International Journal on Advances in Intelligent Systems

2024 vol. 17 nr. 1&2

The *International Journal on Advances in Intelligent Systems* is Published by IARIA.
ISSN: 1942-2679
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Intelligent Systems
Volume 17, Number 1 & 2, 2024

International Journal on Advances in Intelligent Systems
Volume 17, Numbers 1 & 2, 2024

CONTENTS

Anticipating Future Healthy Food Services in Smart Cities Through Systematic Literature Reviews

Anaisa Lucena
I3N & Physics Department
University of Aveiro, 3810-193
Aveiro, Portugal
anaisalucena@ua.pt

Ana Sofia Freitas
Wish and Cook
Aveiro, Portugal
anasofiafreitas@wishandcook.com

Antonio Luis Ferreira
I3N & Physics Depart.
University of Aveiro, 3810-193
Aveiro, Portugal
alf@ua.pt

Fernao Vistulo de Abreu
I3N & Physics Department
& Wish and Cook
Aveiro, Portugal
fva@wishandcook.com

***Abstract*—The onset of the fifth industrial revolution promises to reshape industry-consumer dynamics, particularly impacting the food services sector. This study aims to identify emerging research trends in smart cities, food services, personalized nutrition, and artificial intelligence through a systematic literature review. Natural language processing techniques were employed to extract meaningful terms from titles and abstracts, followed by a modified Non-negative Matrix Factorization (NMF) algorithm to categorize the research into distinct topics. The findings reveal a growing interest in the interdisciplinary field, highlighting smart cities' potential to create seamless connections between consumers, farms, restaurants, and health systems. This connectivity is poised to reduce waste and enable personalized recommendations tailored to individual preferences and health needs. Key factors shaping the future of smart cities include urban mobility, food transportation, dietary monitoring, and waste reduction strategies. Integrating technologies such as blockchain, the internet of things, artificial intelligence, and machine learning is set to enhance efficiency and connectivity, fostering a more sustainable and consumer-centric food ecosystem.**

Keywords-personalized nutrition; food services; smart cities; artificial intelligence; NMF.

I. Introduction

The ongoing Fourth Industrial Revolution is marked by widespread digitalization and automation across industries. It leverages cutting-edge technologies like artificial intelligence (AI), the Internet of Things (IoT), and Big Data to drive innovation and reshape the landscape of various sectors [1], [2].

An important sector that gained from the Industry 4.0 is the food industry and the kitchenware. As other industries, this was associated with an overall increase in innovation, associated with a more automatized, digitalized and data-driven approach of the available food services. Examples range from precision agriculture, efficient supply chain management, personalized nutrition apps and kitchenware with wifi connection to recipe databases in the cloud [3], [4].

Anticipating the advent of the Fifth Industrial Revolution, an unprecedented level of automation, harmonious human–machine collaborations and sustainable practices is to be expected. A cornerstone and defining feature of the future and Industry 5.0 will be smart cities. Due to its primarily urban-based operations and the concepts and characteristics of smart cities, this opens the possibility that the food services sector can be highly impacted. It is likely that customers and consumers will interact in a more engaging and connected way through the offering of more nutritious and efficient meal choices. This can still be combined with sustainability practices, such as optimizing food production and reduce food waste [5], [6].

To gain insight on the multiple dimensions on how this sector can evolve, a search on papers relating smart cities and food services was conducted. However, since this is still an emerging field, only a limited number of papers was available. For that reason, two complementary literature reviews were used to gain more profound results.

This article is structured as follows: first, in Section II, a description of the quantitative literature review methodology is introduced. Then, in Section III, the topics that emerged from this analysis are presented and, finally, in Section IV, the implications of the findings are discussed. These are summarized in the concluding section of the article, Section V.

II. Methodology

This chapter outlines the systematic approach taken for the literature review and topic modeling process. It begins with information gathering, detailing the criteria and sources used for collecting relevant literature. This is followed by data processing, which involves preparing the text data through various preprocessing techniques. The third section discusses the application and modifications of the NMF algorithm. Finally, the chapter concludes with the method used to determine the optimal number of topics.

A. Information Gathering

In order to extract relevant articles, two search engines *PubMed* [7] and *Web of Science* [8] were used. Search keywords were used to unveil research addressing human concerns that could gain from new technologies. In *PubMed*, the search was oriented to research in the health science community, while in *Web of Science* a broaden search was used.

