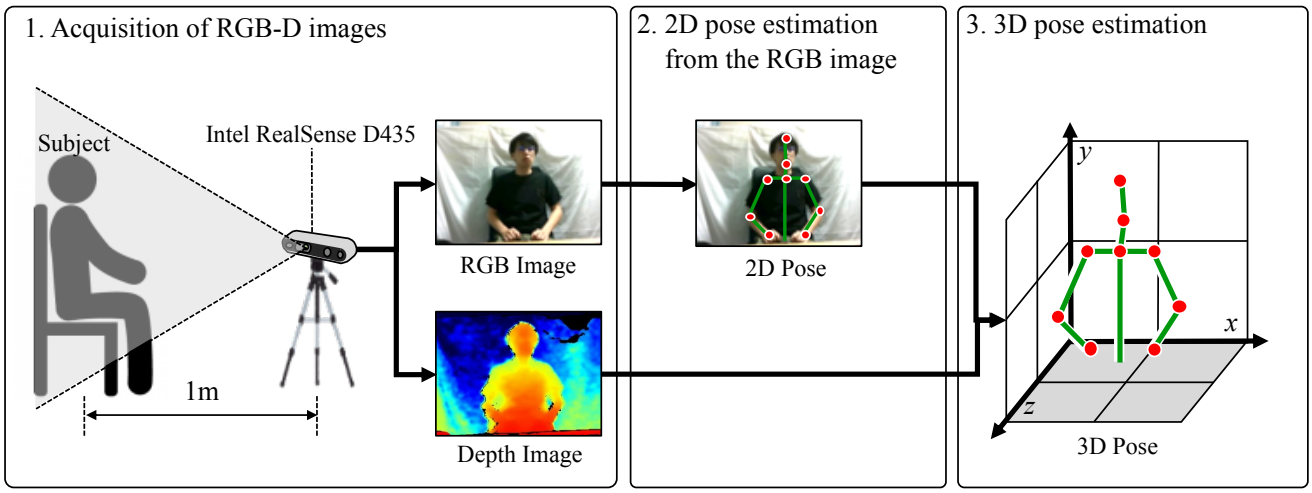


Figure 2. 2D and 3D human pose estimation using an Intel RealSense D435 to build the sitting posture dataset in this study.

method that uses 3D human pose datasets. Chen and Ramanan (2017) proposed a method to generate 3D human poses using a 3D human pose library consisting of pairs of 2D and 3D human poses [12]. Their results suggest that such a simple baseline should be used as a benchmark for future work in 3D human pose estimation. Martinez et al. (2017) constructed a relatively simple deep neural network that converts 2D human pose data to 3D human pose data [13]. Moon et al. (2019) used the correlation between 2D and 3D human poses to estimate the position and posture of the human body in real environment [14]. However, since the 3D human pose estimation from a single camera assumes that the whole-body joints are completely visible in the input image, if any part of the body joints is hidden, the estimation accuracy for the whole-body joints may be degraded.

Methods for 3D human pose estimation considering occlusion have been proposed. Vosoughi et al. (2018) proposed a deep Convolutional Neural Network (CNN) that regresses 3D human pose from an RGB image and a CNN that detects the presence of human body joints from an RGB image [15]. Sárándi et al. (2018) evaluated the robustness for occlusion in 3D human pose estimation using Human3.6M dataset [16] with synthetic occlusions [17]. However, these methods have been evaluated using only existing datasets, and have not been tested for the accuracy of 3D human pose estimation in the case where occlusion is caused in real world. In addition, these methods have not been tested for the case where self-occlusion occurs.

This study attempts to experimentally construct a partial 3D human pose estimation model to accurately estimate only the visible body joints of the human body even if some joints of the human body are hidden by other objects or self-occlusion. To verify the effectiveness of the model, we constructed a model that estimates only the body joints of the upper body and verified the estimation accuracy of the model using a modified Human3.6M dataset and two datasets originally created using RGB-D cameras. In addition, we developed a system for detecting deterioration in sitting posture using the model and verify the usability of the system.

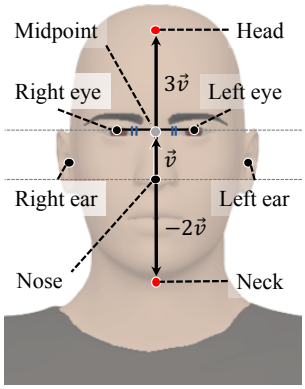


Figure 3. Calculation of Head and Neck points (red dots) using OpenPose face landmarks (black dots) in the sitting posture datasets.

This paper is organized as follows. Section II describes the methodology of constructing the 3D human pose estimation model for partial body joints and three human pose datasets to verify the estimation accuracy of the model. In Section III, we present our evaluation results evaluated using these datasets. Section IV describes how to construct a posture deterioration detection system using the model and verify the usability of the system. Finally, Section V summarizes the results of this study.

II. METHODOLOGY

In this study, we attempt to experimentally construct a partial 3D human pose estimation model to accurately estimate only visible body joints of the human body, even if some of them are hidden. Our model is constructed by improving the existing 3D human posture estimation model [13]. Figure 1 shows the nine upper body joints of the human body to be estimated by the partial 3D human pose estimation model in this study.

To evaluate the performance of our model, we compare its estimation accuracy with that of a whole-body joints model

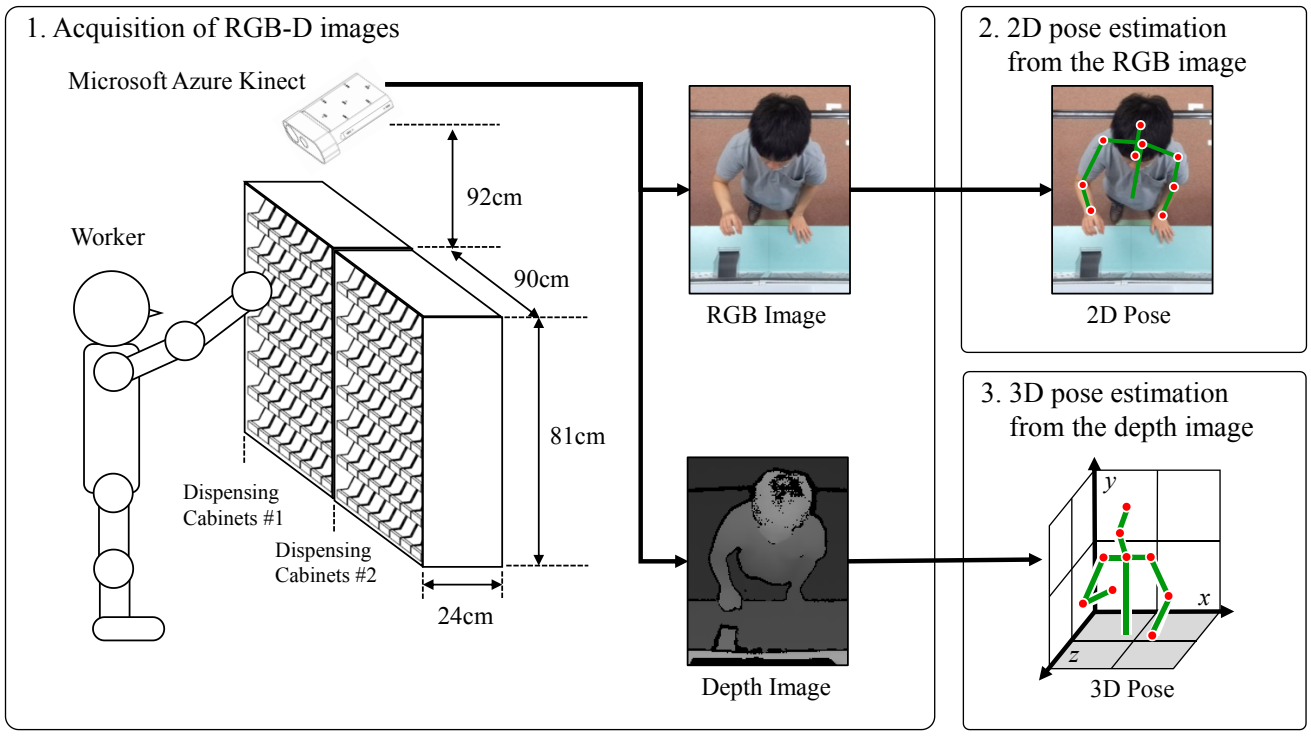


Figure 4. 2D and 3D human pose estimations using a Microsoft Azure Kinect to build the high-angle captured standing posture dataset in this study.

[13] using three different 3D human pose datasets. For convenience, the model of the body joints of the whole body is referred to as the "whole-body model" and the model of only the body joints of the upper body as the "upper-body model". The evaluation involves the following procedures. First, we simulate a scene in which only the upper body is visible using the Human3.6M dataset to evaluate the estimation accuracy of each model. Second, in order to evaluate the estimation accuracy of the 3D human pose of the upper body in the real world, two 3D human pose datasets, such as sitting postures and high-angle captured standing postures are constructed independently using RGB-D cameras.

A. Building a Partial 3D Human Pose Estimation Model

To train the upper-body model, we use the Human3.6M dataset. The dataset consists of 3.6 million 3D human poses and their corresponding images measured by 11 professional actors (5 females and 6 males) in 17 different daily activities. We adopted the method of Martinez et al. [13] to build the upper-body model. This method uses a relatively simple deep feed-forward neural network to efficiently estimate the 3D human pose of the whole body. Following Martinez et al., the model is trained from the measurement data of subjects 1, 5, 6, 7 and 8, and validated from the measurement data of subjects 9 and 11. The neural network receives the 2D human pose as input, and the input data is extended to 1024 dimensions by the fully connected layers. The weights of all the layers are initialized using the method of He et al. (2015) [18]. After that, the network performs batch normalization, Rectified Linear Unit (ReLU), dropout rate of 0.5, and a residual connection. The model is trained for 200 epochs with a learning rate of 0.001 and a batch size of 64.

B. Building Human Pose Datasets Using RGB-D cameras

We create two 3D human pose datasets, as ground truth data, using RGB-D cameras in real-world situations where only the upper body is visible due to the presence of other objects or due to the extreme capture angle and evaluate the performance of models using these datasets. RGB images and depth data of the participants in their postures are recorded.

1. Sitting Posture Dataset

An Intel RealSense D435 [19] was used to measure the sitting posture. The resolution of the RGB-D camera is 640×480px and the number of Frame Per Second (FPS) is 30. Figure 2 shows the procedure for generating human pose data in the sitting posture dataset. Participants were asked to sit 1m from the device and move their hands and bodies.

The procedure for generating 3D human pose data in a sitting posture is shown as follows. First, using the Intel RealSense D435 SDK, we calculated the depth value corresponding to each pixel of the measured RGB image. Then OpenPose library was used to estimate the 2D body joints from RGB images as shown in Figure 2. We also calculated head and neck points using OpenPose facial landmarks. Figure 3 shows calculations of these points. First, we calculated the 2D vector from nose to the midpoint of both eyes. Then the head point was obtained by extending this vector upward by three times its length. Similarly, the neck point was obtained by extending this vector downward by

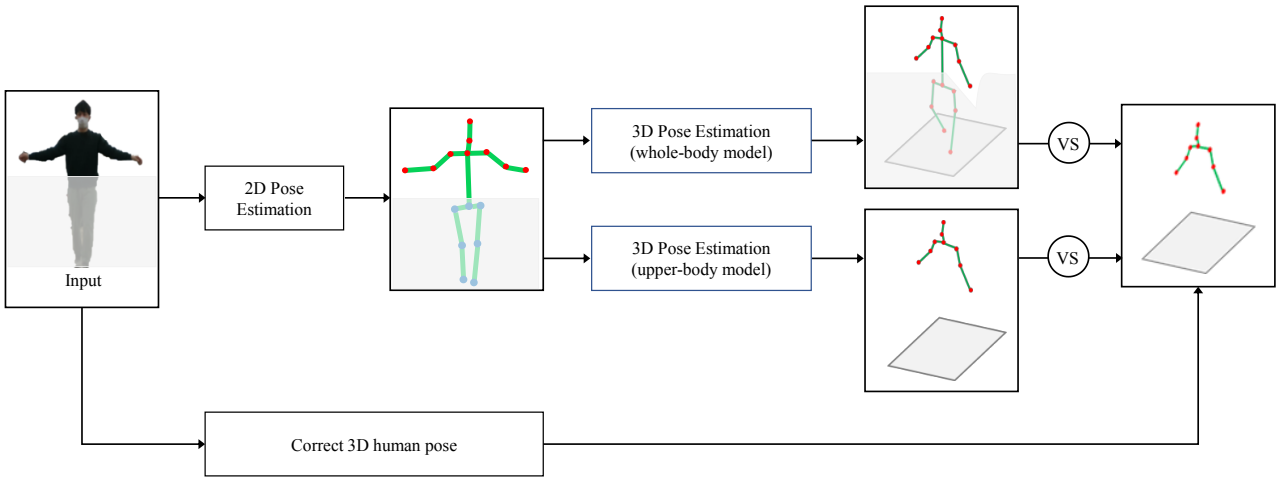


Figure 5. The process of evaluating the estimation accuracy of the upper-body model and the whole-body model in this study.

TABLE I. THE ESTIMATION ACCURACY OF THE UPPER-BODY MODEL AND THE WHOLE-BODY MODEL FOR THREE HUMAN POSE DATASETS.

No.	Joints	Human3.6M [cm]		Sitting Posture [cm]		High-Angle Captured Standing Posture [cm]	
		Whole-Body	Upper-Body	Whole-Body	Upper-Body	Whole-Body	Upper-Body
1	Thorax	16.3	5.0	11.4	7.5	8.8	9.6
2	Neck	3.8	2.7	18.2	14.5	16.3	11.2
3	Head	8.1	4.6	9.7	12.2	19.6	12.4
4	Left Shoulder	8.1	3.7	10.4	9.1	6.9	10.5
5	Left Elbow	9.1	4.5	10.3	9.2	14.2	11.8
6	Left Wrist	10.9	7.0	14.9	10.8	14.2	14.1
7	Right Shoulder	7.3	3.5	10.8	11.3	4.7	8.1
8	Right Elbow	9.0	4.7	10.3	10.9	12.3	11.0
9	Right Wrist	10.6	7.0	15.8	11.1	17.7	19.0
Mean (M)		9.24	4.74	12.42	10.73	12.74	11.97
Standard Deviation (SD)		3.363	1.464	3.064	2.009	5.038	3.132

twice its length. The corresponding 3D body joints might not be measured accurately due to missing depth values. In such a case, the Hampel filter [20] was applied to the time-series values to interpolate the concerned joints by computing the Median Absolute Deviation (MAD) over a specified range of analysis targets. For this study, the window size was set to 30. These 3D body joints are treated as ground truth data.

2. High-Angle Captured Standing Posture Dataset

To evaluate the estimation accuracy of the upper body model when the human body is measured from extreme angles, the motion of a worker picking drugs was captured from the ceiling. A Microsoft Azure Kinect was used to capture the motion of a worker. 2D body joints are estimated by OpenPose library from RGB images. In the standing posture dataset, the head point was obtained as the midpoint of both ears. The neck point was the midpoint of nose and neck estimated by OpenPose library. 3D body joints were measured using the Software Development Kit (SDK) of the Azure Kinect Body Tracking. The head point was derived from the midpoint of both ears and the neck point was derived from the midpoint of the nose and neck measured by the Azure Kinect SDK.

Figure 4 shows the procedure for generating human pose data in the standing posture dataset. Two dispensing cabinets were used to simulate the drug-picking environment. These cabinets can hold 63 shelves (7 rows by 9 columns). These cabinets were placed side-by-side at a height of 85.5 cm from the floor, and the Microsoft Azure Kinect was placed 92 cm above the dispensing cabinet. The resolution of the color camera is 1920×1080px and the angle of view is 90°×59°. The depth sensor has a resolution of 512×512px and the angle of view is 120°×120°.

The procedure for generating the dataset of the standing human posture is as follows. First, the subject stands in front of two dispensing cabinets and manipulates the shelves

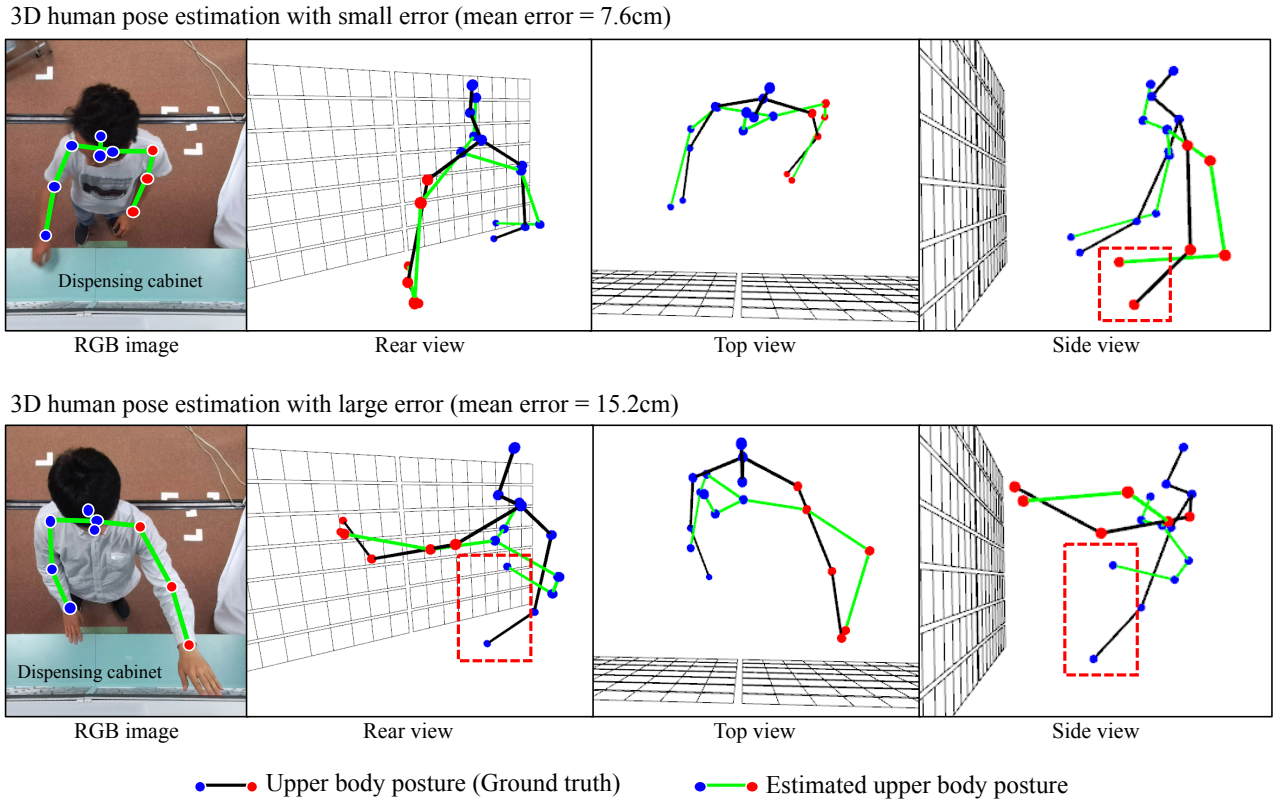


Figure 6. Examples of a 3D human pose estimated by the upper-body model on the standing posture dataset.

indicated by the experimenter. This operation is a series of movements from pulling out the target shelf to putting it back. The subject should operate the shelves on the left side with the left hand and the shelves on the right side with the right hand. After this series of operations, the subject returns to the initial standing position. The subject repeats the above procedure until all the shelves have been manipulated.

C. Evaluation

Using the Human3.6M, the sitting posture dataset, and the standing posture dataset, we evaluate the estimation accuracy of the upper-body model against the whole-body model. Figure 5 shows the process of evaluating the estimation accuracy of these models in this study. 3D human pose of the upper body is estimated using the 2D body joints from each dataset. For the upper-body model, all these body joints of the upper body were used, but for the whole-body model, the coordinates of the body joints corresponding to the lower body were set to (0,0). Finally, the error between the 3D human pose estimated by each model and the corresponding ground truth was calculated.

In this study, Procrustes Analysis, a shape-preserving Euclidean transform, is used to evaluate the differences in the pose data. This analysis eliminates the variation in movement, rotation, and scaling between the pose data while preserving the shape.

III. Result

We evaluated the difference in estimation accuracy between the whole-body model and the upper-body model using the 3D human pose datasets. For this evaluation, we randomly selected 548,800 human poses from the Human3.6M datasets, which were not used for training and validation of these models. For the sitting posture datasets, we generated 7,350 3D human pose from a subject using the Intel RealSense D435. For the standing posture dataset, we randomly selected 10,000 3D human pose from five subject's picking movement data.

Table 1 shows the estimation accuracy of the upper-body model and the whole-body model for the Human3.6M dataset, the sitting posture dataset, and the standing posture dataset. In the Human3.6M dataset, the error from the upper-body model was significantly smaller than that from the whole-body model ($M = 9.24$, $SD = 3.363$, $t(8) = 4.91$, $p < 0.001$), resulting in an improvement of about 4.5cm in estimation accuracy. Likewise, for the sitting posture dataset, the upper-body model improved the estimation accuracy by about 2 cm compared to the whole-body model ($M = 12.4$, $SD = 3.064$, $t(8) = 1.98$, $p < 0.05$). However, the error of the upper-body model was larger than that of the whole-body model at some body joints, such as head and right shoulder. The reason for this may be due to the accuracy of the constructed ground truth data. Finally, for

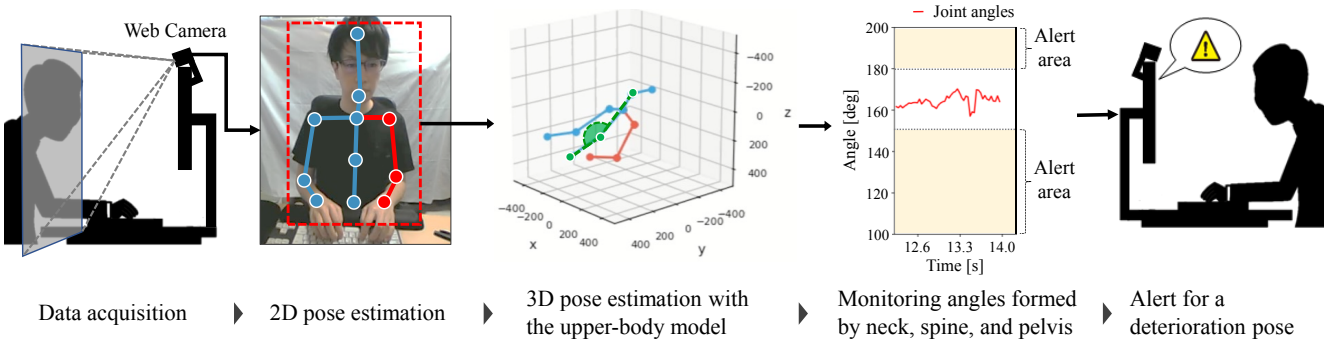


Figure 7. The illustration of the posture deterioration detection system using the upper-body model.

the standing posture dataset, the upper-body model improved the estimation accuracy by 0.8cm compared to the whole-body model, but the difference was not significant ($M = 12.7$, $SD = 5.038$, $t(8) = 0.63$, $p > 0.05$).

Figure 6 shows examples of a 3D human pose estimated by the upper-body model on the standing posture dataset. The top row shows an example where the estimation error is small for all joints of the upper body. The average error for the body joints of the upper body is 7.6 cm. However, the error is particularly large for the wrist joint. This reason for this error may be that the upper-body model is not sufficiently trained to estimate the body joints of the human body measured by the high-angle camera. The bottom row shows an example where the estimation error is large for all joints of the upper body. In this example, because the body joints estimated by the upper-body model differed significantly from the ground truth, the errors are particularly large for the left elbow and the right wrist joints. The dotted box in the figure indicates the body joint with large estimation error in each example.

IV. Detection of Deterioration in Sitting Postures

Workers tend to unconsciously hunch over while seated at a desk or other workstation. Such a posture may cause physical health problems because it puts a burden on the shoulders and hips. Techniques using pressure sensors [21] [22] and IMU sensors [23] have been proposed to detect such an improper posture. However, these techniques require these sensors to be attached to a human body or a chair.

In this study, we attempt to extend an upper-body model to a posture deterioration detection system to solve these problems. This system estimates the 3D human pose of a person using an upper-body model from RGB images measured by a web camera. The system then detects the deterioration of posture based on the body angle calculated from the estimated 3D human pose. In addition to the body joints of the upper body shown in Figure 1, the upper-body model estimates the Spine and the Pelvis. Our system detects deterioration in sitting posture without contact and encourages improvement in posture.

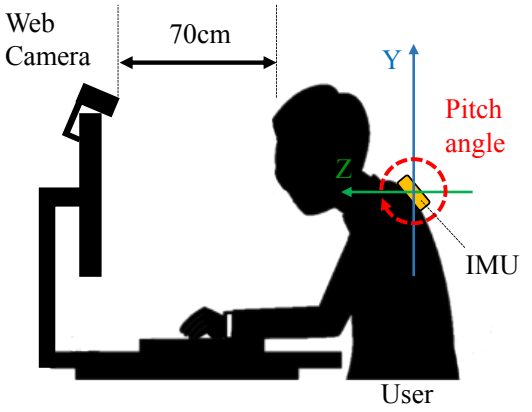
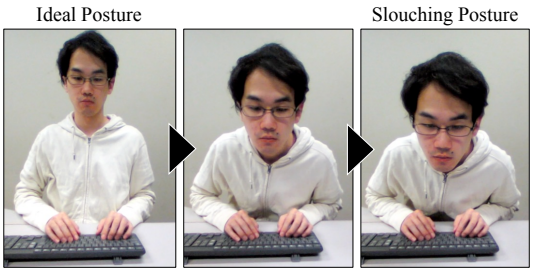
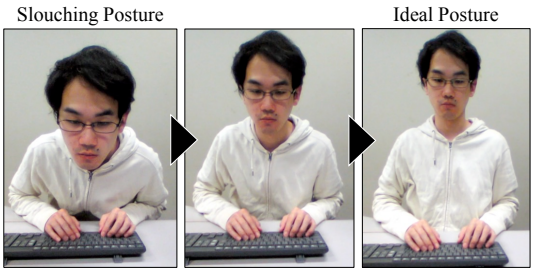


Figure 8. The placement of the web camera and the IMU.

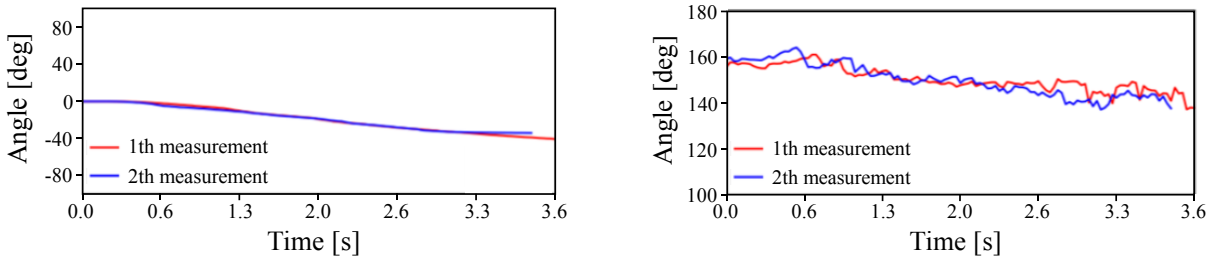


(a) Forward leaning movement



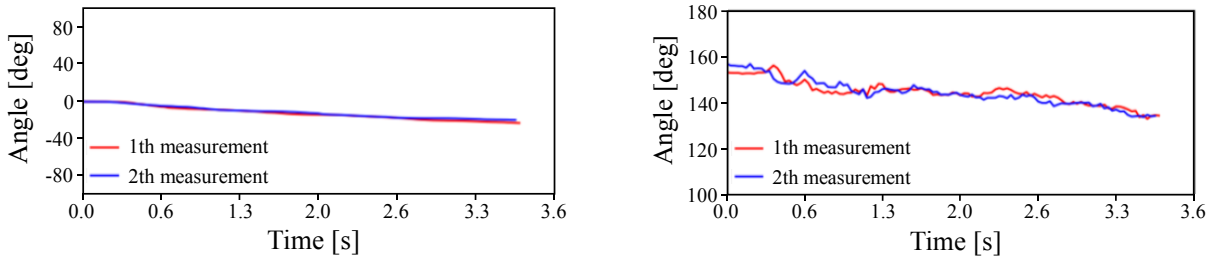
(b) Raising movement

Figure 9. Two motions measured for evaluating the posture deterioration detection system in this study.



(a-1) The IMU measurement (a-2) The proposed system

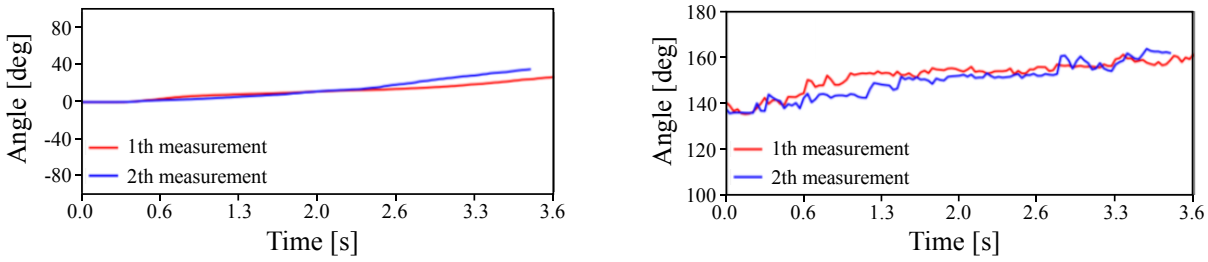
(a) Subject A



(b-1) The IMU measurement (b-2) The proposed system

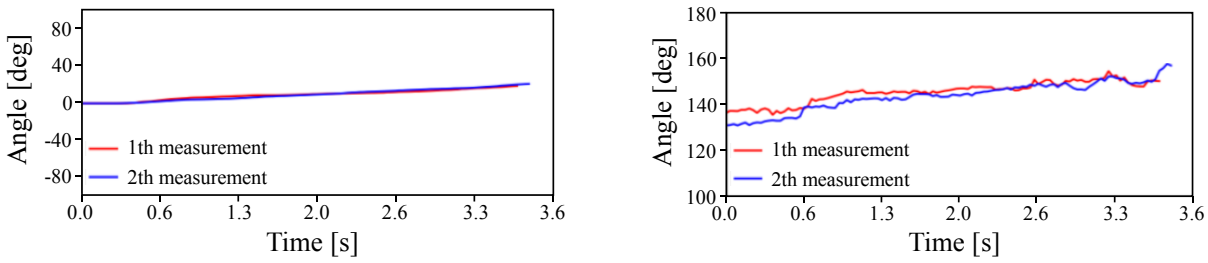
(b) Subject B

Figure 10. The changes in the body angle of two subjects measured by the proposed system and the IMU during forward leaning movement.



(a-1) IMU Measurement (a-2) The proposed system

(a) Subject A



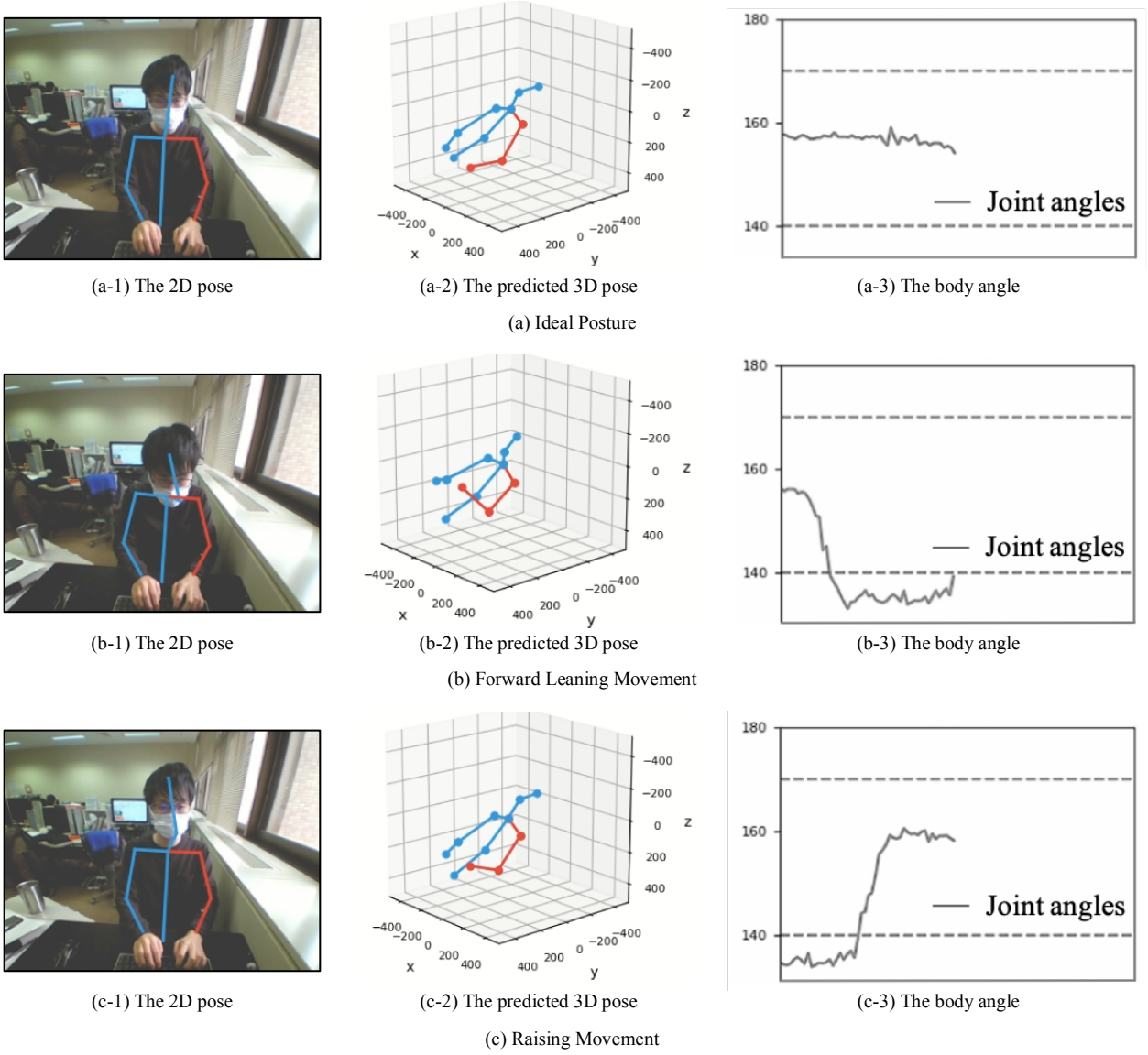
(b-1) IMU Measurement (b-2) The proposed system

(b) Subject B

Figure 11. The changes in the body angle of two subjects measured by the proposed system and the IMU during raising movement.

A. Posture Deterioration Detection System

Figure 7 shows a schematic diagram of the posture deterioration detection system using the upper-body model. First, we use a web camera to capture the frontal view of a worker at a desk. Next, the 2D body joints of the worker is estimated from the RGB images using the OpenPose library. Then, we estimate the 3D human pose of the upper-body from these body joints using the upper-body model. Finally, we detect the deterioration of the posture based on the calculation of the body angle calculated using the estimated 3D body joints.



(a-1) The 2D pose (a-2) The predicted 3D pose (a-3) The body angle

(a) Ideal Posture

(b-1) The 2D pose (b-2) The predicted 3D pose (b-3) The body angle

(b) Forward Leaning Movement

(c-1) The 2D pose (c-2) The predicted 3D pose (c-3) The body angle

(c) Raising Movement

Figure 12. The working example of our system.

Our system focuses on the changes in the body angles formed by neck, spine, and pelvis to detect the deterioration of the posture. In this system, the user decides the upper and lower limits of the body angle to be used for judging the deterioration of posture. When the body angle measured by the system exceeds the threshold value, the system detects the posture as a deteriorated posture.

B. Evaluation of Posture Deterioration Detection System

To evaluate the usability of the posture deterioration detection system, we evaluate the system from two perspectives: whether the system can measure the change in body angle appropriately and whether the system can properly detect a deteriorated posture based on the threshold value set from the measured change in body angle.

For this evaluation, we assume that the system is used during desk work and detects posture deterioration in the sitting posture. We use a 57cm×43.5cm display and place it on a desk. The web camera is Buffalo BSW200MBK. This camera is placed at the center of the upper part of the display to measure the subject from the front. The angle of view of the camera is 120°×67°, the resolution is 640×480px, and the frame rate is set to 30FPS.

To evaluate the angle of the subject's body, an Inertial Measurement Unit (IMU) is used as reference data. We use the Adafruit BNO05 (100Hz) IMU, which is attached to the back of the subject's neck. Figure 8 shows the placement of the web camera and the IMU. Five subjects participated in this evaluation.

The procedure for evaluating the posture deterioration detection system is as follows. Figure 9 shows the two

motions measured in this evaluation. First, the subject is seated 70cm from the web camera. At this point, the subject is instructed to put his hands on the desk, raise his head, and straighten his back. We refer to this posture as the ideal posture. Next, the subject takes 5 seconds from the ideal posture to tilt the upper body forward at a certain speed. After that, the subject returns to the ideal posture at a certain speed for 5 seconds. This series of motions is performed twice. In this evaluation, we refer to the movement from the ideal posture to the forward leaning posture as the forward leaning movement. In addition, we refer the movement from the forward leaning posture to the ideal posture as the rising movement. The body angle is measured by the web camera and the IMU. Finally, we calculate the correlation between the changes in the body angles measured by the web camera and the IMU.

The correlation coefficients of the body angle changes measured by the web camera and the IMU were 0.933 on average. Figures 10 and 11 show the changes in the body angle measured by the web camera and the IMU during forward leaning and rising movements. Figure 12 shows the working example of our system. As shown in Figure 12(b), when the worker's head moves forward and the posture becomes hunched, the body angle changes significantly. The dotted line in the figure indicates an empirical threshold that indicates the acceptable range of an appropriate sitting posture. When the seating posture deteriorates, the measured angle does not fall within this range.

V. Conclusion and Future Work

The purpose of this study is to accurately estimate the body joints of the human body that can be measured if a part of body joints is hidden due to the presence of other objects or the position and angle of the vision camera. For this purpose, we developed a partial 3D human pose estimation model that estimates the upper body joints, assuming that only the upper body is visible. We evaluated the performance of the model using three different 3D human pose datasets to examine the estimation accuracy of the model. And we also attempted to extend the model to a system for detecting deterioration in sitting posture.

To verify the estimation accuracy of the model, we evaluated the estimation accuracy of the model against the whole-body model using three different 3D human pose datasets. At first, we conducted an evaluation using the Human3.6M dataset. Next, we constructed two datasets independently using RGB-D cameras and evaluated the estimation accuracy of both models to confirm the estimation results of 3D human pose with occlusion in the real world. The first dataset is a sitting posture dataset, which was constructed to evaluate the pose of the human body when occlusion by other objects occurs. The second dataset is a high-angle captured standing posture dataset. This dataset was constructed to evaluate the estimation accuracy of the pose of a human body measured by a camera from extreme angles. The results show that the proposed model can estimate the 3D human pose more accurately from frontal human images than the whole-body model. However, when the human body was measured from extreme angles, there was no significant difference in the estimation accuracy between the two models. The reason for this is that the model did not learn sufficiently about the pose of the 3D human body from extreme angles.

We attempted to extend the partial 3D human pose estimation model to a posture deterioration detection system for sitting posture in order to promote the improvement of the sitting posture. In this system, the posture of the person is measured using a web camera, and the pose of the upper body is estimated by the model. The system then calculates the body angle to detect the deterioration of the posture. We confirmed that the system can measure changes in the body angle and is effective in detecting posture deterioration in a sitting posture. In addition, the system has the advantage of being able to detect posture deterioration without contact using only a widely used web camera.

In the future, we will construct other partial body joint models, such as the left half of the body and the lower half of the body and verify the estimation accuracy for each model. In addition, we will study the improvement of the model to increase the accuracy of human pose estimation from extreme measurement angles. Furthermore, we will improve the reliability of posture deterioration detection to build a more practical posture deterioration detection system.

References

[1] O.D.A. Prima and K. Hosogoe, "3D Human Pose Estimation of a Partial Body from a Single Image and Its Application in the Detection of Deterioration in Sitting Postures," The Thirteenth International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED2021, pp. 1-5, 2021.

[2] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3D Human Pose Estimation: A Review of the Literature and Analysis of Covariates," Computer Vision and Image Understanding, vol. 152, pp. 1-20, Nov. 2016, doi: 10.1016/j.cviu.2016.09.002.

[3] J. Wang et al., "Deep 3D human pose estimation: A review," Computer Vision and Image Understanding, vol. 210, pp. 1-21, Sept. 2021, doi: 10.1016/j.cviu.2021.103225.

[4] T. Liu, Y. Song, Y. Gu, and A. Li, "Human Action Recognition Based on Depth Images from Microsoft Kinect," 2013 Fourth Global Congress on Intelligent Systems, vol. 1, pp. 200-204, 2013, doi: 10.1109/GCIS.2013.38.

[5] Microsoft Azure, "Azure Kinect DK," https://azure.microsoft.com/en-us/services/kinect-dk/ [retrieved: Nov, 2021]

[6] Y. Ono and O.D.A. Prima, "Assessment of Drug Picking Activity using RGB-D Camera," The Fourteenth International Conference on Advances in Computer-Human Interactions, ACHI 2021, pp. 6-11, 2021.

[7] F. Zhang et al., "MediaPipe hands: on-device real-time hand tracking," CVPR Workshop on Computer Vision for Augmented and Virtual Reality, pp. 1-5, Jun. 2020, arXiv:2006.10214.

[8] O.D.A. Prima et al., "Evaluation of Joint Range of Motion Measured by Vision Cameras," International Journal on Advances in Life Sciences, 11, 3 & 4, pp. 128-137, 2019.

[9] J. Ziegler, K. Nickel, and R. Stiefelhagen, "Tracking of the Articulated Upper Body on Multi-View Stereo Image Sequences," Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 1-8, 2006, doi: 10.1109/CVPR.2006.313.

[10] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Shikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using

Part Affinity Fields," arXiv preprint, pp. 1-14, 2018, arXiv:1812.08008v2.

[11] N. Nakano et al., "Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras," Front. Sports Act. Living, May. 2020, doi: 10.3389/fspor.2020.00050.

[12] C. Chen and D. Ramanan, "3D Human Pose Estimation = 2D Pose Estimation + Matching," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7035-7043, 2017.

[13] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation," arXiv preprin, pp. 1-10, 2017, arXiv:1705.03098.

[14] G. Moon, J. Y. Chang, and K. M. Lee, "Camera distance-aware top-down approach for 3d multi-person pose estimation from a single RGB-image," pp. 1-15, Aug. 2019.

[15] S. Vosoughi and M. A. Amer, "Deep 3D Human Pose Estimation Under Partial Body Presence," 2018 25th IEEE International Conference on Image Processing (ICIP), Oct. 2018, doi: 10.1109/ICIP.2018.8451031.

[16] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," IEEE Transaction on Pattern Analysis and Machine Intelligence, 36, 7, pp. 1325-1339, Dec. 2014, doi: 10.1109/TPAMI.2013.248.

[17] I. Sárándi, T. Linder, K. O. Arras, and B. Leibe, "How Robust is 3D Human Pose Estimation to Occlusion?," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1-5, Aug. 2018, arXiv:1808.09316.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification," Proceedings of the IEEE International Conference on Computer Vision, pp. 1026-1034, 2015.

[19] Intel Corporation, "Intel RealSense Depth Camera D435," https://www.intelrealsense.com/depth-camera-d435/ [retrieved: Nov, 2021]

[20] F. Hampel, "The Influence Curve and Its Role in Robust Estimation," Journal of the American Statistical Association, Vol. 69, pp. 383-393, Jun. 1974, doi: 10.2307/2285666.

[21] B. Mutlu, A. Krause, J. Forlizzi, C. Guestrin, and J. Hodgins, "Robust, Low-cost, Non-intrusive Sensing and Recognition of Seated Postures," Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 149-158, Oct. 2007, doi: 10.1145/1294211.1294237.

[22] L. Martins et al., "Intelligent Chair Sensor – Classification and Correction of Sitting Posture," XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, pp. 1489-1492, 2014, doi: 10.1007/978-3-319-00846-2_368.

[23] A. Petropoulos, D. Sikeridis, and T. Antonakopoulos, "SPoMo: IMU-based Real-time Sitting Posture Monitoring," 2017 IEEE 7th International Conference on Consumer Electronics, pp. 5-9, Sept. 2017, doi: 10.1109/ICCE-Berlin.2017.8210574.

Prediction of Authors' Personality Types and Traits in Modern Greek Essays Using Stylometric Features

Gagiatsou Sofia
Department of Linguistics, School of Philosophy
National and Kapodistrian University of Athens
Athens, Greece
e-mail: sgagiats@phil.uoa.gr

Markopoulos Georgios
Department of Linguistics, School of Philosophy
National and Kapodistrian University of Athens
Athens, Greece
e-mail: gmarkop@phil.uoa.gr

Mikros George
College of Humanities and Social Sciences
Hamad Bin Khalifa University
Doha, Qatar
e-mail: gmikros@hbku.edu.qa

Abstract—**We present a study focused on the prediction of the author's personality based on natural language processing techniques applied to essays written in Modern Greek by high-school students. Each writer has been profiled by filling in two personality questionnaires, one based on the typology of Carl Jung and the other based on the Model of Five Factors. In addition, personality prediction is being discussed under the general research framework of author profiling by examining the effectiveness of several stylometric features to predict students' personality types. The feature set we employed was a combination of the word and sentence length, the most frequent part-of-speech tags, most frequent character/word bigrams and trigrams, most frequent words, as well as hapax/dis legomena. Since personality prediction represents a complex multidimensional research problem, we applied various machine learning algorithms to optimize our model's performance after extracting the stylometric features. We compared nine machine learning algorithms and ranked them according to their cross-validated accuracy. The best results in predicting the Jung's Typology types were obtained by the Naive Bayes algorithm. In contrast, for the prediction of personality features based on the Five Factors Model, the Generalized Linear Model (Binomial method) algorithm prevailed. According to the personality classification based on the Jung Typology Test, the author's personality prediction accuracy reached 80.7% on Extraversion, 79.9% on Intuition, 68.8% on Feeling, 75.7% on Judging, according to the personality classification. In the Big Five personality classification, the prediction accuracy reached 85.9% on Openness, 71.2% on Conscientiousness, 67.6% on Extraversion, 70.2% on Agreeableness, and 65.6% on Neuroticism. The reported results show a competitive approach to the personality prediction problem. Furthermore, our research revealed new combinations of stylometric features and corresponding computational techniques, giving interesting and satisfying solutions to the author's personality prediction problem for Modern Greek.**

Keywords-Author profiling; stylometry; Personality prediction; Jung Typology Test; Big Five model; corpus processing; computational stylistics; machine learning.

I. INTRODUCTION

Authorship identification represents one of the emerging text mining fields at the intersection of Machine Learning, Information Retrieval, and Natural Language Processing. Under the stylometric framework, the author's identity is a multidimensional construct based mainly on writing patterns scattered across multiple linguistic levels and expressed quantitatively. The specific research domain splits into three subdomains: attributing a text to a particular author among a finite set of authors (Authorship Attribution), attributing a text to an author that does not belong to a closed group (Authorship Verification), and specifying the author's metadata such as demographic and psychological traits of the author (Authorship Profiling), including gender, age, personality [1], etc.

Language as a communication mechanism denotes the diversity of every individual. Therefore, the quantitative study of linguistic features can lead to predictions regarding the individual's character. The subject of Computational Personality Prediction (CPP) through natural language processing techniques constitutes a relatively new research field with many applications.

One critical application domain of this field is Forensic Linguistics. Criminals can be identified by the way they write. Moreover, conclusions can be drawn regarding their personalities and the way they think. The example of identifying students' personality that carry guns and participate in school shootings is typical [2]. CPP can highlight their psychological traits, which can be exploited in successfully identifying potential perpetrators.

Apart from the apparent contribution provided in Behavioural Psychology by connecting personality traits to human behaviour, CPP can also function in many other fields. For instance, companies utilize personality analysis of users/consumers in the marketing domain to adopt effective recruitment techniques and customer service techniques. Even in human resources management, predicting the personality can affect or facilitate the selection and determine the eligibility of candidates for a particular job. Moreover, based on the user's personality, dialogic systems can be customized and brought closer to users' temperament making interaction more effective and satisfying.

Another vital analysis domain where automatic personality prediction is used is education. For example, by analyzing students' writings, talented students or students with difficulties could be recognized and thus receive adaptive teaching, addressing the appropriate cognitive level for each group.

One of the most crucial issues in CPP research is developing appropriate linguistic resources enriched with the author's personality metadata. Unfortunately, these resources are challenging to create due to the increased level of manual interaction with the authors and the various privacy and ethical considerations linked with administering psychometric questionnaires to many individuals.

Another issue is that most Natural Language Processing (NLP) tools specialized in psychometric text profiling support only English. Therefore, research in other languages should be done by developing specialized dictionaries and other supporting linguistics resources from scratch (see, for example, the case of Linguistic Inquiry and Word Count-LIWC [3]).

To cover the above-mentioned research gaps, we performed the first CPP study in Modern Greek focused on high-school students. For this reason, we developed a model for predicting students' personality based on Jung's taxonomy and the model of Big-Five factor markers by analyzing their term-essays and applying various machine learning methods to rich document representation based on several stylometric features.

The rest of this paper is organized as follows. In Section II we provide an overview of previous work on personality prediction. Section III describes our researching methods. In Section IV we present the research results. We summarize our findings and discuss future work in Section V.

II. Literature Review

This section presents the two personality questionnaires used to profile the writers (Carl Jung's and Isabel Briggs Myers' Personality Type Questionnaire and Big-Five Personality Test). Then, we review the findings of studies in the field of CPP from the text.

A. Carl Jung's and Isabel Briggs Myers' Personality Type Questionnaire

Research in the field of personality prediction uses Carl Jung's and Isabel Briggs Myers' personality type theory [4][5] or the Five-factor Model of Personality [6], which are the two most utilized personality models, to profile the participating authors. Therefore, the literature review presented in this section is referred to associated research, which involves the above-mentioned personality questionnaires since our students have been profiled with these tests.

According to Jung's theory of psychological types [4], people can be characterized by

- their preference of general attitude as Extraverted (E) or Introverted (I), which signifies the source and direction of a person's energy expression.
- their preference of one of the two functions of perception as Sensing (S) or Intuitive (N) represents how someone perceives information.
- their preference of one of the two functions of judging as Thinking (T) or Feeling (F), which describes how a person processes information.
- their orientation to the outer world as Judging (J) or Perceiving (P), which reflects how a person implements the information he/she has processed.

The Jung Typology Test classifies psychological personality differences in four dichotomies that yield 16 different combinations or personality types. Each personality type can be assigned a 4-letter acronym of the corresponding combination of preferences: ESTJ, ISTJ, ENTJ, INTJ, ESTP, ISTP, ENTP, INTP, ESFJ, ISFJ, ENFJ, INFJ, ESFP, ISFP, ENFP, INFP.

B. Big-Five Personality Test

One of the most widely accepted personality theories in psychology is the Five-Factor model. According to the Five-Factor Model of Personality, most human personality traits can be boiled down to five broad dimensions of personality, regardless of language or culture. There has been much research on how people describe others, and five major dimensions of human personality emerged. They are often referred to as the OCEAN model of personality because of the acronym from the names of the five dimensions. Openness to Experiences, Conscientiousness, Extraversion, Agreeableness, and Neuroticism are the five most essential personality traits [7]. More specifically:

- Openness to Experience

High scorers tend to be original, creative, curious, complex; Low scorers tend to be conventional, down to earth, have narrow interests, be uncreative.

- Conscientiousness

High scorers tend to be reliable, well-organized, self-disciplined, careful; Low scorers tend to be disorganized, undependable, negligent.

- Extraversion

High scorers tend to be sociable, friendly, fun-loving, talkative; Low scorers tend to be introverted, reserved, inhibited, quiet.

- Agreeableness

High scorers tend to be good-natured, sympathetic, forgiving, courteous; Low scorers tend to be critical, rude, harsh, callous.

- Neuroticism

High scorers tend to be nervous, high-strung, insecure, worrying; Low scorers tend to be calm, relaxed, secure, hardy.

C. Personality and Language

The way a person uses language as a communication code reveals much information for his/her personality. The selection of specific morphological, syntactic structures and lexical choices can indicate his/her age, gender, social class, and feelings. Moreover, we can understand whether the speaker or author of a text is extraverted, emotional, or distant. So, a critical element that needs to be examined is the relationship between personality and language.

In general, the dominant opinion is that personality affects and directs our behavior, thoughts, feelings, interpersonal relationships, and of course, language production. People speak and write in different ways, even if they want to express the same content. The language user chooses the appropriate level of speech depending on the specific instance of linguistic communication, thus shaping a personalized way of speaking or writing. Researchers in this field support that every human has a characteristic way of using the language, i.e., a kind of authorial fingerprint [8]. Since the idiolect is constructed through the selective use of specific linguistic elements and their differentiated usage frequency, we can infer that also a correlation between personality traits and language features, such as lexical categories, n-grams is evident.

The above is confirmed by current research; [9] supported that language reveals each person's temperament and investigates how it is linked with his/her linguistic individuality. [10] emphasizes that all linguistic levels (phonology, morphology, syntax, semantics, pragmatics) affect the message recipient. Research shows that personality traits impact each person's language production [11]. [12] points out that personality is projected through language, but that personality may also become perceivable to the recipient through language. Moreover, he mentions that different personality traits affect different levels of language production. [13] talks about the psychological aspect of language and focus on the choice of words by the language user as an indicative element of its character. Social psychologists have pointed out that the use of words, intonation, accent, and other language elements reveal their social, financial, and psychological position [14].

Although we perceive the importance of the connection between language and personality traits of the speaker or the author, the field has not been studied sufficiently, as most research focuses on verbal speech and the trait of Extraversion. According to [15], this is due to paralinguistic elements of the verbal speech, such as accent and intonation, as well as the fact that speech between family members and friends from a sociolinguistic point of view offers more useful linguistic data, since it is more spontaneous. Finally, Extraversion as a trait is more easily recognizable in somebody's speech, and therefore in combination with the above, research has focused on identifying the language features denoting this speaker's personality trait; as a result, it has been studied more than the other traits, both in the Five-Factor model as well as in Jung's typology. Finally, the dominant language in written data is English; this makes the comparative study of findings in other languages more difficult.

Our research attempts to cover the void in this field by creating specialized corpora and utilizing natural language processing techniques in order to research all types and all traits of both personality theories, and thus showing that the relation between language and personality can be determined computationally.

D. Personality Research from Text

We briefly present previous research that involves either Jung Typology Test or Big-Five Personality Test in the author's personality prediction task from the text.

1) Jung Typology Test

One of the first studies related to the author's personality prediction problem [16] defined the research problem as a text categorization task. They developed a corpus consisted of essays written in Dutch by 145 students (BA level). By selecting syntactic features and training machine learning algorithms, the experiments in personality prediction suggested that the personality dimensions Introverted-Extraverted and Intuitive-Sensing) can be predicted accurately.

CPP studies have also expanded to social media texts with an emphasis on Twitter. A study for predicting Twitter users' personality type [17] showed that the classifier's performance on training data was quite good. Still, the classifier failed to achieve satisfying results for the test data. Another study [18] describes a logistic regression classifier's training process to predict each of the four dimensions of Jung Typology. Their results showed that linguistic features are the most predictive features. Although they successfully distinguished between the personality dimensions Introverted-Extraverted and Feeling-Thinking, the other two dimensions were hard to predict.

In a study of a multilingual corpus of tweets [19], based on six languages (Dutch, German, French, Italian, Portuguese, and Spanish), the researchers extracted the most frequent word and character n-grams. Their results confirmed the findings of the previous work in that particular personality distinctions could be predicted from social media data with success. In another study focused on tweets [20], the researchers used a Naive Bayes classifier achieving 80%

accuracy for Introverted-Extraverted and 60% for the other dimensions.

CPP has also being applied to languages with a different graphemic organization compared to Western languages. For example, in [21], researchers investigate the personality prediction of Twitter users in Japanese and conclude that the textual information of user behaviors is more valuable than the users' cooccurrence behavior information such as the likes. In this study, the problem of author personality prediction was treated as a set of binary classification tasks using Support Vector Machines.

2) Big-Five Personality Test

Another study [22], which also treated personality prediction as a classification problem, has been conducted using student essays data. The corpus consisted of essays written by 198 psychology undergraduates over twenty minutes expressing thoughts and feelings. Each writer has been profiled by filling in a questionnaire testing the "Big Five" personality dimensions. The researchers focused on two of the Big Five traits, Extraversion and Neuroticism. Style and content features were extracted, and they concluded that style features provide a significant amount of information about personality.

In [23], authors developed classification, regression, and ranking models to recognize Big Five personality traits. They extracted a set of linguistic and psycholinguistic features from essays written by 2,479 psychology students, who were told to write whatever came through their minds for 20 minutes. The LIWC lexicon provided 88 word categories with syntactic and semantic information, while the Medical Research Council (MRC) Psycholinguistic Database [24] was used to extract 14 features. These features were used to train machine learning algorithms. The LIWC features outperformed the MRC features for every trait, and the LIWC features on their own always perform slightly better than the full feature set.

Using a publicly available dataset [11] consisting of essays, the authors of [25] developed a personality prediction model. They used psycholinguistic indices and language embeddings as features. Their results showed that language embeddings consistently outperform conventional psycholinguistic features.

In recent years, CPP studies have focused on corpora of social network data written in English and other languages. One of the most successful research initiatives in this area is the Author Profiling Task organised at PAN 2015. The specific task aimed to identify Twitter users' personality traits considering multilingual data (English, Spanish, Italian, and Dutch) [26].

III. CORPUS

To test our research hypothesis, that is, whether it is possible to detect personality traits of the authors of written Modern Greek texts, it is necessary to have a corpus of Modern Greek texts and at the same time to connect each author of these texts to a psychological profile. Due to the lack of such material, the first step was to collect primary textual data from native speakers of Modern Greek. In particular, the corpus that we developed consists of essays of 198 high school students and comprises 250.000 words in total. It is balanced in size (number of words per student) and students' demographics (gender and age).

The participating students of three different high schools were asked to write three essays to achieve our goal of collecting at least 1,000 words from each student. The task was voluntary, lasted three school years, and the writing was held in the classroom. The experiment was repeated three times at different periods. The authors had to write spontaneously and continuously for 60 minutes an essay. The volunteers were many more than 198, but their data have been ignored because they did not provide in their linguistic production the required text size. The mean length of the essays was 1,255 words. The topics, which were not given in advance, were related to the benefits of art, the role of school in raising environmental awareness, and fighting against child labor. Finally, since the provided texts were handwritten, we had to digitize them by manually typing all of them.

IV. METHODOLOGY

The following section describes the approach used to predict the personality types of students.

A. Approach

In the literature, two approaches stand out for an automatic author's personality prediction. In a bottom-up approach, personality labels are predicted from linguistic features that are being extracted from the corpora used using standard NLP document representations (e.g., Bag-of-Words - BoW models, etc.) [27]-[29]. In a top-down approach, instead, specialized dictionaries with custom entries are used to check the potential correlation with personality traits [30]-[32]. Both approaches have advantages, as well as restrictions. Therefore, modern techniques are oriented towards hybrid methods that combine the use of a dictionary with extended document representations trained on machine learning algorithms to exploit the best from both approaches, i.e., speed and precision, respectively. In this study, we followed the bottom-up approach, which among other benefits explained above, is also language-independent.

B. Feature extraction

The features used in our research can be considered part of a broader feature set characterized as stylometric, i.e., models quantitatively the text's style. The linguistic features that have been used previously as stylometric indices are numerous. They increase continuously and belong to the whole range of linguistic levels. Stylometric features are compact, information-rich signaling linguistic devices. They are correlated with many different textual functions and carry multilevel information related to both the author's identity and his/her metadata. In CPP, stylometric features can unchain the

hidden link between linguistic production and its correlation with specific personality types. This is because our personality traits are defining and be defined by our socio-cognitive and psychological conditions. In that sense, aspects of our linguistic behavior reflect these personality traits indirectly and amplify them using identity perceptions.

We processed the corpus with natural language processing tools during the pre-processing phase, i.e., tokenizer, lemmatizer, and POS tagger. The output (Figure 1) of the preprocessing phase (matrix of stylometric features) was submitted to the data mining platform Rapidminer [33]. The text preprocessing pipeline was initially applied to the original texts of the students. However, we observed that various language errors were scattered across all linguistic levels and inserted significant bias in the modeling process negatively affecting the prediction results. Therefore, the essays were corrected manually without loss of information on the morphosyntactic level.

	AverageWord Length	PercentageOf TopMostFreq TriGramsCove rageInFile	VerbsFreq	PercentageOf AllStopWords CoverageInFile	PercentageOf TokensAppear ingOnceCover ageInFile	Functional Density	BigFive Classification
2	6.64	11.622	15.405	55.405	33.514	0.805	Agreeable
3	6.377	11.854	14.679	54.407	29.483	0.838	Agreeable
4	6.514	11.166	14.392	53.102	29.777	0.883	Agreeable
5	6.944	12.5	12.921	47.443	30.966	1.108	Agreeable
6	6.624	10.986	13.803	53.239	34.366	0.878	Agreeable
7	6.84	10.563	13.732	51.408	38.028	0.945	Agreeable
8	6.755	13.937	16.115	53.592	19.971	0.866	Agreeable
9	6.884	13.191	16.578	55.615	27.629	0.798	Agreeable
10	6.5	12.698	15.584	56.277	22.655	0.777	Agreeable
11	6.39	13.0	17.25	60.75	28.5	0.646	Agreeable
12	6.596	12.745	17.445	59.314	25.245	0.686	Agreeable
13	6.575	12.148	17.354	59.219	20.824	0.689	Agreeable
14	6.907	10.516	14.484	50.595	35.119	0.976	Agreeable
15	6.935	12.607	17.597	54.915	30.983	0.821	Agreeable
16	6.962	12.716	15.87	51.724	34.267	0.933	Agreeable
17	6.574	14.935	10.82	51.948	23.052	0.925	Agreeable
18	6.768	12.275	15.569	53.593	24.551	0.866	Agreeable

Figure 1. Output of the preprocessing phase.

We designed and ran multiple experiments in order to extract and quantify many different subsets of stylometric features from the corpus. We extracted the most frequent character bigrams and trigrams, words bigrams, and trigrams, mean word and sentence length, the occurrence frequency of content and functional words, the most and less frequent words, the occurrence frequency of parts of speech, as well as hapax and dis legomena. These features have been proven effective in the field of authorship attribution [34] and gender identification [35], and we tested them for author personality prediction as well. A list of the stylometric features extracted from the textual data is reported in Table I.

C. Classification Algorithms

In this project, the problem of predicting the personality type and personality traits was treated as a binary classification task among the four dimensions of personality, **E**xtraversion-**I**ntroversion, **S**ensing-i**N**tuition, **T**hinking-**F**eeling, and **J**udging-**P**erceiving and on the other hand, the Five Factors of personality, **O**penness to Experience, **C**onscientiousness, **E**xtraversion, **A**greeableness, and **N**euroticism. The extracted stylometric features matched the texts whose authors clearly belonged to a positive or negative category to have a valid prediction.

TABLE I. STYLOMETRIC FEATURES EXTRACTED FROM CORPUS

1. Frequency of Verbs	13. Frequency of Active Voice Verbs	25. Functional Density
2. Frequency of Nouns	14. Frequency of Passive Voice Verbs	26. Average Word Length
3. Frequency of Adjectives	15. Percentage of all Stop Words	27. Average Sentence Length
4. Frequency of Articles	16. Percentage of Top Most Frequent Tokens	28. Percentage of Top Most Frequent Word Bigrams
5. Frequency of Pronouns	17. Percentage of Top Most Frequent Non Stop Words	29. Percentage of Top Most Frequent Word Trigrams
6. Frequency of Adverbs	18. Percentage of Bottom Least Frequent Tokens	30. Percentage of Top Most Frequent Character Bigrams
7. Frequency of Prepositions	19. Percentage of Bottom Least Frequent Non Stop Words	31. Percentage of Top Most Frequent Character Trigrams
8. Frequency of Conjunctions	20. Number of Single Non Stop Words per all Words Occurrences	32. Percentage of 100 Most Frequent Words
9. Frequency of Personal Pronouns	21. Percentage of Tokens Appearing Once	33. Percentage of 100 Most Frequent Word Bigrams
10. Frequency of Coordinative Conjunctions	22. Percentage of Tokens Appearing Twice	34. Percentage of 100 Most Frequent Character Bigrams
11. Frequency of Subordinative Conjunctions	23. Ratio of Twice over Once Appearing Tokens	35. Percentage of 100 Most Frequent Character Trigrams
12. Frequency of Personal And Possessive Pronouns	24. Percentage of all Non Stop Words	

Since personality detection presents a complex classification task, we decided to use several different machine learning algorithms to find the best approach in terms of model performance. We compared nine machine learning methods, i.e., Naive Bayes, Generalized Linear Model (Binomial Method), Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees, Support Vector Machines, and we ranked them according to their cross-validated accuracy (10-fold). We evaluated the machine learning algorithms in terms of their predictive ability using the students' essays as training data. Their personality type and traits had been defined before using the appropriate psychometric questionnaires.

V. RESULTS

This section presents the results of the procedure that we followed to automatically classify the students' essays based on the personality type and personality traits defined by the personality questionnaires they filled in. From the nine algorithms trained in the textual data, we present the evaluation metrics of the most effective algorithm (Table II and Table III) along with the corresponding weights that

positively affected the prediction of the personality type and traits depending on the psychological theory used.

A. *Jung Typology Test*

Regarding the prediction of all personality types of Jung's typology, the algorithm with the best results was Naive Bayes. The accuracy rate revealed a range from 68.8% to 80.7%, with an average of 76.5%. Extraversion type was predicted with 80.7%, the Intuition type with 79.9%, the Feeling with 68.8%, and the Judging type with 75.7% [36]. A more detailed list of evaluation metrics (accuracy, precision, and recall) is reported in Table II.

TABLE II. NAIVE BAYES MODEL PERFORMANCE

Personality Type	**Naive Bayes Classifier**		
	Accuracy	***Precision***	***Recall***
Extraversion	80.7%	80.5%	100%
Intuition	79.9%	81.3%	92.6%
Feeling	68.8%	67.7%	96.7%
Judging	75.7%	76.2%	95.2%

The remaining algorithms that were trained in the corpus produced the following results in terms of classification accuracy: Regarding the Extraversion type, the Generalized Linear Model (Binomial Method), Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest and Gradient Boosted Trees algorithms have the same percentage of accuracy being 80.0%, and the Support Vector Machine algorithm has 79.0%. The Intuition type was predicted with 75.0% by Gradient Boosted Trees algorithm, with 71.9% by Deep Learning and 71.7% by Generalized Linear Model (Binomial Method) and Logistic Regression. For the Feeling type, the Decision Tree algorithm exhibits the second-best performance with 63.2%. Random Forest is in the third position with 63.1%. The next best result was 63% using Gradient Boosted Trees. The algorithms with the best performance for the Judging type were Support Vector Machine, Fast Large Margin, and Deep Learning with calculated accuracies of 71.1%, 71.0%, and 70.3%, respectively.

The study aimed to classify the essays of the students in personality types by using stylometric indices. Therefore, we had to check whether and which of these features are the most useful and contribute to the prediction accuracy of the algorithm. For this reason, we extracted the weights from the Naive Bayes model that measure the importance of each stylometric feature to the classification decisions of the algorithm for each personality type separately.

For Extraversion (Figure 2), verb types in active voice had a significant impact. In addition, the mean length of the sentence in words of all sentences, the words that occur only twice in one text, the most frequent content words, and finally, the personal pronouns complete the list with the five most important stylometric features.

Figure 2. Weights for Extraversion.

Figure 3 depicts the prediction ability of the stylometric features for Intuition used by the algorithm. The word's mean length in characters had the most significant impact. The features that follow are the most frequent trigrams of characters, the hapax legomena, the personal pronouns, the content words, the most frequent word bigrams, the rarest words, the most frequent word trigrams, and all content words.

Figure 3. Weights for Intuition.

The stylometric features that affected the result of the classification of the essays in terms of Feeling are the verbs, the adjectives, the most frequent content words, the personal and the possessive pronouns, the nouns, and the adverbs (Figure 4).

Figure 4. Weights for Feeling.

Finally, in Figure 5, the eight stylometric features that contributed to the prediction of the Judging type were in descending order: The most common word trigrams, the most common word bigrams, the mean length of the sentence in words, the most common character bigrams and the most common character trigrams with the same percentage, the

personal and possessive pronouns, the articles, and the mean length of the word in characters.

Figure 5. Weights for Judging.

B. *Big-Five Personality Test*

Regarding the prediction of all Big Five personality traits, the algorithm with the best results was the Generalized Linear Model (Binomial Method). The accuracy rate revealed a range from 65.6%% to 85.9%, with an average of 72.1%. Openness to Experience was predicted with 85.9%, Conscientiousness with 71.2%, Extraversion with 67.6%, Agreeableness with 70.2%, and the trait of Neuroticism with 65.6% [36]. It clearly emerges that Openness to Experience is the easiest trait that can be predicted from the textual data, followed by Conscientiousness and Agreeableness. Table III reports a more detailed list of evaluation metrics (accuracy, precision, and recall).

TABLE III. GENERALIZED LINEAR MODEL PERFORMANCE (BINOMIAL METHOD)

Personality Trait	**Generalized Linear Model (Binomial Method)**		
	Accuracy	***Precision***	***Recall***
Openness	85.9%	85.4%	100%
Conscientiousness	71.2%	68.6%	80.0%
Extraversion	67.6%	66.7%	86.7%
Agreeableness	70.2%	67.9%	98.7%
Neuroticism	65.6%	64.8%	71.9%

In terms of classification accuracy, the next best algorithms trained in the corpus produced the following results: Regarding Openness to Experience, Logistic Regression achieved 85.2%, Fast Large Margin, Decision Tree, Random Forest and Support Vector Machine 81.4% and Deep Learning 80.4%. The algorithms with the best performance for the trait of Conscientiousness were Naive Bayes, Gradient Boosted Trees, and Random Forest with calculated accuracies of 61.8%, 57.2%, and 56.0%, respectively. Extraversion was predicted with 65.1% by Fast Large Margin algorithm, 60.0% by Decision Tree and Random Forest, and 57.5% Deep Learning and Support Vector Machine. For Agreeableness, Random Forest algorithm exhibits the second-best performance with 69.1%, Deep Learning, Decision Tree, Gradient Boosted Trees, and Support Vector Machine being in the third position with 62.1%. The algorithms with the best performance for the trait of Neuroticism were Deep Learning, Naive Bayes, and Logistic Regression with calculated accuracies of 59.2%, 58.0%, and 57.6%, respectively.

In the following paragraphs, we present the weights we extracted from the Generalized Linear Model (Binomial Method) that measure the importance of each stylometric feature to the classification decisions of the algorithm for each personality trait separately with the aim to classify the essays of the students in personality traits by using stylometric indices. Therefore, we had to check whether and which of these features are the most useful and contribute to the prediction accuracy of the algorithm.

For Openness to Experience (Figure 6), the use of personal pronouns had a significant impact. In addition, the use of verbs, dis legomena, adjectives, prepositions, pronouns, articles, subordinative conjunctions, nouns, conjunctions, adverbs, and coordinative conjunctions complete the list with the twelve most important stylometric features.

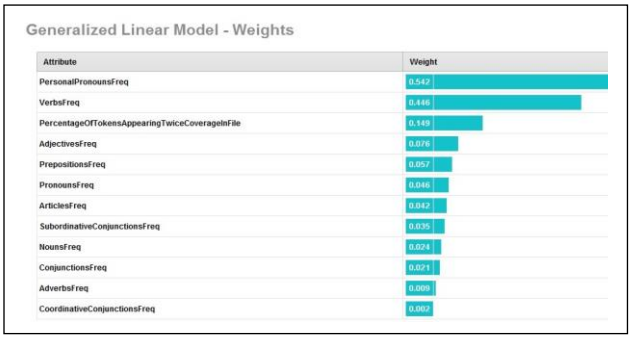


Figure 6. Weights for Openness to Experience

The stylometric features that contributed to the prediction of Conscientiousness were in descending order: functional density, non stop words, stop words, dis legomena, ratio of twice over once appearing tokens, the top most frequent tokens, the average word length, the top most frequent word bigrams, the hapax legomena, subordinative conjunctions, the bottom least frequent tokens, and the bottom least frequent non stop words (Figure 7).

Figure 8 depicts the prediction ability of the stylometric features for Extraversion used by the algorithm. The average sentence length had the most significant impact. The features that follow are ratio of twice over once appearing tokens, personal and possessive pronouns, the top most frequent word bigrams, adverbs, the bottom least frequent non stop words, conjunctions, prepositions, and dis legomena.

The stylometric features that affected the result of the classification of the essays in terms of Agreeableness are the verbs, ratio of twice over once appearing tokens, dis legomena, the use of verb types in active voice, personal and possessive pronouns, the top most frequent word bigrams, the average word length in characters, the average sentence length

in words, prepositions, the top most frequent tokens, and the top most frequent character trigrams (Figure 9).

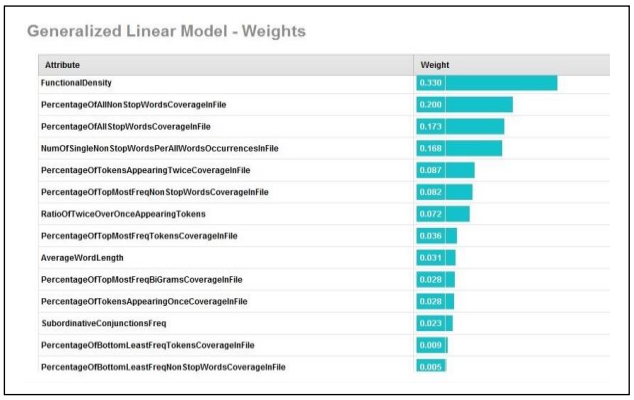


Figure 7. Weights for Conscientiousness

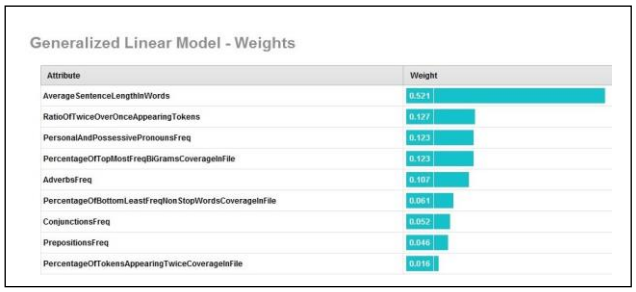


Figure 8. Weights for Extraversion

Figure 9. Weights for Agreeableness

Finally, in Figure 10, the stylometric features that contributed to the prediction of Neuroticism are many of the 100 most frequent character trigrams, which were extracted from the whole corpus in contrast to the other features extracted from the subcorpora depending on the personality trait. Additionally, personality prediction was affected by subordinative conjunctions, adverbs, nouns, and the top most frequent word trigrams.

Figure 10. Weights for Neuroticism

It is evident that the most important features extracted from the model vary considerably for each personality type and trait. Therefore, we can infer that each type and trait is based on a different combination of linguistic features and these subsets are different between the different personality types and traits.

It also becomes clear that the predictive accuracy of the proposed classification model is high compared to the existing literature on the field of personality prediction. Regarding Jung's Typology Test, we got an average accuracy of 76.5%,

compared to the 68.62% reported for Dutch [8]. On the other hand, research on textual data from essays using the Big Five model achieved an average accuracy of 60.6% [25], while we got 72.2%. The other studies mentioned [9]-[13][26] implemented machine learning techniques in textual data that were retrieved from social media. Therefore, their results can't be directly compared since they involve research with textual data from adults written under different circumstances and in a different language.

VI. Conclusion and Future Work

To summarize, in this paper, we presented the results of our research in the field of personality prediction. We applied CPP for the first time in texts written by high-school students, making our dataset unique. Our results confirmed our initial research hypothesis that stylometric features could be used as reliable prediction indices for the author's psychological profile.

It is essential, of course, to emphasize that in the research field in which this research belongs, there are no reference data measuring and comparing the performance of different personality traits prediction methods objectively. None of the existing research uses comparable methods that have been applied to identical or comparable sets of textual data in the same language. Therefore, the percentages of accuracy from literature involve research with textual data, but not those of students but those of adults, written under other circumstances and in a different language; and, of course, with other features, not always stylometric.

Our findings further support the latent link of personality traits with a wide array of linguistic behaviour aspects. Different personality types correlate with different stylometric features that belong to different linguistic levels. Therefore, the personality prediction through text demands a highly dynamic feature set to capture the widest possible spectrum of linguistic structures.

A basic target for continuing the research work in CPP is the investigation of new traits but also testing more stylometric features. In this study, we utilized only linguistic stylometric features. In this direction, future research will employ experimentation with new linguistic features or features already examined in the literature, such as content features, psycholinguistic, and syntactic features. We plan to localize well-known psychometric lexicons in Modern Greek (e.g., LIWC) and use them to complement our feature sets. In addition, we need to select features depending on the corpus, as, for instance, a person writes differently in a school essay and differently on social media.

We have ascertained the need to develop high-volume representative data since this constitutes a prerequisite for any relevant research. We need to create specialized corpora of Greek education texts by means of the used algorithmic methods, as well as the respective reference corpora to review the performance of the methods. Moreover, we plan to increase the size of the used corpus with additional students' essays. To draw more reliable conclusions, the growth of the corpus needs to ensure a balance between textual genres and different personality profiles. To achieve this, the corpus could be enriched with essays of different topics and textual genres and sufficient data for every psychological type.

References

[1] S. Gagiatsou, G. Markopoulos, and G. Mikros, "Using Stylometric Features to Predict Author Personality Type in Modern Greek Essays" The Fifteenth International Conference on Digital Society (ICDS 2021), Jul. 2021, pp. 34-39, ISBN: 978-1-61208-869-3

[2] Y. Neuman, D. Assaf, Y. Cohen, and L. J. Knoll, "Profiling school shooters: Automatic text-based analysis," Front. Psychiatry, vol. 6, p. 86, 2015.

[3] J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn, "The Development and Psychometric Properties of LIWC2015". Austin, TX: University of Texas at Austin. (www.LIWC.net), 2015. [retrieved: November, 2021]

[4] C. G. Jung, "Psychological Types," Princeton, New Jersey: Princeton University Press, 1971.

[5] I. Briggs Myers and P. B. Myers, "Gifts Differing: Understanding Personality Type," Mountain View, CA: Davies-Black Publishing, 1980.

[6] Jr. P. T. Costa and R. R. McCrae, "NEO-PI-R: Professional Manual," Odessa, Fla.: Psychological Assessment Resources, 1993.

[7] The Big Five Project Personality Test https://www.outofservice.com/bigfive. [retrieved: November, 2021].

[8] P. Juola, "Authorship Attribution," Foundations and Trends in Information Retrieval, vol. 1, no. 3, pp. 233-334, 2008.

[9] F. H. Sanford, "Speech and Personality: A Comparative Case Study," Journal of Personality, vol. 10, pp. 169-198, 1942.

[10] J. J. Bradac, "Language Attitudes and Impression Formation", In H. Giles and W.P. Robinson (eds.), Handbook of Language and Social Psychology, pp. 387-412, 1990.

[11] J. W. Pennebaker and L. A. King, "Linguistic Styles: Language Use as an Individual Difference," Journal of Personality and Social Psychology, vol. 77, pp. 1296-1312, 1999.

[12] A. J. Gill, "Personality and Language: The Projection and Perception of Personality in Computer-mediated Communication", Ph.D. Thesis, University of Edinburgh: Scotland, 2003.

[13] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological Aspects of Natural Language Use: Our Words, our Selves," Annual Review of Psychology, vol. 54, pp. 547-577, 2003.

[14] J. W. Pennebaker and L. D. Stone, "Words of Wisdom: Language Use over the Life Span," Journal of Personality and Social Psychology, vol. 85, no. 2, pp. 291-301, 2003.

[15] J. Oberlander and A. J. Gill, "Language with Character: A Stratified Corpus Comparison of Individual Differences in e-mail Communication," Discourse Processes, vol. 42, pp. 239-270, 2006.

[16] K. Luyckx and W. Daelemans, "Personae: A corpus for Author and Personality Prediction from Text" The Sixth International Language Resources and Evaluation Conference (LREC 2008), 28-30 May 2008, pp. 2981-2987.

[17] D. Brinks and H. White, "Detection of Myers - Briggs Type Indicator via Text based Computer-mediated Communication," CS 229 Machine Learning Projects, Stanford, 2012.

[18] B. Plank and D. Hovy, "Personality Traits on Twitter-or-how to get 1,500 Personality Tests in a Week" The Sixth Workshop on Computational Approaches to Subjectivity, Sentiment and

Social Media Analysis, Association for Computational Linguistics, 2015, pp. 92-98.

[19] B.Verhoeven, W. Daelemans, and B. Plank, "Creating TwiSty: Corpus Development and Statistics," Computational Linguistics and Psycholinguistics Research Center CLiPS Technical Report Series, University of Antwerp, Belgium, CTRS-006, 2016.

[20] L. C. Lukito, A. Erwin, J. Purnama, and W. Danoekoesoemo, "Social Media User Personality Classification using Computational Linguistic" The Eighth International Conference on Information Technology and Electrical Engineering, Oct. 2016, pp. 1-6.

[21] K. Yamada, R. Sasano, and K. Takeda, "Incorporating Textual Information on User Behavior for Personality Prediction" The 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Jul.-Aug. 2019, pp. 177-182.

[22] S. Argamon, M. Koppel, J. W. Pennebaker, and J.Schler, "Automatically Profiling the Author of an Anonymous Text," Communications of the Association for Computing Machinery, vol. 52, no. 2, pp. 119-123, 2009.

[23] F. Mairesse and M. A. Walker, "Words Mark the Nerds: Computational Models of Personality Recognition through Language" The 28th Annual Conference of the Cognitive Science Society, Jul. 2006, pp. 543-548.

[24] M. Wilson, "MRC Psycholinguistic Database: Machine Usable Dictionary, Version 2.00," Behavioural Research Methods, Instruments and Computers, vol. 20, pp. 6-11, 1988.

[25] Y. Mehta et al., "Bottom-up and top-down: Predicting Personality with Psycholinguistic and Language Model Features" 20th IEEE International Conference on Data Mining (ICDM), Nov. 2020, pp. 1184-1189.

[26] F. Rangel et al., "Overview of the 3rd Author Profiling Pask at PAN 2015" In L. Cappellato, N. Ferro, J.Gareth, and E. San Juan (Eds), CLEF 2015 Evaluation Labs and Workshop - Working Notes Papers, Sep. 2015.

[27] J. Oberlander and S. Nowson, "Whose Thumb is it anyway? Classifying Author Personality from Weblog Text" The 44th Annual Meeting of the Association for Computational Linguistics ACL, Jul. 2006, pp. 627-634.

[28] F. Iacobelli, A. J. Gill, S. Nowson, and J. Oberlander, "Large Scale Personality Classification of Bloggers" The Fourth International Conference on Affective Computing and Intelligent Interaction, 2011, Heidelberg: Springer-Verlag, pp. 568-577.

[29] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and Patterns of Facebook Usage" The Fourth Annual ACM Web Science Conference, 2012, pp. 36-45.

[30] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, "Lexical Predictors of Personality Type" The Joint Annual Meeting of the Interface and the Classification Society of North America, 2005, pp. 1-16.

[31] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," Journal of Artificial Intelligence Research, vol. 30, pp. 457-500, 2007.

[32] J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media" The 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems, pp. 253-262.

[33] I. Mierswa and R. Klinkenberg. Rapidminer Studio (9.1) [Data science, machine learning, predictive analytics]. (https://rapidminer.com). [retrieved: November, 2021].

[34] G. K. Mikros and G. Markopoulos, "Using Multiword Sequences as Features in Authorship Attribution: Experiments based on Greek Blog Texts," In A. Christofidou (Ed.), Aspects of Corpus Linguistics: Principles, applications and challenges Vol. 14, pp. 56-67, 2017. Athens: Academy of Athens: Research Center for Scientific Terms and Neologisms.

[35] G. K. Mikros, "Authorship Attribution and Gender Identification in Greek Blogs," In I. Obradović, E. Kelih & R. Köhler (Eds.), Selected papers of the VIIIth International Conference on Quantitative Linguistics (QUALICO), Apr. 2013, Belgrade: Academic Mind, pp. 21-32.

[36] S. Gagiatsou, "Automatic author profiling based on natural language processing techniques", Ph.D. Thesis, National and Kapodistrian University of Athens: Greece, 2021.

Transcending Two-Path Impedance Spectroscopy with Machine Learning: A Computational Study on Modeling and Quantifying Electric Bipolarity of Epithelia

Benjamin Schindler
Universität Leipzig
Machine Learning Group
Leipzig, Germany
bschindler@informatik.uni-leipzig.de

Dorothee Günzel
Charité - Universitätsmedizin Berlin
Institute of Clinical Physiology
Berlin, Germany
dorothee.guenzel@charite.de

Thomas Schmid
Lancaster University in Leipzig /
Universität Leipzig
Leipzig, Germany
schmid@informatik.uni-leipzig.de

***Abstract*—Quantifying tissue permeability is a central task in assessing pathophysiology of intestinal epithelia. A common and convenient approach for this task is to determine electric properties like resistance and capacitance of the epithelial tissue by applying impedance spectroscopy. While the measurement technique itself is well-established, analysis tools and strategies are still subject to ongoing research in epithelial physiology. Estimations of electric parameters are known to be particularly imprecise for models where apical and basolateral sides of the tissue differ significantly from each other. One-sided application of substances such as Nystatin play an important role here, as they alter membrane conductivity on one side of the tissue while leaving other properties unchanged. Here, we present a novel method that considers two functional states of the cells, namely before and after apical addition of the substance Nystatin. To this end, an extensive dataset modeled after the epithelial cell lines HT-29/B6, IPEC-J2, and MDCK I was synthesized. In a broad study, we show that considering features from two distinct tissue states leads to significantly better regressions by decision trees, random forests, and multilayer perceptrons. Therein, we extend previous work in order to progress from a two-path to a more revealing three-path model of electric tissue properties. Parameters of a corresponding equivalent circuit could be determined with less than five percent deviation from the known target value on average. In a post-processing step, predictions by independent machine learning regressions are employed to initialize a least squares parameter fitting, where the associated impedance spectrum is aligned with the originally observed spectrum, reducing the residual sum of squares by 99% on average.**

Keywords–Physiology; Epithelia; Impedance Spectroscopy; Machine Learning; Least Squares; Neural Networks; Random Forests.

I. Introduction

Epithelia play a crucial role in animal and human bodies. They are tissues that form barriers between different compartments of an organism as well as between the organism and the environment. Stratified epithelia, such as the skin, consist of multiple cell layers, whereas simple epithelia (e.g., intestinal, kidney, lung, or glandular epithelia) consist of a single cell layer that is attached to the basal lamina, a matrix of extracellular proteins. The results presented here extend previous work aiming at modeling electric properties of these simple epithelia [1] and reflects properties of animal and human colon, intestinal and kidney tissue.

Fully differentiated epithelial cells are polarized, i.e., the basolateral plasma membrane attached to the basal lamina or facing the neighboring cells functionally differs from the opposite, apical plasma membrane (facing the external environment or the lumen of body cavities). Both, apical and basolateral membranes contain a multitude of different ion channels, carriers and pumps (for review see, e.g., [2][3]) and their apical/basolateral compartmentalization is a prerequisite for vectorial transepithelial transport. Apical and basolateral membranes are separated by the tight junction (TJ). TJs are a meshwork of protein strands that consist of integral membrane proteins. These proteins bind to their counterparts within the same plasma membrane as well as within the plasma membrane of the neighboring cells and thus serve two functions:

1) *Fence function.* TJs act as a diffusion barrier for proteins within the plasma membrane and prevent membrane proteins from diffusing from the apical to the basolateral membrane compartment.
2) *Gate function.* TJs limit and regulate the passage of water and solutes between cells, i.e., along the paracellular pathway [4].

Since their discovery in the 1990s, members of the TJ-associated MARVEL proteins (TAMP), including, e.g., occludin or tricellulin [5][6] and the claudin protein family [7] have been recognized to be the main constituents of this paracellular barrier. Claudins are of special structural relevance as they are able to spontaneously assemble into a meshwork of TJ-like strands at the contact areas with neighboring, even when overexpressed in non-epithelial cells [8]. As a consequence, many members of the claudin protein family strengthen the barrier function of the TJ, as they seal off the paracellular cleft. However, some claudins (e.g., claudin-2, -10, -15, -17) were found to act as paracellular cation or anion channels within these TJ strands [9].

The barrier strength of an epithelium can be quantified by measuring the transepithelial resistance (TER). TER is the sum of the resistances of the subepithelial connective tissues (subepithelial resistance, R^{sub}) and the epithelial resistance (R^{epi}). The latter, again, consists of two major resistances: the resistance of the transcellular pathway (R^{trans}) and the resistance of the paracellular pathway (R^{para}), which are connected in parallel. Thus, alterations in TER may reflect alterations in any of these three components:

$$TER = R^{sub} + R^{epi} = R^{sub} + \frac{R^{trans} \cdot R^{para}}{R^{trans} + R^{para}} \quad (1)$$

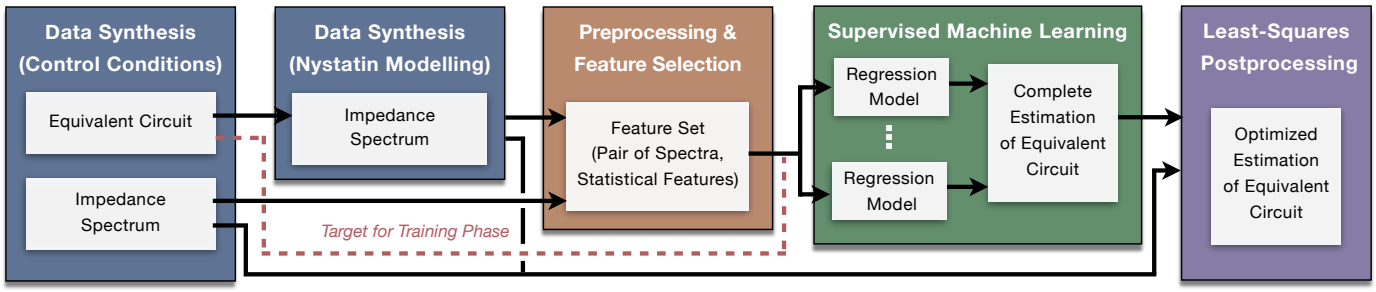


Figure 1. Overview of the presented approach for pairwise analysis of epithelial impedance spectra. The main machine learning tasks (preprocessing, feature selection, supervised learning) are preceeded by data synthesis and followed by a least squares-based postprocessing.

Activation or inactivation of ion channels in the apical or basolateral membrane alters the corresponding membrane conductivities or membrane resistances R^{ap} and R^{bl}, respectively. Any change in R^{ap} and R^{bl} also affects R^{trans} and thereby ultimately alters TER. A prominent example is the over-stimulation of cAMP-regulated Cl^- channels in the apical membrane of enterocytes, e.g., through activation of the adenylate cyclase by toxins such as cholera toxin, toxins from enterotoxigenic *Escherichia coli* (ETEC) or by substances such as caffeine or theophylline. The resulting secretion of Cl^- into the gut lumen is accompanied by osmotic water flux and hence leads to secretory diarrhea.

Alterations in the paracellular barrier of the intestine may be caused by an upregulation of channel-forming TJ proteins or a downregulation of barrier-forming TJ proteins. Both result in a back-leakage of electrolytes and consequently of water into the gut lumen and thus to leak-flux diarrhea (leaky gut syndrome). Dysregulation of TJ proteins is observed in inflammatory diseases, such as chronic inflammatory bowel diseases (Crohn's disease; ulcerative colitis), immune-related diseases (e.g., celiac disease) or in protozoal, bacterial or viral infection (for review see, e.g., [10][11]). Thus, a hall-mark of secretory as well as of leak flux diarrhea is a reduction in TER. Therefore, alterations in TER of a tissue may be due to alterations in the transcellular or the paracellular pathway (or both) and a detailed knowledge of the underlying processes is crucial for an understanding and treatment of these diseases.

Alterations in R^{sub} are observed in many inflammatory diseases that cause increased proliferation of subepithelial cells and thus a thickening of the subepithelial tissue. The resulting increase in R^{sub} is of experimental, not of physiological relevance: in the experiment, TER is increased, suggesting an increase in barrier function. In situ, however, capillaries traverse the subepithelium so that the distance between the capillaries and the basolateral membrane of the epithelium remains unaltered. Thus, the observed increase in TER does not affect the barrier function of the epithelium. Exact knowledge of R^{sub} is thus necessary to correctly assess alterations in R^{epi}.

Many different cell culture models are established to study regulatory aspects of epithelial barrier functions and to investigate the effects of over-expression or knockdown of TJ components. Among the most widely studied epithelial cell lines are human colonic Caco-2, T84, HT-29/B6; human bronchial 16HBE14o-; porcine jejunum IPEC-J2; porcine kidney LLC-PK1; canine kidney MDCK; mouse kidney IMCD-3; mouse mammary HC11. These cell lines fully differentiate into tight cell monolayers when grown on filter supports, as judged by their transepithelial resistances of several hundreds or even thousands of $\Omega \cdot cm^2$. As in tissue, TER is the sum of R^{epi} and R^{sub}. While R^{epi}, again, consists of R^{trans} and R^{para}, R^{sub} in the cell culture system is the resistance of the filter support on which the cells are grown and thus constant.

In order to describe, explain and understand epithelial functionality, a variety of computational models has been developed for this tissue. To this end, we have established a machine learning-based approach for quantifying electrical properties of epithelia. In particular, we could not only improve estimates for R^{epi} [1], but also the one for the epithelial capacity C^{epi} [12]. Based on an early proof-of-concept study [13], we here introduce an improved, systematic and reliable approach to quantify transcellular properties discerned into apical and basolateral resistances (R^{ap}, R^{bl}) and capacitances (C^{ap}, C^{bl}). The key concept to achieve this is to model machine learning training data, not for individual impedance measurements but a combination of two measurements during distinct tissues states. Moreover, we introduce an additional post-processing step to further improve parameter quantifications. Figure 1 shows an overview of the approach.

The rest of the paper is organized as follows. First, the current state of research is presented (Section II). In the following sections, a new method for the determination of physiological properties of epithelia is introduced. Figure 1 shows this method as a block diagram. The basis is a data synthesis where impedance spectra are synthesized using an equivalent circuit as a cell model (Section III). From previous estimates of physiological properties, epithelial tissues are modeled under control conditions as well as after the addition of Nystatin. From these, two impedance spectra are generated for each sample. In a preprocessing step, new statistical features are constructed from this pair of spectra. Supervised learning is then used to train the prediction of circuit parameters (Section IV). In this step, only a single parameter of the cell model is determined at a time. By combining the individual parameter values, a complete prediction of the model parameters is obtained. Using the originally observed spectra from the data set, the circuit parameters are further optimized in a least squares post-processing (Section V). In a comprehensive study, this method is applied to the three cell lines HT-29/B6, IPEC-J2 and MDCK I. Subsequently, the results of the evaluation are discussed (Section VI) and summarized (Section VII).

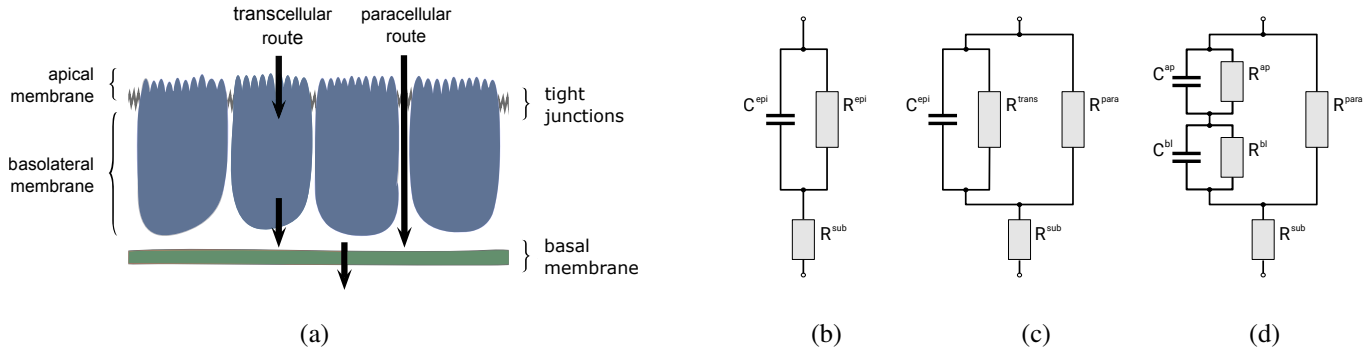


Figure 2. a) Schematic drawing of a simple epithelium. b) A resistor-capacitor (RC) circuit reflects general properties of the epithelium. c) A RC circuit with a parallel resistor reflects the paracellular pathway. d) A 2-RC circuit reflects apical and basolateral membrane properties.

II. State of the Art

Over the last century, a variety of techniques has been established to assess the functionality of epithelia. The most common way to determine tissue permeability is a direct measurement of fluxes, e.g., by using radioactive isotopes or labeled substances. Alternatively, the permeabilities of the two major extracellular ion species, Na^+ and Cl^-, are determined by measuring tissue conductance or its reciprocal, resistance, e.g., by using "chopstick electrodes" or Ussing chambers [14]. Both techniques, however, simply measure TER and are not able to discriminate between any partial resistances (cf. Eq. (1)).

In contrast to DC resistance measurements, impedance spectroscopy measurements allow us to distinguish between R^{sub} and R^{epi}. Further, impedances reflect not only conductive but also capacitive properties and allow to derive the epithelial capacitance, which directly depends on the epithelial surface area [12]. Impedance spectroscopy uses the fact that the plasma membrane of epithelial cells acts as a capacitor (C^{epi}) that causes a phase shift between current and voltage under alternate current (AC) conditions. C^{epi} is short-circuited by R^{epi}, resulting in a time constant $\tau = R^{epi} \cdot C^{epi}$. With R^{epi} of different cell types and tissues varying between about 10 and 10.000 Ωcm^2 and C^{epi} values of about 1 to 10 $\mu F/cm^2$, τ ranges between 10^{-5} and 10^{-1} s. In terms of angular frequency ($\omega = 1/\tau$) this is equal to a range from 10 Hz to 100 kHz. During a typical experiment, up to 50 different frequencies covering this range are used to obtain impedance values $|Z|$ and corresponding phase angles ϕ [15].

Impedance spectroscopy was already used during the first half of the 20th century [16][17], and gained popularity in epithelial research since the 1970s (see e.g., [18][19][20][21]). To analyze samples, an equivalent electric circuit of appropriate complexity is modeled [22]. The simplest circuit that incorporates R^{epi} is a resistor-capacitor (RC) circuit (Figure 2b). In a previous publication, the evaluation of impedance spectra based on this equivalent circuit was dubbed one-path impedance spectroscopy (1PI) to distinguish it from the more recently developed two-path impedance spectroscopy (2PI, [23][13]) that additionally allows the separation of the two major transepithelial transport pathways (trans- and paracellular). In 2PI, a circuit with an RC subcircuit and a resistor in parallel is used to represent the physiological polarity of epithelial cells (Figure 2c). In both cases, the subepithelium can be represented by an additional resistor in series.

In theory, R^{sub} and R^{epi} can be derived by physical relationships. In particular, $\lim_{\omega\to\infty} \Re(Z) = R^{sub}$ and $\lim_{\omega\to 0} \Re(Z) = R^{sub} + R^{epi}$ can be exploited for this task. In previous work, we have demonstrated that traditional ways to estimate R^{epi} from these relations, like circle fits or visual extrapolation from Nyquist diagrams [22], can lead to serious errors in the analysis of epithelial characteristics [13]. This holds true, in particular, where spectra deviate from a semicircular shape. We have also demonstrated that estimations can be improved by applying machine learning techniques on complex-valued impedances of error-prone [13] or on extracted features of ideal impedance spectra, respectively [24]. For R^{epi}, e.g., we achieved less than ten percent deviation from the known target value on average with outliers, however, exhibiting significantly larger maximum deviations [24]. Recently, we have tested a novel approach to quantify electric properties of epithelia based on extracted features of error-prone impedance spectra [25].

C^{epi} is established by the transepithelial pathway or the hydrophobic part of the lipid bilayers of its cell membranes, respectively. Under the AC conditions induced by impedance measurements, charging and discharging of the epithelial capacitor depends on frequency. At very high frequencies the impedance of the capacitor approaches zero [26]. In previous work, we have shown that this physical relation can be used to approximate C^{epi} from the cartesian representation of an impedance spectrum [12]. The main advantage of this is the minimal requirements for additional equipment. Alternative methods employ square pulses and the analysis of resulting output transients [17][27] or very brief pulses or simultaneously imposing different sine waves and using Fourier analysis [28][29]. Currently, however, estimations of C^{epi} by machine learning are less precise than those by more complex and more time-consuming methods [12].

The term machine learning generally comprises a variety of algorithms that are characterized by learning a function from given samples. Depending on the employment or non-employment of target values for each sample, these algorithms are either regarded as supervised or unsupervised learning algorithms [30]. To this end, quantification of epithelial properties is a supervised learning task. Among the most widely used supervised machine learning techniques are neural net-

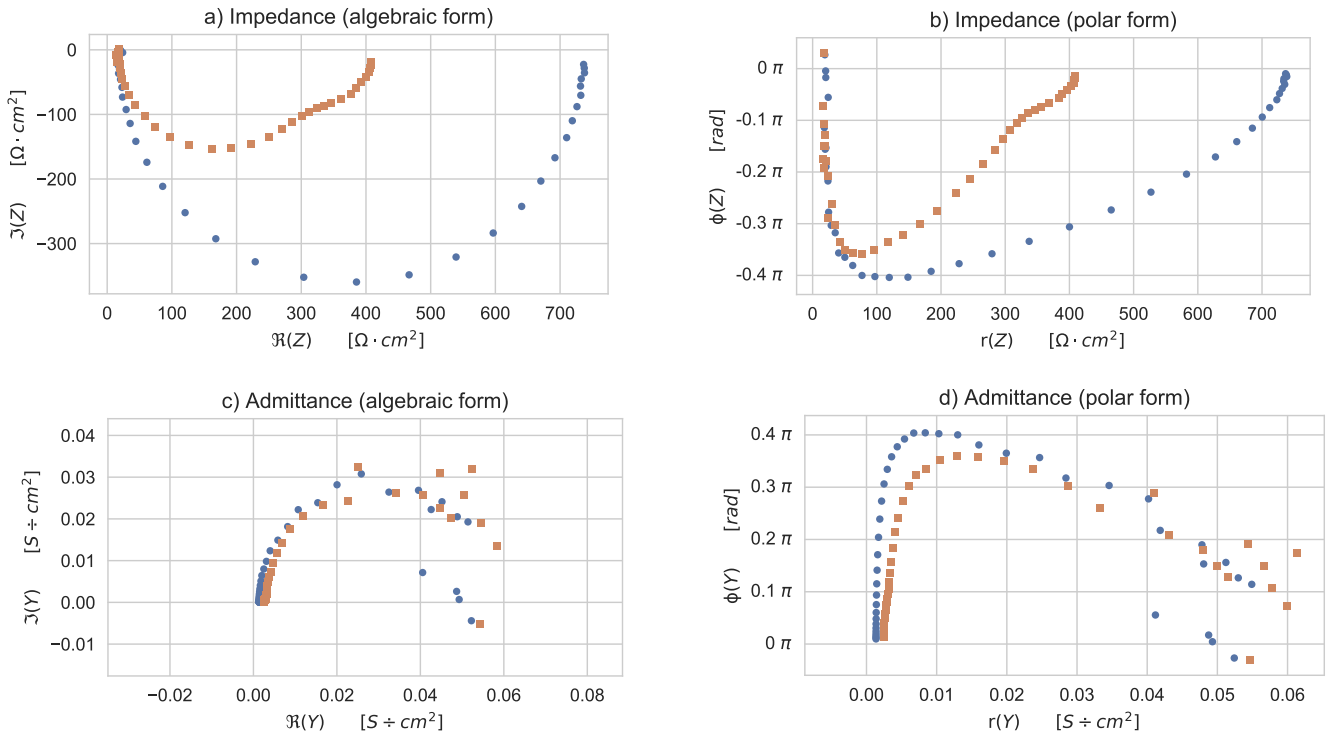


Figure 3. Overlay of two impedance measurements of modeled HT29B6 at 42 frequencies using different representations. Colors indicate functional condition before (●) and after (■) Nystatin addition. a) Nyquist representation: impedance in algebraic form ($\Re(Z)$, $\Im(Z)$) b) impedance in polar form ($r(Z)$, $\phi(Y)$) c) admittance in algebraic form ($\Re(Y)$, $\Im(Y)$) (Nyquist representation) d) admittance in polar form ($r(Y)$, $\phi(Y)$).

works, which are biologically motivated, and random forests, which are statistically motivated. Neural networks model the characteristics of biological neurons as mathematical functions [31] and have been proven to be universal approximators [32]. Random Forests are based on the concept of decision trees [33] and exploit the mathematical effects of induced variability among a large number of decision trees [34].

As a routine preprocessing task, many machine learning applications employ an extraction and/or selection of input features. While feature extraction typically aims at transforming or increasing existing features [35], feature selection can be considered as an algorithmic dimension reduction process. Feature selection techniques are typically either purely statistically motivated filter approaches or so-called wrapper approaches that make use of supervised learning algorithms [36]. Well-known wrapper approaches are, e.g., Sequential Forward Selection or Recursive Feature Elimination [37]. Apart from these general strategies, embedded feature selection approaches are known as effective strategies as they are implemented directly into supervised learning algorithms [36]. Random Forests are prominent example for embedded feature selection [34].

To achieve reliable predictions, machine learning requires large amounts of training data. As measurements are typically only available in small numbers ($n \leq 1.000$), employing this approach to epithelial analysis requires modeling realistic impedance data. Therefore, we have established a modeling methodology, which is based on an equivalent electric circuit reflecting epithelial polarity (Figure 2d). This includes not only parameter ranges for circuit components and a mathematical model of setup-specific data scatter, but also an algorithmic approach to compare modeled and measured data [13][25]. Detailed and realistic models have been developed for three epithelial cell lines under various conditions [1]: the human colon carcinoma cell line HT-29/B6, the porcine jejunum cell line IPEC-J2, and the canine kidney cell line MDCK I.

Figure 3a shows an example of the impedance spectrum of a modeled HT29B6 cell culture under control conditions and after apical addition of Nystatin. Note, how the shapes of the two curves differ: Whereas under control conditions an approximately semicircular curve is seen, the shape after addition of Nystatin can be described as two overlapping semicircles representing the two cell membrane sides. This is due to the reduction of the apical membrane resistance R^{ap} which leads to a difference in the time constants of the apical membrane τ^{ap} and basolateral membrane τ^{bl}. As an indicator of the curve shape, previous work has predicted the τ ratio q, where impedance spectra can be considered semicircular for $q < 5$ and non-semicircular for $q > 5$ [38].

In addition to the use of Cartesian coordinates, i.e., the algebraic form, other forms of representation can be considered for the analysis of impedance measurements. Therefore, polar coordinates of impedance have been calculated in previous studies to predict epithelial properties using machine learning (e.g., in [39][12][1]). In addition to impedance, its complex inverse, the admittance, can also be considered, which can also be represented in coordinates and polar form. In the case of admittance, in particular, there is a lack of more detailed research on how well it can be used for machine-learning-based analyses. Figure 3 shows the four representations and the influence of an apical addition of Nystatin.

III. Modeling Impedance Spectra

For the present study, an extensive dataset was synthesized that mimics impedance measurements on the three cell lines HT-29/B6, IPEC-J2, and MDCK I. The data is freely available online at [40]. For each sample, two different functional states are considered together: control conditions and application of Nystatin. By the modeling procedures described in the following, we produced 150,000 samples per cell line and tissue condition (cf. Table I). As described in the following, the modeling process comprised assumptions about an underlying equivalent electric circuit for both functional states as well as calculating impedances for given frequencies and modeling equipment-specific measurements biases.

TABLE I. SAMPLE SIZES FOR MODELED EPITHELIAL CELL LINES AND THEIR FUNCTIONAL CONDITIONS.

Condition	HT-29/B6	IPEC-J2	MDCK I
Control	150.000	150.000	150.000
Nystatin	150.000	150.000	150.000

A. Modeling Impedance Spectra

For all three modeled cell lines, an equivalent circuit consisting of two RC subcircuits a (R_a, C_a) and b (R_b, C_b) located in series, a resistor in parallel (R_p) and a concluding resistor in series (R_s) is assumed.(cf. Figure 2). In accordance with Kirchhoff's laws, the corresponding impedance Z at an angular frequency ω can be derived from the complex-valued impedance of the circuit components:

$$Z(\omega) = \frac{R_p(R_a + R_b) + i\omega[R_p(R_a\tau_b + R_b\tau_a)]}{R_a + R_b + R_p(1 - \omega^2\tau_a\tau_b) + i\omega[R_p(\tau_a + \tau_b) + R_a\tau_b + R_b\tau_a]} + R_s \quad (2)$$

where $i = \sqrt{-1}$, and $\tau_a = R_aC_a$ and $\tau_b = R_bC_b$.

In the measurements to be modeled, ten frequencies per decade are used. Based on a lowest frequency f_1 of 1.3 Hz, higher frequencies f_i with $1 < i \leq n = 42$ are multiples by a factor of $10^{0.1}$ (~1.26). Note, that the value of f_1 is a chosen to a avoid obtaining multiples of 50 Hz (mains frequency) and that for application with Eq. (2), $\omega_i = 2\pi/f_i$ is calculated.

From calculations, n=42 complex-valued impedances $(Z_0, \ldots, Z_{n-1})$ are obtained. Real and imaginary parts of a spectrum can be regarded as separate feature sets $S_{\Re(Z)}$ and $S_{\Im(Z)}$:

$$S_{\Re(Z)} = \{\Re(Z_0), \ldots, \Re(Z_{n-1})\} \quad (3)$$
$$S_{\Im(Z)} = \{\Im(Z_0), \ldots, \Im(Z_{n-1})\} \quad (4)$$

As an alternative representation, these complex values were transformed to polar coordinates, i.e., to phase angle ϕ and magnitude r. This results in two alternative sets S_ϕ and S_r:

$$S_{\phi(Z)} = \{\phi(Z_0), ..., \phi(Z_{n-1})\} \quad (5)$$
$$S_{r(Z)} = \{r(Z_0), ..., r(Z_{n-1})\} \quad (6)$$

Furthermore, the admittances $Y_i = Z_i^{-1}$, i.e., the complex reciprocals of the impedances, were computed for the whole spectrum. Feature sets $S_{\Re(Y)}, S_{\Im(Y)}$ correspond to their real and imaginary parts.

$$S_{\Re(Y)} = \{\Re(Y_0), \ldots, \Re(Y_{n-1})\} \quad (7)$$
$$S_{\Im(Y)} = \{\Im(Y_0), \ldots, \Im(Y_{n-1})\} \quad (8)$$

As for the impedance, the polar form of the admittance was derived from the algebraic coordinates which leads to feature sets $S_{\phi(Y)}, S_{r(Y)}$

$$S_{\phi(Y)} = \{\phi(Y_0), ..., \phi(Y_{n-1})\} \quad (9)$$
$$S_{r(Y)} = \{r(Y_0), ..., r(Y_{n-1})\} \quad (10)$$

Thus, we obtain four different representations of a spectral impedance measurement, which can be seen in Figure 3.

B. Modeling Data Scatter

Further, we model and apply synthetic data scatter reflecting deviations from the theoretical impedance value caused by the electrophysiological measurement setup. This scatter is modeled based on relative deviations of the real part $\Re$ and imaginary part $\Im$ of measured impedances from theoretical values. For a given impedance Z at frequency $f = \omega/2\pi$, the relative deviation σ of $\Re(Z)$ is approximated as second-order Fourier series (n=2) and relative deviation of $\Im(Z)$ is approximated as fourth-order polynomial function (n=4):

$$\sigma_{\Re}(f) = a_0 + \sum_{i=1}^{n} a_i \cdot cos(nwf) + b_i \cdot sin(nwf) \quad (11)$$

$$\sigma_{\Im}(f) = a_0 + \sum_{i=1}^{n} a_i \cdot f^i \quad (12)$$

where coefficients w, a_0, a_1, b_1, a_2, b_2 or a_0, a_1, a_2, a_3 and a_4 were determined by function fitting. a_0 is modeled as dependent on R^T. For further details on the data scatter model, see [13] or [41].

For all synthetic samples used in the following, complex impedances are calculated according to the model parameters. Subsequently, data scatter is added, and finally, polar impedances and admittances are calculated from the scattered complex impedances.

Note, that for the imaginary parts of the impedance, the modeled scatter can lead to positive values at measurement points that were already close to zero (cf. Figure 3a). This in turn leads to negative imaginary parts of the admittance (cf. Figure 3c). This modeling agrees with our observations from laboratory practice, although in theory such values should not occur (cf. Eq. (2)).

C. Sampling Cell Lines for Control Conditions

While IPEC-J2 and MDCK I cells typically show relatively high R^{epi} values under physiological conditions, for HT-29/B6 cells relatively low values are measured. Based on this, as well as on further published measurement results, parameter ranges of the components of the equivalent circuit had been defined in previous work [25]. An overview of the parameter ranges for the synthesis is shown in Table II. In an early synthesis phase, samples with biologically implausible values were obtained. Under the control conditions it is therefore presumed that $R^{ap} > R^{bl}$ applies to all cell lines and $\frac{1}{4} < \frac{R^{para}}{R^{trans}} < 4$ applies to HT-29/B6 and MDCK I.

From the observation that impedance spectra from epithelia usually show a symmetric semi-circular shape under control conditions, it can be concluded that apical and basolateral polarity outbalance each other [22]. In other words, the τ ratio q is expected to be close to 1. Therefore, impedance spectra, whose equivalent circuits have a τ-ratio greater than five, were

TABLE II. MODELED PARAMETER RANGES FOR CELL LINES UNDER CONTROL CONDITIONS

Parameter	HT-29/B6		IPEC-J2		MDCK I		unit
	min	*max*	*min*	*max*	*min*	*max*	
R^{sub}	8	30	8	50	8	25	$\Omega \cdot cm^2$
R^{epi}	150	1498	900	8567	100	4495	$\Omega \cdot cm^2$
R^{para}	152	30000	944	15000	102	10000	$\Omega \cdot cm^2$
R^{trans}	151	20000	3000	20000	101	15000	$\Omega \cdot cm^2$
R^{ap}	1	19500	2500	19000	10	14500	$\Omega \cdot cm^2$
C^{epi}	1.0	5.5	0.7	2.0	0.8	3.3	$\mu F/cm^2$
C^{ap}	1.3	8.1	1.1	3.3	0.9	5.7	$\mu F/cm^2$
C^{bl}	5.0	86.2	1.5	9.3	2.9	9.5	$\mu F/cm^2$

filtered out. Visual inspection confirmed that this yields only impedance spectra with an almost semicircular shape.

The obtained dataset shows different distributions of the equivalent circuit parameters for the three cell lines, as can be seen in Figure 4. For example, it is clear that for HT-29/B6 modeling, as in reality, significantly smaller values for R^{epi} occur than for MDCK I and IPEC-J2 (cf. Figure 4a). Note, that for HT-29/B6 higher capacitances occur, especially on the basolateral side (cf. Figure 4d). MDCK I have on average slightly smaller resistances than IPEC-J2 (cf. Figure 4a-c), but slightly larger capacitances (cf. Figure 4d).

D. Modeling Nystatin

For estimating apical and basolateral cell membrane properties, a one-sided addition of the substance Nystatin has been used, as it increases the conductance on one side of the tissue without noticeably changing the conductance on the other side [42]. For the equivalent circuit, the apical application of Nystatin means a decrease of the apical resistance R^{ap}. To model the application of the drug to our cells, the Nystatin factor $d_{nyst} \in [0; 1]$ is introduced. The apical resistance after Nystatin application R_2^{ap} is calculated according to the following formula:

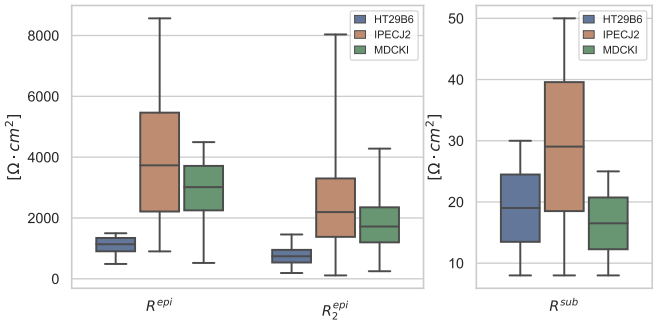
$$R_2^{ap} = R^{ap} \cdot d_{nyst} \tag{13}$$

Note, that the stronger the effect of Nystatin, the lower d_{nyst}. In reality, the effect of Nystatin depends on the amount of substance added and the time between application and measurement. The reduction of the apical resistance also results in a reduced transcellular resistance R_2^{trans}, epithelial resistance R_2^{epi} and apical time constant τ_2^{ap}.

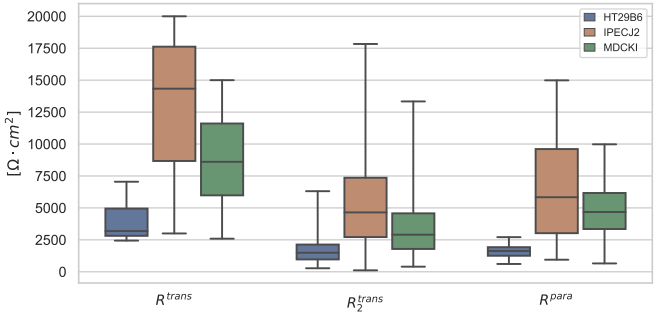
In laboratory practice, Nystatin is applied until two distinguishable semicircles are visible in the Nyquist representation. In this study this procedure is modeled as a multi-step Nystatin application. For a series of k Nystatin applications, the Nystatin factor is calculated from the product of the individual applications.

$$d_{nyst} = \prod_{i=1}^{k} d_{nyst}^{i} \tag{14}$$

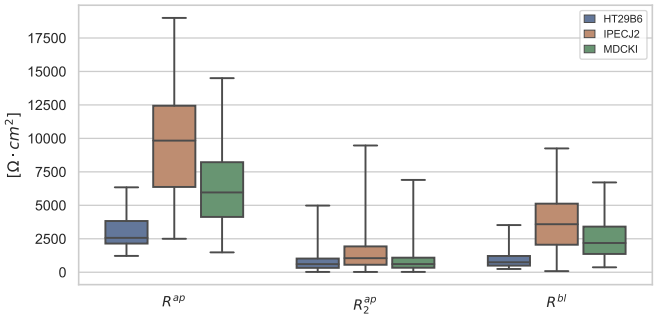
Note, that $d_{nyst}^{i} = 1.0$ means no change and a value of 0 means complete removal of apical resistance. Therefore, at each step, a random, uniformly distributed value is chosen from the interval $[0.2; 0.8]$ before updating the apical resistance R_2^{ap}. This is repeated until a τ ratio above five is reached, ensuring two distinguishable semicircles in the Nyquist representation.



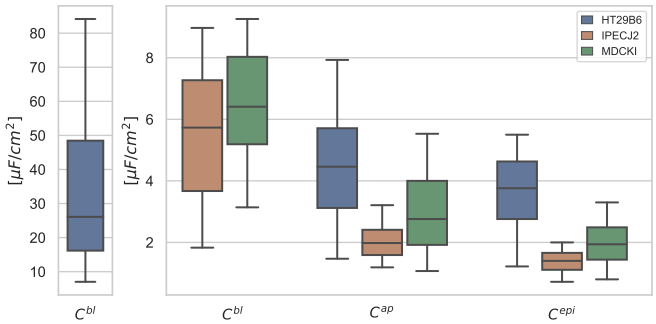
a) Epithelial resistance before (R^{epi}) and after Nystatin application (R_2^{epi}) and subepithelial resistance (R^{sub}).



b) Transcellular resistance before (R^{trans}) and after Nystatin application (R_2^{trans}) and paracellular resistance (R^{para}).



c) Apical resistance before (R^{ap}) and after Nystatin application (R_2^{ap}) and basolateral resistance (R^{bl}).



d) Apical capacity (C^{ap}), basolateral capacity (C^{bl}) and epithelial capacity (C^{epi}).

Figure 4. Overview of the distribution of the target parameters for the modeled cell lines as box plots with whiskers showing the minimum and maximum values.

TABLE III. INPUT FEATURE SETS COMPRISING IMPEDANCE SPECTRA AND STATISTICAL FEATURES.

Feature Set	Definition	Approach	Size	Quality	Form	Functional States	Ratios and Differences*
$I^1_{\perp(Z)}$	$S^*_{\Re(Z)} \times S^*_{\Im(Z)}$	1	106	Impedance	algebraic	Control	–
$I^1_{\angle(Z)}$	$S^*_{\phi(Z)} \times S^*_{r(Z)}$	1	106	Impedance	polar	Control	–
$I^1_{\perp(Y)}$	$S^*_{\Re(Y)} \times S^*_{\Im(Y)}$	1	106	Admittance	algebraic	Control	–
$I^1_{\angle(Y)}$	$S^*_{\phi(Y)} \times S^*_{r(Y)}$	1	106	Admittance	polar	Control	–
$I^2_{\perp(Z)}$	$I^1_{\perp(Z)} \times N^*_{\Re(Z)} \times N^*_{\Im(Z)}$	2	212	Impedance	algebraic	Control + Nystatin Affected	–
$I^2_{\angle(Z)}$	$I^1_{\angle(Z)} \times N^*_{\phi(Z)} \times N^*_{r(Z)}$	2	212	Impedance	polar	Control + Nystatin Affected	–
$I^2_{\perp(Y)}$	$I^1_{\perp(Y)} \times N^*_{\Re(Y)} \times N^*_{\Im(Y)}$	2	212	Admittance	algebraic	Control + Nystatin Affected	–
$I^2_{\angle(Y)}$	$I^1_{\angle(Y)} \times N^*_{\phi(Y)} \times N^*_{r(Y)}$	2	212	Admittance	polar	Control + Nystatin Affected	–
$I^3_{\perp(Z)}$	$I^2_{\perp(Z)} \times D^*_{\Re(Z)} \times D^*_{\Im(Z)} \times R^*_{\Re(Z)} \times R^*_{\Im(Z)}$	3	318	Impedance	algebraic	Control + Nystatin Affected	included
$I^3_{\angle(Z)}$	$I^2_{\angle(Z)} \times D^*_{\phi(Z)} \times D^*_{r(Z)} \times R^*_{\phi(Z)} \times R'^*_{r(Z)}$	3	318	Impedance	polar	Control + Nystatin Affected	included
$I^3_{\perp(Y)}$	$I^2_{\perp(Y)} \times D^*_{\Re(Y)} \times D^*_{\Im(Y)} \times R^*_{\Re(Y)} \times R'^*_{\Im(Y)}$	3	318	Admittance	algebraic	Control + Nystatin Affected	included
$I^3_{\angle(Y)}$	$I^2_{\angle(Y)} \times D^*_{\phi(Y)} \times D^*_{r(Y)} \times R^*_{\phi(Y)} \times R'^*_{r(Y)}$	3	318	Admittance	polar	Control + Nystatin Affected	included

*as calculated in section III-E

Figure 4 shows the distributions of equivalent circuit parameters for all cell lines after modeling Nystatin in this manner.

After updating the circuit parameters, new impedance spectra were calculated based on the modified equivalent circuit. Thus, for each sample, we obtain two impedance spectra at two different functional states. As for the state under control conditions, the feature sets for the state after Nystatin application were computed according to the formulas 3-10, but are denoted as N_κ in the following.

E. Extracted Features

New feature sets can be constructed by considering the impedance spectra before and after Nystatin application. In order to describe the change of spectra due to the application of Nystatin, differences and ratios of the measuring points were calculated for all representations, resulting in the feature sets D_κ and R_κ.

$$D_\kappa = \{n - s \mid n \in N_\kappa \wedge s \in S_\kappa \wedge \omega(n) = \omega(s)\} \quad (15)$$

$$R_\kappa = \{n/s \mid n \in N_\kappa \wedge s \in S_\kappa \wedge \omega(n) = \omega(s)\} \quad (16)$$

where $\kappa \in \{\Re(Z), \Im(Z), \phi(Z), r(Z), \Re(Y), \Im(Y), \phi(Y), r(Y)\}$ and $\omega(x)$ is the angular frequency for the measurement of x.

Previous work has shown that the use of additional statistical features can lead to an improved determination of epithelial properties [1]. According to this principle, 11 statistical properties were selected (cf. Table IV). The function stat : $\mathbb{R}^{42} \rightarrow \mathbb{R}^{11}$ returns these statistical features for a given feature set. For all previously defined feature sets S_κ, N_κ, D_κ, and R_κ, the inclusion of statistical features results in new feature sets S^*_κ, N^*_κ, D^*_κ, and R^*_κ.

$$S^*_\kappa = S_\kappa \times \text{stat}(S_\kappa) \quad (17)$$

$$N^*_\kappa = N_\kappa \times \text{stat}(N_\kappa) \quad (18)$$

$$D^*_\kappa = D_\kappa \times \text{stat}(D_\kappa) \quad (19)$$

$$R^*_\kappa = R_\kappa \times \text{stat}(R_\kappa) \quad (20)$$

where $\kappa \in \{\Re(Z), \Im(Z), \phi(Z), r(Z), \Re(Y), \Im(Y), \phi(Y), r(Y)\}$

TABLE IV. STATISTICAL PROPERTIES FOR FEATURE SET S_i.

Feature	Definition with $x_j \in S_i$ and $n = \#(S_i)$	Description
$min(S_i)$	$\{x : x \leq x_j \ \ \forall\ x_j \in S_i\}$	Minimum
$P_{10}(S_i)$	$x_{\lfloor 0.1 \cdot n+1 \rfloor}$	10th Percentile
$P_{25}(S_i)$	$x_{\lfloor 0.25 \cdot n+1 \rfloor}$	1st Quartile
$med(S_i)$	$\begin{cases} x_{(\frac{n+1}{2})}, & \text{if } n \text{ odd} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{if } n \text{ even} \end{cases}$	2nd Quartile (Median)
$P_{75}(S_i)$	$x_{\lfloor 0.75 \cdot n+1 \rfloor}$	3th Quartile
$P_{90}(S_i)$	$x_{\lfloor 0.9 \cdot n+1 \rfloor}$	90th Percentile
$max(S_i)$	$\{x : x \geq x_j \ \ \forall\ x_j \in S_i\}$	Maximum
$\bar{x}(S_i)$	$\frac{1}{n}\sum_{j=1}^{n} x_j$	Arithmetic mean
$s(S_i)$	$\sqrt{s^2(S_i)}$	Standard deviation
$s^2(S_i)$	$\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x}(S_i))^2$	Variance
$R_{MM}(S_i)$	$med(S_i) - \bar{x}(S_i)$	Distance between median and arithmetic mean

F. Comparison with Measured Data

As outlined (cf. Table I), modeling the cell lines HT-29/B6, IPEC-J2 and MDCK I under two distinct conditions, six modeling scenarios were considered. For all six scenarios, modeled impedance spectra were compared to measured spectra. As true values can not be known for measured parameters, an indirect approach needs to be chosen that involves differences between estimations. Such an indirect approach has been proposed [13] and optimized [1] by us in previous work: Each parameter is estimated by three alternative techniques, normalized and resulting estimation differences were plotted in a two-dimensional diagram for modeled and measured data, respectively. If data points from measured spectra lie within the area covered by the data obtained from model spectra, it is assumed that the modeling reflects realistic epithelial properties. By calculating hexagonal bins for both diagrams, the overlap or similarity was quantified. For further details on the calculation and quantification of similarity, see [41].

IV. Quantifying Electric Properties by Machine Learning

This section describes how physiological properties of epithelia can be determined by applying machine learning techniques on impedance measurements. In particular, we investigate how much the predictions of the equivalent circuit parameters can be improved by a two-state consideration. First, necessary preprocessing and feature selection steps are described. In the main step, parameters of the tissue-equivalent electric circuit model are determined by separate regression models (cf. Figure 1). Therein, the known parameter values from the modeled impedance data set are used for supervised learning and subsequent evaluation.

A. Preprocessing and Feature Selection

As described in the previous section, varying input feature sets were formed from the cell measurements under control conditions and their statistical properties. In particular, different qualities and forms of measurements were considered and combined. In general, this approach is analogous to our previous work [1].

In addition to features from control conditions data, the pair of spectra, i.e., before and after Nystatin application, and statistical properties were used to extract features. The last approach results from combining the features from the second approach and the differences and ratios of the measurement points under the two states as described before. For each of the three approaches, the four representations of impedance measurements were used.

All in all, this results in 12 input feature sets, which can be seen in Table III. To determine which features are suitable for estimating the physiological properties of epithelia, the feature sets were evaluated by assessing the results of various supervised learning procedures.

To evaluate the feature sets and learning methods, the absolute percentage error *APE* is calculated for all test samples from the target value t and its prediction $\hat{t}$:

$$APE(t, \hat{t}) = \frac{|\hat{t} - t|}{t} \tag{21}$$

Samples from each cell line were randomized and standardized by removing the mean and scaling to unit variance. Subsequently, the dataset was divided to obtain a training set of 100,000 and a test set of 50,000 samples.

B. Supervised Learning

While the general approach aims to determine several circuit parameters, individual circuit parameters were determined instead of multiple regression. With twelve feature sets, six targets, and three machine learning models, the total number of trained models is 216 per cell line. For the whole study, 648 models were trained.

This approach allowed to choose the best model and input feature set for each parameter. Thus, the values of R^{sub}, R^{para}, R^{ap}, R^{bl}, C^{ap}, C^{bl} under control conditions were determined. To quantify the effect of Nystatin addition, d_{nyst} was also determined, which was then used to calculate R_2^{ap} using the prediction of R^{ap}.

Previous work has shown that machine learning methods, such as random forests and multilayer perceptrons, provide more accurate estimates than traditional methods such as circle fit [12][1]. On this basis, three machine learning regression methods of varying complexity were used for this work:

- **Decision Tree (DT).** This simple regressor serves as a baseline regression method. The depth of the tree is not limited and two samples are sufficient to split an inner node. The best split is always chosen based on the mean squared error.
- **Random Forest (RF**.) The statistically motivated regressor was employed using an ensemble of 500 decision trees. As with single decision tree regression, the random forest trees are not limited in-depth and a split requires at least two samples. The best split is chosen based on the mean squared error.
- **Multilayer Perceptron (MLP).** This biologically inspired neural network regressor was used with the seven-layers network architecture $[m, 256, 128, 64, 32, 16, 1]$, where m denotes the dimensionality of the input vector. Thus, the number of input neurons depends on the input feature set. Rectified linear units (ReLU) are used as the activation function. For training, the Adam optimizer of Kingma et al. was used [43]. This is a variant of the stochastic gradient descent method with backpropagation and adaptive learning rate. The initial learning rate was 0.001. In addition, an L2 regularization term with an α of 0.0001 was used. 200 samples were combined to a mini-batch.

Focussing on selecting effective feature sets, extensive optimization of hyperparameters was omitted for the regression methods at this point, although these were chosen based on experience from past work. For all supervised learning tasks, machine learning methods were used as provided by the Python library *scikit-learn* [44].

C. Results

By comparing prediction results from the 648 trained machine learning models, we identified feature sets best suited for the given task of determining R^{ap}, R^{bl}, etc (Table V). The underlying assumption was that prediction accuracies will differ depending on the input feature set and the machine learning model used.

From all 648 trained models, we show results for the best 162 models in Table V. For all target variables, regression models were considered that estimate with a MAPE of less than 5%. For every single approach and target, the feature set that leads to the minimum *MAPE* value was selected. The arithmetic mean over the APE values of all test samples *MAPE* is a relative metric independent of actual target values and ranges. It is particularly well suited as a measure to compare the prediction accuracies among different target variables.

For all cell lines, targets, and machine learning models, the feature sets of the second approach perform better than the first approach (cf. Table V). This is especially evident in the estimation of apical and basolateral features. In many cases, the *MAPE* can further be reduced by adding extra features using the third approach. For example, in predicting C^{bl} for IPEC-J2, the first approach achieves a *MAPE* of 18.7%, while the *MAPE* for the second approach is 2.5% and 2.0% for the third approach (cf. Table V).

TABLE V. MAPE FOR PREDICTING EPITHELIAL PROPERTIES [±%].

a) Estimation of C^{ap}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	8.43	7.01	6.90
	2	$I^2_{\angle(Z)}$	2.88	1.84	**0.66**
	3	$I^3_{\perp(Y)}$	**2.60**	**1.64**	0.67
IPECJ2	1	$I^1_{\perp(Y)}$	9.77	8.13	7.73
	2	$I^2_{\perp(Y)}$	**3.82**	**2.79**	**1.02**
	3	$I^3_{\perp(Y)}$	2.85	1.92	1.31
MDCKI	1	$I^1_{\perp(Y)}$	5.58	4.79	5.84
	2	$I^2_{\perp(Y)}$	2.87	2.10	0.93
	3	$I^3_{\perp(Y)}$	**2.24**	**1.50**	**0.77**

c) Estimation of R^{sub}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Z)}$	3.24	2.22	1.42
	2	$I^2_{\perp(Z)}$	**2.65**	**1.76**	1.31
	3	$I^3_{\perp(Z)}$	2.67	1.78	**1.27**
IPECJ2	1	$I^1_{\perp(Z)}$	3.75	2.56	2.05
	2	$I^2_{\perp(Z)}$	**3.02**	**1.96**	**1.32**
	3	$I^3_{\perp(Z)}$	3.06	1.98	2.66
MDCKI	1	$I^1_{\perp(Z)}$	4.88	3.36	2.49
	2	$I^2_{\perp(Z)}$	**3.90**	**2.60**	**2.45**
	3	$I^3_{\perp(Z)}$	3.95	2.62	2.49

e) Estimation of R^{bl}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	50.96	44.06	34.81
	2	$I^2_{\perp(Z)}$	5.54	5.26	2.34
	3	$I^3_{\perp(Z)}$	**3.06**	**2.98**	**1.99**
IPECJ2	1	$I^1_{\perp(Y)}$	44.87	36.85	25.07
	2	$I^2_{\perp(Z)}$	10.35	8.19	4.90
	3	$I^3_{\perp(Y)}$	**6.04**	**5.26**	**3.56**
MDCKI	1	$I^1_{\perp(Y)}$	35.08	27.51	22.66
	2	$I^2_{\perp(Z)}$	6.66	5.59	1.93
	3	$I^3_{\angle(Y)}$	**4.42**	**3.78**	**1.78**

b) Estimation of C^{bl}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	45.13	40.73	33.33
	2	$I^2_{\perp(Z)}$	9.25	5.56	**1.65**
	3	$I^3_{\perp(Z)}$	**8.76**	**5.44**	2.02
IPECJ2	1	$I^1_{\perp(Y)}$	24.53	20.49	18.70
	2	$I^2_{\perp(Z)}$	8.37	5.62	2.53
	3	$I^3_{\perp(Y)}$	**5.71**	**3.72**	**2.00**
MDCKI	1	$I^1_{\perp(Y)}$	12.86	11.09	11.53
	2	$I^2_{\perp(Z)}$	6.66	4.38	1.85
	3	$I^3_{\perp(Y)}$	**5.55**	**3.49**	**1.09**

d) Estimation of R^{ap}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	20.73	17.03	13.43
	2	$I^2_{\perp(Z)}$	6.03	6.43	2.58
	3	$I^3_{\perp(Z)}$	**4.46**	**4.76**	**2.21**
IPECJ2	1	$I^1_{\perp(Y)}$	37.64	31.55	22.92
	2	$I^2_{\perp(Y)}$	22.60	18.73	4.85
	3	$I^3_{\perp(Y)}$	**16.45**	**15.11**	**4.25**
MDCKI	1	$I^1_{\perp(Y)}$	31.11	24.39	19.74
	2	$I^2_{\perp(Z)}$	10.48	10.28	**2.93**
	3	$I^3_{\perp(Z)}$	**8.78**	**8.59**	3.04

f) Estimation of R^{para}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Z)}$	7.25	6.95	5.43
	2	$I^2_{\perp(Z)}$	2.57	2.59	**0.75**
	3	$I^3_{\perp(Z)}$	**2.07**	**2.03**	1.07
IPECJ2	1	$I^1_{\perp(Y)}$	17.36	13.82	9.76
	2	$I^2_{\perp(Z)}$	7.81	6.84	**3.38**
	3	$I^3_{\perp(Z)}$	**6.94**	**5.90**	4.26
MDCKI	1	$I^1_{\perp(Y)}$	16.08	13.02	9.57
	2	$I^2_{\perp(Z)}$	5.67	4.93	2.16
	3	$I^3_{\perp(Z)}$	**5.00**	**4.17**	**1.36**

TABLE VI. SELECTED ROBUST FEATURE SETS

Target	HT-29/B6		IPEC-J2		MDCK I	
	Feature Set	*Model*	*Feature Set*	*Model*	*Feature Set*	*Model*
C^{ap}	$I^2_{\angle(Y)}$	*MLP*	$I^2_{\angle(Z)}$	*MLP*	$I^2_{\angle(Z)}$	*MLP*
C^{bl}	$I^2_{\angle(Y)}$	*MLP*	$I^2_{\angle(Z)}$	*MLP*	$I^2_{\angle(Z)}$	*MLP*
R^{sub}	$I^2_{\perp(Z)}$	*MLP*	$I^2_{\perp(Z)}$	*MLP*	$I^2_{\angle(Z)}$	*MLP*
R^{ap}	$I^2_{\angle(Z)}$	*RF*	$I^2_{\perp(Y)}$	*MLP*	$I^3_{\angle(Y)}$	*RF*
R^{bl}	$I^2_{\angle(Y)}$	*MLP*	$I^2_{\perp(Y)}$	*MLP*	$I^3_{\angle(Y)}$	*RF*
R^{para}	$I^2_{\perp(Y)}$	*RF*	$I^2_{\angle(Y)}$	*MLP*	$I^3_{\perp(Y)}$	*RF*
R^{ap}_2	$I^2_{\angle(Z)}$	*MLP*	$I^3_{\perp(Y)}$	*RF*	$I^2_{\angle(Y)}$	*MLP*

TABLE VII. ABSOLUTE PERCENTAGE ERROR FOR MACHINE LEARNING WITH ROBUST FEATURE SETS

	HT-29/B6				IPEC-J2				MDCK I			
	mean	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*
C^{ap}	0.7	0.5	1.8	9.1	1.3	0.9	3.7	20.0	0.8	0.7	2.2	12.0
C^{bl}	1.2	0.9	3.4	35.9	2.1	1.6	5.9	34.2	1.3	1.0	3.4	31.9
R^{sub}	1.3	1.0	3.7	9.9	1.3	0.9	4.1	20.0	2.3	1.7	6.4	20.7
R^{ap}	8.8	4.8	30.5	91.6	4.8	2.5	15.5	303.0	8.6	3.4	36.6	176.9
R^{bl}	1.5	1.0	4.1	38.2	3.3	2.1	10.5	99.8	3.8	1.8	14.6	56.5
R^{para}	4.7	3.7	13.6	29.2	3.3	1.9	10.6	115.4	6.9	4.3	23.1	67.5
R^{ap}_2	10.2	6.4	32.7	152.8	14.9	7.9	52.4	320.4	11.8	7.8	36.4	215.2

Among the machine learning models, the MLPs lead to the smallest *MAPE* values. In comparison to the baseline method of the decision tree, however, the random forest also yielded an improvement of the estimates in the vast majority of cases.

D. Outlier and Robust Feature Sets

Even though it could be shown that machine learning may yield a very accurate determination of epithelial properties, in a few cases very high relative deviations occur. In practice, however, even a few extreme outliers may be more harmful than lower average accuracies. To this end, an approach that avoids extreme outliers is favorable and a reliable method is needed that provides useful estimates even in the worst cases.

Therefore, in an alternative approach, feature sets and machine learning models were selected for which the smallest relative deviations (i.e., APE values) occur in the worst case within the test samples. These alternative feature sets for all cell lines and target variables are displayed in Table VI and referred to as robust feature sets in the following.

Table VII shows the resulting distribution of the absolute percentage error for the selected robust feature sets. C^{ap}, e.g., may be determined for HT-29/B6, IPEC-J2, and MDCK I with a worst-case deviation of 20%. For 11 out of 21 models, maximum APE values are less than 50 percent. For five models, maximum APE values are less than 25 percent. It can be seen, though, that even with robust feature sets, very high maximum deviations may still occur for individual target variables. For example, the determination of R^{ap} of IPEC-J2 has an average deviation of 4.8%, but in the worst case a deviation of 303.0% (cf. Table VII). As percentiles ($P_{95\%}$) indicate, however, such high prediction errors occur very rarely.

V. Postprocessing using Least Squares Approach

After completing the previous step, all components of an assumed equivalent electric circuit are determined by machine learning with the respective accuracy (section IV). Having obtained estimates for all circuit parameters, however, one also obtains the possibility of synthesizing impedance spectra for the given values of the circuit parameters. This is the basis for the postprocessing step that we propose here.

Analogously to the generation of training data (section III), the determined parameter values are used to model two synthetic impedance spectra for control conditions and Nystatin application. Both newly synthesized spectra are then compared to the originally observed spectra. The least-squares method is then employed to adjust the parameter values so that the synthesized pair of spectra matches the original.

A. Approach

Using the machine learning predictions, a new pair of impedance spectra $\hat{Z} = (\hat{Z}_1, \ldots, \hat{Z}_{84})^\top$ is calculated using equation 2. Note, that one does not necessarily obtain an equivalent circuit whose corresponding pair of spectra matches the original observed pair of spectra $Z = (Z_1, ..., Z_{84})^\top$, as described in the next Chapter. The Residual Sum of Squares (*RSS*) is a suitable measure for this purpose, which is calculated according to the following formula:

$$RSS(Z, \hat{Z}) = \sum_{i=1}^{84} (Z_i - \hat{Z}_i)^2 \tag{22}$$

It is desirable that the predicted equivalent circuit exhibit similar electrical behavior to the measured cell, i.e., have a minimum *RSS*. For this purpose, the nonlinear least-squares curve fit method from the Python library SciPy was applied [45]. The *Trust Region Reflective* variant was chosen because of its capability to specify upper and lower bounds for each parameter when calling the method. The predictions of the machine learning models with Robust Feature Sets serve as initial values for the optimization from which a local minimum of the *RSS* is searched for. The bounds from Table II were used and the measurement points were weighted with the standard deviations from the error modeling of section III-B.

Note, that both improvement and deterioration of the values are possible. In previous attempts, no improvement of R^{sub} could be achieved. For this reason and because the predictions of the machine learning models were already sufficiently good, R^{sub} was fixed to its initial value.

B. Results

We first present results on how well the *RSS* could be minimized by the least-squares approach. Secondly, we evaluate relative deviations (APE) of the equivalent circuit parameters.

1) Reduction of the Residual Sum of Squares: For all test samples, the solver was able to obtain parameter values with a lower *RSS* compared to the initial values of the machine learning. For all cell lines, the average reduction of *RSS* by optimization is greater than 99%. (cf. Figure 5). An example of a pair of impedance spectra before and after post-processing is shown in Figure 6. Looking at the 84 complex measuring points, the average *RSME* could be reduced from 20.2 $\Omega \cdot cm^2$ to 2.0 $\Omega \cdot cm^2$ for HT-29/B6, from 67.6 $\Omega \cdot cm^2$ to 2.6 $\Omega \cdot cm^2$ for IPEC-J2 and from 61.1 $\Omega \cdot cm^2$ to 2.5 $\Omega \cdot cm^2$ for MDCK I.

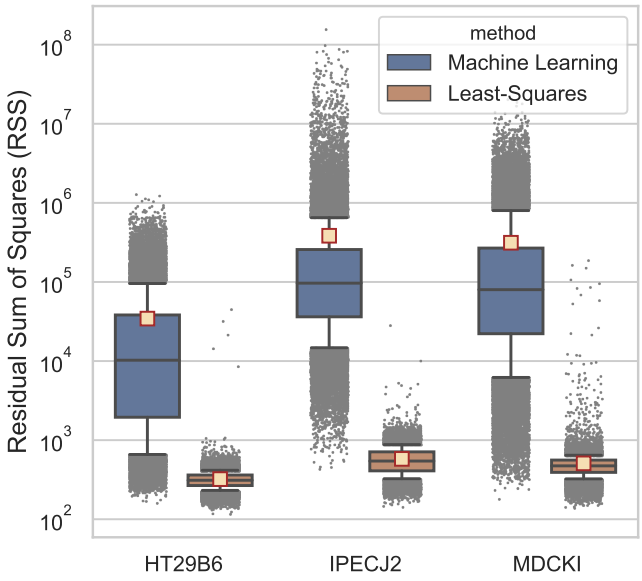


Figure 5. Residual Sum of Squares (*RSS*) of predicted impedance spectra pair before and after least-squares postprocessing as box plots with arithmetic mean (□) and whiskers showing the 10th and 90th percentiles.

2) Circuit parameters: Table VIII shows a detailed overview of derived *APE* values after least-squares postprocessing. Note, that the deviations for R^{sub} do not change because the parameter was fixed during execution. For the targets R_2^{ap}, R^{para} and R^{ap}, a significant improvement is observed for all cell lines. The largest improvement is seen for R_2^{ap}, where, for example, for MDCK I the *MAPE* is reduced from 11.8% to 0.4% and also the maximum *APE* is 99.7% instead of 215.2%. It becomes clear that post-processing delivers better results especially when good initial values are already available, whereas hardly any improvements can be found for the worst-case samples with maximum *APE*. This can be seen in the statistical distribution of *APE*, where, for example, for C^{bl} the median improves for all cell lines, whereas the maximum *APE* increases. The median *APE* worsens in only one case, namely in the determination of R^{bl} for HT/29-B6 by 0.1%. Otherwise, clear improvements can be observed.

TABLE VIII. ABSOLUTE PERCENTAGE ERROR AFTER POSTPROCESSING

	HT-29/B6				IPEC-J2				MDCK I			
	mean	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*
C^{ap}	0.3	0.2	1.0	9.5	0.3	0.1	1.0	17.7	0.3	0.1	1.2	18.3
C^{bl}	1.3	0.8	4.2	45.1	0.6	0.2	2.4	50.7	0.7	0.2	2.7	44.3
R^{sub}	1.3	1.0	3.7	9.9	1.3	0.9	4.1	20.0	2.3	1.7	6.4	20.7
R^{ap}	4.2	1.8	17.2	76.2	3.0	0.7	12.0	309.7	3.4	0.8	14.6	343.1
R^{bl}	2.2	1.1	7.9	36.6	1.3	0.3	5.3	98.8	1.4	0.3	6.4	94.3
R^{para}	1.7	0.7	6.7	38.3	1.0	0.3	3.6	84.5	1.7	0.4	6.9	120.5
R_2^{ap}	0.9	0.5	3.1	99.6	0.4	0.2	1.2	212.9	0.4	0.3	1.2	99.7

TABLE IX. FEATURE SUBSET IMPORTANCE FOR PREDICTIONS WITH RANDOM FOREST AND $I^3_{\perp(Y)}$

Features	C^{bl}			R^{bl}			R^{para}			R^{sub}		
	HT-29/B6	IPEC-J2	MDCK I	HT-29/B6	IPEC-J2	MDCK I	HT-29/B6	IPEC-J2	*MDCK I*	HT-29/B6	IPEC-J2	MDCK I
$S_{\Re(Y)} \times S_{\Im(Y)}$	1.3	1.6	2.8	2.5	0.7	3.1	78.6	68.6	73.1	0.2	0.2	1.6
$N_{\Re(Y)} \times N_{\Im(Y)}$	6.5	4.8	24.5	5.3	1.4	4.7	1.2	1.3	3.2	0.2	0.2	0.6
$D_{\Re(Y)} \times D_{\Im(Y)}$	42.9	35.0	23.3	69.4	87.7	78.9	1.9	2.5	4.1	0.1	0.1	0.4
$R_{\Re(Y)} \times R_{\Im(Y)}$	26.5	5.6	7.3	7.6	3.4	2.1	4.4	5.1	11.1	0.1	0.1	0.4
$\mathrm{stat}\left(S_{\Re(Y)}\right) \times \mathrm{stat}\left(S_{\Im(Y)}\right)$	2.3	2.0	1.4	1.3	2.0	1.7	10.2	19.5	3.2	93.3	88.3	10.4
$\mathrm{stat}\left(N_{\Re(Y)}\right) \times \mathrm{stat}\left(N_{\Im(Y)}\right)$	1.3	47.9	37.6	1.5	0.3	1.3	0.4	0.5	0.8	6.0	11.1	86.3
$\mathrm{stat}\left(D_{\Re(Y)}\right) \times \mathrm{stat}\left(D_{\Im(Y)}\right)$	8.5	1.8	1.4	9.6	1.7	6.5	0.7	0.6	1.0	0.0	0.0	0.2
$\mathrm{stat}\left(R_{\Re(Y)}\right) \times \mathrm{stat}\left(R_{\Im(Y)}\right)$	10.8	1.3	1.7	2.8	2.8	1.8	2.7	1.9	3.5	0.0	0.0	0.2

VI. Discussion

A. Relevance for Biomedical Applications

Impedance spectroscopic techniques are increasingly gaining importance in biomedical applications like monitoring the growth of cultured epithelial and endothelial cells (e.g., retinal pigment epithelium, gastrointestinal tract cells, pulmonary cells, blood-brain-barrier models [46][47]), or alterations of barrier function during pharmacological studies [48][49][46]. Furthermore, impedance spectroscopy is the only technique that allows functional distinction between epithelial and subepithelial properties of ex vivo tissue, such as intestinal biopsies of patients with suspected barrier impairment.

Reliable, automated determination of epithelial properties is becoming increasingly important, as, during the past decade, impedance spectroscopy of epithelia has started to move from pure basic research or investigations on biopsies to clinical applications in patients. Several groups have developed various endoscopic devices that allow in vivo impedance measurements, e.g., in the GI tract of critically ill patients that may develop mucosal ischemia [50] or in the esophagus from patients suffering from gastroesophageal reflux disease [51][52][53].

If the technique is to be used on a routine basis, however, reliable automatization for the evaluation of impedance spectra is indispensable. On one hand, manual evaluation of impedance spectra to extract the physiologically relevant parameters requires extensive user training and is time-consuming, as individual spectra need to be fitted by complex equations [24]. On the other hand, currently available systems usually only record R^{trans} (i.e., the sum of the subepithelial and epithelial resistance) or even only relative alterations in R^{trans} over the time-course of an experiment, as estimation of absolute R^{trans} values is too error-prone [54][46].

B. Combined Consideration of Two-State Measurements

The accurate determination of apical and basolateral properties with one-path impedance spectroscopy has been considered an unsolved problem in clinical physiology. In this study, we have been able to demonstrate that this issue can be solved by observing the cell under two different conditions, or states, namely before and after the apical addition of Nystatin. It became clear that machine learning techniques are well suited to process pairs of impedance spectra, with a significant improvement compared to determination based on a single measurement under control conditions alone (see Table V). First attempts to construct new features from two different spectra were successful and showed that this step improves the determination of epithelial properties by machine learning.

C. Feature Extraction

From the results shown in section IV, we can draw some preliminary conclusions regarding the usefulness of the extracted features and feature sets.

1) Representations: In the context of this study, none of the selected representations showed to be consistently superior. However, among the regression models with the smallest MAPE, algebraic forms of impedance and admittance are noticeably more frequent than polar forms (cf. Table V). The average improvement of the MAPE by the right choice of representation is about 0.9 percentage points. Therefore, we assume that the representation does not play a major role in the method presented here.

2) Feature Importance: Scikit-learn's Random Tree Regressor provides the output of feature importances. The value of feature importance lies in the interval [0; 1] and is the larger, the greater the influence on the regression. Notice that the sum of all feature importances of an input feature set is equal to 1. The input feature set can be divided into subsets and compared to each other by summing the individual importances. Table IX shows the feature subset importances for the random forest regressions of C^{bl}, R^{bl}, R^{para} and R^{sub} using $I^3_{\perp(Y)}$ as input.

3) Statistical Features: The extraction of statistical features has worked well in previous work related to epithelial impedance analysis, which is why the approach was retained for this work [1]. The benefit of using statistical features varies for different targets. In Table IX, e.g., it can clearly be seen that the statistical features are of high importance in predicting R^{sub} and C^{bl} through random forest. This is particularly evident in the prediction of R^{sub}. Here, with a summed feature importance of $\geq 97\%$, decisions are made almost entirely based on the statistical features. For C^{bl} of IPEC-J2 and MDCK I, the statistical features in $I^3_{\perp(Y)}$ obtain high importance and could contribute to a $0.9 - 1.9$ percentage point improvement in prediction (cf. Table Vb). Because of the ease of calculation and the possible positive effect on prediction, the use of descriptive statistics for feature extraction in the analysis of impedance spectra can be further recommended.

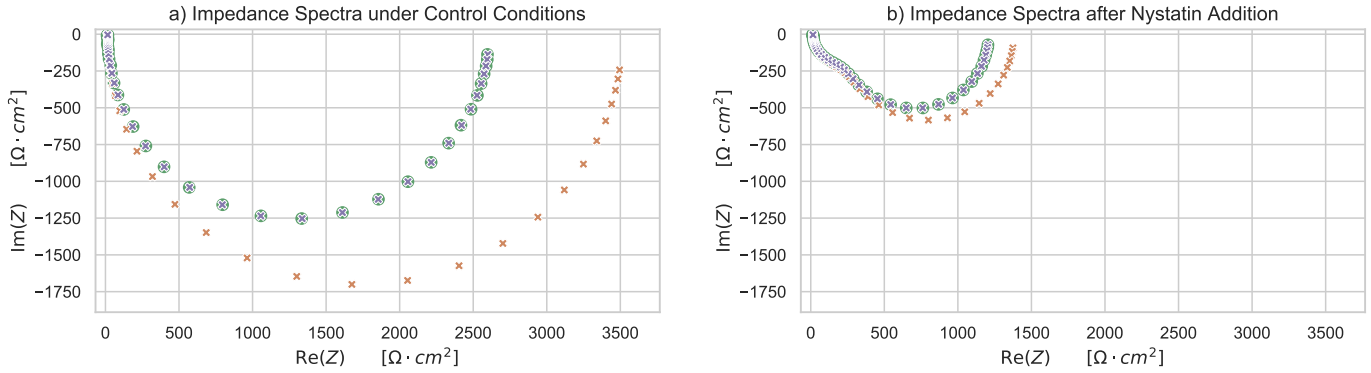


Figure 6. Pair of impedance spectra before and after Nystatin addition: Original measurement from data set (●), pair of spectra derived from Machine Learning predictions (✖), pair of spectra derived from predictions after least-squares postprocessing (✖)

4) Differences and Ratios: The frequency-wise calculation of the differences and ratios from the pairs of spectra is a completely new approach resulting from the two-state consideration. Looking at the feature importance from Table IX, it becomes very clear that the features extracted in this way work very well for transcellular targets such as C^{bl} and R^{bl} with an feature importance between 31% and 91%. On the other hand, for the quantities that do not belong to the transcellular route, i.e., R^{para} and R^{sub}, this approach is of little use. Differences and ratios not only yielded high feature importance for random forests but were also able to contribute to a 1.8−2.8 percentage point improvement in prediction (see Table Ve). Additionally, more complex features could be extracted from the two-state consideration to further improve prediction.

D. Machine Learning and Least-Squares Optimization

In this work, machine learning methods following statistical and neuro-inspired paradigms were used for the pairwise analysis of epithelial impedance spectra. In particular, the *MAPE* values from Table V show that with the right combination of feature set and machine learning model, very good predictions of all sought targets are possible. With a maximum training time of a regression model of a few hours on a 4-core CPU, the overall training period was acceptable despite the large sample size. Therefore, no further time measurements were taken during training. A drawback of the presented study is the brute-force approach in the search for optimal feature sets. Here, an elaborated search for optimal feature sets would have saved training effort, for example in the unprofitable use of different impedance representations.

This computational study was able to show the promising influence of least-squares postprocessing for machine-learning-based analysis. Figure 6 shows the original impedance spectra of a measurement from the data set and the spectra derived from the machine learning predictions with and without least-squares optimization. It can be seen that the machine learning predictions without least-squares optimization may result in impedance spectra that are significantly different from the measured ones. In this respect, posterior optimization using a least-squares approach has been shown to be a suitable approach to better align a predicted pair of curves to the observed ones. The greatest empirical evidence represents the average reduction of the *RSS* by more than 99% for all cell lines (see Figure 5). Due to the much better alignment, the equivalent circuit obtained after optimization is much better suited to describe the electrical behavior of the cell. Moreover, with the help of the least-squares optimization, an improvement in the predictions of R^{para}, R^{ap} and R_2^{ap} by several percentage points could be achieved.

Least-squares methods, as used in our postprocessing, usually require suitable initial values from which a local minimum is found. The choice of suitable initial values for optimization problems is a frequent topic in various disciplines [55][56]. In the field of epithelial impedance spectroscopy, the use of randomly chosen initial values may lead to poor convergence behavior. Manual initialization by an expert, on the other hand, offers no possibility of automation. The presented method could fill this gap by using machine learning to find suitable initial values in an automated way. Accordingly, e.g., in the field of crystal lattice determination, machine learning techniques have recently been used to find suitable initial values for subsequent optimization [57].

E. Limitations

A notable limitation of the presented method is the occurrence of highly increased prediction errors in a few cases, especially for R^{ap} with a maximal relative deviation of more than 300% for MDCK I and IPEC-J2 (see Table VIII). Because the apical resistance decreases after apical addition of Nystatin, R^{ap} is the only parameter that changes during the experimental setup, while the others remain the same. For a reliable approximation of R^{ap}, basolateral addition of Nystatin might be useful to keep R^{ap} constant during both measurements.

Looking at the absolute percentage error plotted against the Nystatin factor d_{nyst}, we see for all targets and cell lines that the highly increased prediction errors occur only for $d_{nyst} < 0.4$. In Figure 7, this can be seen for R^{ap}, R^{bl}, and R^{para} as examples. Note, that in the given experimental setup, the amount of Nystatin added can freely be chosen. Based on this, we filtered and re-evaluated the test data set for samples with $d_{nyst} \geq 0.4$ (see Table X). For the reduced test data set, the overall performance of the proposed method is significantly better with a maximum APE of 36.5%. However, the assumption $d_{nyst} \geq 0.4$ is not applicable to the general case, because smaller d_{nyst} values, derived from a substantial

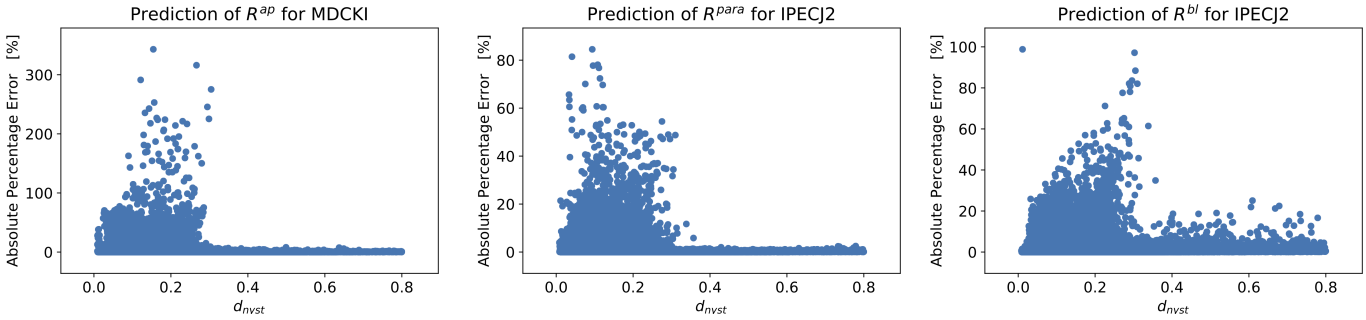


Figure 7. Absolute percentage error plotted against the Nystatin factor d_{nyst}.

addition of Nystatin, are often required to distinguish apical and basolateral membranes in the impedance curve.

As the presented method to determine electric properties of epithelia involves modeled impedance spectra, applicability is naturally depending on the modeling and sampling of the data. The previously modeled data that is used here is in good accordance with data obtained from measurements on epithelial tissue [25]. While a variety of cell lines and functional states is considered in our work, results and estimations obtained here are still limited to the specific scenarios modeled. Also, characteristics of the training and test data influence characteristics of estimation methods. For example, the precision of supervised learning methods tends to decrease if the number of samples decreases. To this end, e.g., the distribution of target values needs to be considered (cf. Figure 4).

VII. Conclusions and Future Work

While impedance spectroscopy is a convenient measurement technique, quantifying the electric bipolarity of epithelial tissue with traditional approaches is challenging. For some parameters, it may not only be error-prone, but practically impossible with reasonable efforts. In this computational study, we have shown that this aim may be achieved with good results by combining machine learning regression with a least squares-based postprocessing step. Key concepts were the combination of measurements obtained under two different tissue conditions as well as a systematic feature extraction approach. Due to detailed and realistic modeling, we suggest this approach is valid for the epithelial cell lines HT-29/B6, IPEC-J2 and MDCK I under control conditions, as well as under the influence of Nystatin and EGTA.

All in all, the method outlined and evaluated in the present study represents a significant progress in order to achieve a routine evaluation of impedance spectra obtained in a clinical measurement setting. Several challenges are envisioned, however, that need to be adressed in future work:

1) Compared to impedance spectra recorded from cultured cell layers, data obtained from patients will be subject to a considerably larger scatter within one impedance spectrum, as electrical shielding is limited.
2) There will be a substantial variability between individual spectra, due to variation in the positioning of the electrodes.
3) Larger artifacts have to be recognized, e.g., caused by movements of the patient.
4) In contrast to the GI tract, esophageal epithelia are stratified. Thus, valid equivalent circuits for stratified epithelia have to be developed and tested, before data can be modeled accurately.

A promising strategy is to solve these open questions in our current approach would be the application of generative machine learning methods, which are designed to rebuild characteristics of given training data. Although techniques such as Generative Adversarial Networks (GAN) were originally designed to mimic image data [58], we see much potential in employing them to model spectral data. In future work, we will therefore investigate the usage of such state-of-the-art generative machine learning methods for advanced and highly automated modeling of impedance spectra.

Appendix

A. *Prediction for samples with $d_{nyst} \geq 0.4$*

In the discussion section, we have pointed out some notable effects of the Nystatin application on the reliability of predictions. As can be seen in Figure 7, high prediction errors occur exclusively in samples with a small Nystatin factor d_{nyst}. Therefore, the test samples were filtered by $d_{nyst} \geq 0.4$ and re-evaluated. From the original 50,000 samples, we obtained reduced data sets with 11043 Samples for HT29B6, 3526 for IPECJ-2, and 2448 for MDCK I. The results of the evaluation are shown in Table X.

TABLE X. ABSOLUTE PERCENTAGE ERROR AFTER POSTPROCESSING FOR SAMPLES WITH $d_{nyst} \geq 0.4$

	HT-29/B6				IPEC-J2				MDCK I			
	mean	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*
C^{ap}	0.2	0.1	0.5	4.7	0.2	0.1	0.6	4.6	0.1	0.1	0.4	2.4
C^{bl}	1.4	1.0	4.2	45.1	0.8	0.4	2.7	21.2	0.6	0.4	1.7	9.7
R^{sub}	1.4	1.0	3.9	8.3	1.5	1.0	4.5	18.2	2.3	1.7	6.9	14.9
R^{ap}	1.8	1.0	6.0	36.5	0.6	0.3	2.2	13.0	0.4	0.3	1.4	8.1
R^{bl}	2.4	1.4	7.8	33.0	1.0	0.5	3.7	25.1	0.8	0.5	2.5	17.6
R^{para}	0.9	0.5	2.9	15.0	0.2	0.2	0.7	2.5	0.3	0.2	0.9	3.1
R_2^{ap}	1.0	0.6	3.6	17.3	0.3	0.2	1.2	8.8	0.2	0.2	0.7	5.0

References

[1] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the resistance of epithelial cell layers from an impedance spectrum," in Proceedings of the Tenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 2018, pp. 8—13.

[2] V. Singh et al., "Translating molecular physiology of intestinal transport into pharmacologic treatment of diarrhea: stimulation of na+ absorption," Clinical Gastroenterology and Hepatology, vol. 12, no. 1, 2014, pp. 27–31.

[3] S. Priyamvada et al., "Mechanisms underlying dysregulation of electrolyte absorption in inflammatory bowel disease–associated diarrhea," Inflammatory bowel diseases, vol. 21, no. 12, 2015, pp. 2926–2935.

[4] J. M. Diamond, "Twenty-first bowditch lecture. the epithelial junction: bridge, gate, and fence," Physiologist, vol. 20, no. 1, Feb 1977, pp. 10–18.

[5] M. Furuse et al., "Occludin: a novel integral membrane protein localizing at tight junctions," Journal of Cell Biology, vol. 123, no. 6, 1993, pp. 1777–1788.

[6] J. Ikenouchi et al., "Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells," Journal of Cell Biology, vol. 171, no. 6, Dec 2005, pp. 939–945.

[7] M. Furuse, H. Sasaki, K. Fujimoto, and S. Tsukita, "A single gene product, claudin-1 or -2, reconstitutes tight junction strands and recruits occludin in fibroblasts," The Journal of Cell Biology, vol. 143, no. 2, 1998, pp. 391–401.

[8] M. Furuse, H. Sasaki, and S. Tsukita, "Manner of interaction of heterogeneous claudin species within and between tight junction strands," The Journal of Cell Biology, vol. 147, no. 4, 1999, pp. 891–903.

[9] S. M. Krug et al., "Charge-selective claudin channels," Annals of the New York Academy of Sciences, vol. 1257, 2012, pp. 20–28.

[10] J. Luettig, R. Rosenthal, C. Barmeyer, and J. Schulzke, "Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation," Tissue Barriers, vol. 3, no. 1-2, 2015, p. e977176, pMID: 25838982.

[11] C. Barmeyer, J. D. Schulzke, and M. Fromm, "Claudin-related intestinal diseases," Seminars in Cell & Developmental Biology, vol. 42, 2015, pp. 30 – 38, claudins Time, Space and the Vertebrate Body Axis.

[12] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the capacitance of epithelial cell layers from an impedance spectrum," in Proceedings of the 7th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 2015, pp. 27–32.

[13] T. Schmid, M. Bogdan, and D. Günzel, "Discerning apical and basolateral properties of HT-29/B6 and IPEC-J2 cell layers by impedance spectroscopy, mathematical modeling and machine learning," PLOS ONE, vol. 8, no. 7, 2013, p. e62913.

[14] B. Jovov, N. Wills, and S. Lewis, "A spectroscopic method for assessing confluence of epithelial cell cultures," American Journal of Physiology, vol. 261, no. 6 Pt 1, 1991, pp. C1196–203.

[15] J. R. Macdonald and W. B. Johnson, Fundamentals of Impedance Spectroscopy. John Wiley & Sons, Inc., 2005, pp. 1–26.

[16] K. S. Cole, "Electric phase angle of cell membranes," The Journal of general physiology, vol. 15, no. 6, 1932, p. 641.

[17] T. Teorell, "Application of "square wave analysis" to bioelectric studies1," Acta Physiologica Scandinavica, vol. 12, no. 2-3, 1946, pp. 235–254.

[18] E. Schifferdecker and E. Frömter, "The ac impedance of necturus gallbladder epithelium," Pflügers Archiv, vol. 377, no. 2, 1978, pp. 125–133.

[19] G. Kottra and E. Frömter, "Rapid determination of intraepithelial resistance barriers by alternating current spectroscopy. i. experimental procedures," Pflügers Archiv, vol. 402, no. 4, 1984, pp. 409–420.

[20] M. Fromm, C. E. Palant, C. J. Bentzel, and U. Hegel, "Protamine reversibly decreases paracellular cation permeability in necturus gallbladder," The Journal of membrane biology, vol. 87, no. 2, 1985, pp. 141–150.

[21] W. Van Driessche et al., "Interrelations/cross talk between transcellular transport function and paracellular tight junctional properties in lung epithelial and endothelial barriers," American Journal of Physiology-Lung Cellular and Molecular Physiology, vol. 293, no. 3, 2007, pp. L520–L524.

[22] D. Günzel et al., "From TER to trans-and paracellular resistance: lessons from impedance spectroscopy," Annals of the New York Academy of Sciences, vol. 1257, no. 1, 2012, pp. 142–151.

[23] S. M. Krug, M. Fromm, and D. Günzel, "Two-path impedance spectroscopy for measuring paracellular and transcellular epithelial resistance." Biophys J, vol. 97, no. 8, 2009, pp. 2202–2211.

[24] T. Schmid, D. Günzel, and M. Bogdan, "Efficient prediction of x-axis intercepts of discrete impedance spectra," in Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2013, pp. 185–190.

[25] T. Schmid, "Automatisierte Analyse von Impedanzspektren mittels konstruktivistischen maschinellen Lernens," Ph.D. dissertation, Universität Leipzig, Germany, 2018.

[26] M. E. Orazem and B. Tribollet, Eds., Electrochemical Impedance Spectroscopy. Wiley, 2008, ch. Methods for Representing Impedance, pp. 309–331.

[27] K. Suzuki, G. Kottra, L. Kampmann, and E. Frömter, "Square wave pulse analysis of cellular and paracellular conductance pathways in necturus gallbladder epithelium," Pflügers Archiv, vol. 394, 1982, pp. 302–312.

[28] C. Bertrand, D. Durand, G. Saidel, C. Laboisse, and U. Hopfer, "System for dynamic measurements of membrane capacitance in intact epithelial monolayers," Biophysical Journal, vol. 75, no. 6, 1998, pp. 2743–2756.

[29] W.-M. Weber, H. Cuppens, J.-J. Cassiman, W. Clauss, and W. Van Driessche, "Capacitance measurements reveal different pathways for the activation of cftr," Pflügers Archiv, vol. 438, no. 4, 1999, pp. 561–569.

[30] I. El Naqa and M. J. Murphy, "What is machine learning?" in Machine Learning in Radiation Oncology. Springer, 2015, pp. 3–11.

[31] A. Krogh, "What are artificial neural networks?" Nature Biotechnology, vol. 26, no. 2, Feb 2008, pp. 195–197.

[32] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural networks, vol. 2, no. 5, 1989, pp. 359–366.

[33] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, no. 1, 1986, pp. 81–106.

[34] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, 2001, pp. 5–32.

[35] H. Liu and H. Motoda, Feature Extraction, Construction and Selection: A Data Mining Perspective. Springer Science & Business Media, 1998, vol. 453.

[36] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, 2003, pp. 1157–1182.

[37] J. Kittler, "Feature selection and extraction," Handbook of Pattern Recognition and Image Processing, 1986.

[38] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the relation between resistor-capacitor subcircuits from an impedance spectrum," in Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies, ser. BIOSTEC 2014, 2014, p. 141–148.

[39] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the relation between resistor-capacitor subcircuits from an impedance spectrum," in Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, 2014, pp. 141–148.

[40] B. Schindler, D. Günzel, and T. Schmid, "Synthesized Impedance Spectra Measurements of Epithelial Tissue," Nov. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5718939

[41] T. Schmid, "Automatisierte analyse von impedanzspektren mittels konstruktivistischen maschinellen lernens," Ph.D. dissertation, Leipzig, 2018.

[42] S. Lewis, D. Eaton, C. Clausen, and J. Diamond, "Nystatin as a probe for investigating the electrical properties of a tight epithelium," The Journal of general physiology, vol. 70, 11 1977, pp. 427–40.

[43] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 12 2014.

[44] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.

[45] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," Nature Methods, 2020.

[46] B. Srinivasan et al., "TEER measurement techniques for in vitro barrier model systems," Journal of Laboratory Automation, vol. 20, no. 2, 2015, pp. 107–126.

[47] N. Onnela et al., "Electric impedance of human embryonic stem cell-derived retinal pigment epithelium," Medical & Biological Engineering & Computing, vol. 50, no. 2, 2012, pp. 107–116.

[48] F. A. Atienzar, H. Gerets, K. Tilmant, G. Toussaint, and S. Dhalluin, "Evaluation of impedance-based label-free technology as a tool for pharmacology and toxicology investigations," Biosensors, vol. 3, no. 1, 2013, pp. 132–156.

[49] S. Ramasamy, D. Bennet, and S. Kim, "Drug and bioactive molecule screening based on a bioelectrical impedance cell culture platform," International Journal of Nanomedicine, vol. 9, no. 1, 2014, pp. 5789–5809.

[50] N. E. Beltran and E. Sacristan, "Gastrointestinal ischemia monitoring through impedance spectroscopy as a tool for the management of the critically ill," Experimental biology and medicine, vol. 240, no. 7, 2015, pp. 835–845.

[51] D. Sifrim and K. Blondeau, "Technology insight: the role of impedance testing for esophageal disorders," Nature Clinical Practice Gastroenterology & Hepatology, vol. 3, no. 4, 2006, pp. 210–219.

[52] F. Ates et al., "Mucosal impedance discriminates gerd from non-gerd conditions," Gastroenterology, vol. 148, no. 2, 2015, pp. 334–343.

[53] T. Matsumura et al., "Evaluation of esophageal mucosal integrity in patients with gastroesophageal reflux disease," Digestion, vol. 97, no. 1, 2018, pp. 31–37.

[54] Z. Maherally et al., "Real-time acquisition of transendothelial electrical resistance in an all-human, in vitro, 3-dimensional, blood–brain barrier model exemplifies tight-junction integrity," The FASEB Journal, vol. 32, no. 1, 2018, pp. 168–182.

[55] P. Tsiotras, E. Bakolas, and Y. Zhao, "Initial guess generation for aircraft landing trajectory optimization," 08 2011.

[56] Q. Yao and H. Tong, "Quantifying the influence of initial values on non-linear prediction," Journal of the Royal Statistical Society: Series B (Methodological), vol. 56, no. 4, 1994, pp. 701–725.

[57] K. Kaufmann et al., "Crystal symmetry determination in electron diffraction using machine learning," Science, vol. 367, 2020, pp. 564–568.

[58] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," IEEE Access, vol. 8, 2020, pp. 63 514–63 537.