To find the most relevant articles in *PubMed*, MeSH terms (i.e., vocabulary terms used to index and categorize articles in the MEDLINE database) were used [9]. The search looked for articles with the "*nutritional physiological phenomena*" MeSH term, together with requiring the presence of the expression "artificial intelligence" OR "machine learning". This search aimed to unveil research that could use new technologies to improve human healthy eating.

In the *Web of Science* engine, the condition used for the search was the presence of keywords *("smart cities" OR "smart city") AND ("nutrition" OR "food")*. These search keywords were selected so that a statistically significant number of articles could be found linking new technologies (artificial intelligence or machine learning in one case, and smart cities in the other) to food and nutrition.

With these search criteria, 630 articles were found in *PubMed* and 464 in *Web of Science*. For every article, the title, the complete abstract and the year of publication were recorded for data processing analysis, as described next.

B. Data Processing

After collecting the information on the selected articles, the title and the abstract of every article were processed using Natural Language Processing (NLP) techniques. This involved converting every letter to lowercase, removing non letter characters, tokenizing the text (in order to analyze words individually), removing stop-words (the most common English words, such as '*the*', '*or*', '*what*', etc.) and stemming (reducing inflected words to their word stem or root form; for example, the words '*simulation*', '*simulator*' and '*simulate*' can all be reduced to their stem '*simul*').

Another important operation consists in removing words that appear frequently in scientific texts but are not specific to this particular field of interest, such as '*prove*', '*demonstrate*', '*experiment*', etc. In order to do this, 647 articles were collected using *Semantic Scholar* [10] after searching with the keyword "*biotechnology*". The former NLP techniques were also applied on these articles, and the frequency of occurrence for all the words in the two collections of articles (those related to this analysis, and those related to the biotechnology search) was calculated.

Then it was possible to calculate the frequency ratio, FR, as seen in Eq. (1).

$$FR(word) = \frac{Counts_S(word)/Tot_S}{Counts_B(word)/Tot_B} \quad (1)$$

Here, $Counts_X(word)$ represents how many times $word$ appeared in the abstract and title of every article in the nutrition and artificial intelligence and the smart cities and food/nutrition searches (X=S) or on the articles in the biotechnology search (X=B). Tot_X represents the total number of words in the abstract and title of every article on each search. The ratio $FR(word)$ was calculated for every word found on both searches. In the case of the words that did not show up in the biotechnology search, $Counts_B(word)$ was considered to be 1.

With this approach, the 100 words with the highest $FR(word)$ ratio were selected, which allows for pinpointing specific terms that are likely to convey essential information related to the topics of interest.

C. Modified Non-negative Matrix Factorization Algorithm

In order to identify the different thematic areas for each search, a NMF algorithm was used. NMF algorithms find an approximate factorization of an input matrix V into two other non-negative matrices, W and H, such that $V \approx WH$ [11].

In this case, the input matrix V is defined as a binary matrix, where each row is associated with an article, and each column corresponds to one of the top 100 words. The values in the matrix are either 0 or 1, indicating the absence or presence of a word in an article. Being N the number of articles, V is a $N \times 100$ matrix.

The NMF algorithm helps compress the information in V by decomposing it into two smaller matrices, W and H, with dimensions $N \times k$ and $k \times 100$ respectively. Typically, k is no larger than 20, so $N \times k + k \times 100$ is significantly smaller than $N \times 100$.

According to this decomposition, each row in V (representing the words in an article) is expressed as a linear combination of the rows in H. Since the number of rows in H (k) is much smaller than the number of articles (N), this breakdown suggests that articles can be grouped into topics. These topics are characterized by the specific set of words listed in each row of H.

To apply the NMF algorithm it is then necessary to choose the number of topics k. Then the NMF algorithm produces, after an optimization algorithm, the matrices W and H, as represented in Figure 1.

While the entries in W reveal the degree of association between articles and topics, the entries in H indicate the contribution of the words to each topic. Therefore, examining the H matrix reveals connections between words associated with each topic, facilitating the definition of diverse areas where artificial intelligence and smart cities influence nutrition. Meanwhile, from matrix W it possible to understand which articles pertain to specific topics, multiple topics, or none at all.

Given that the number of articles used in these searches is relatively small (in statistical terms), some topics (lines in the H matrix) have one entry (word) dominating all others, as shown in the example in Figure 2. There, the word '*urban*' has a significantly higher value than the others. This would mean that the presence of the word '*urban*' in an article would be enough to associate it to a topic, which is a typical example of over-fitting due to the small number of articles used.

To overcome this problem, an iterative reformulation of the NMF algorithm was developed such that whenever the weight associated with a word in the matrix H represents more than 20% of all the weights in the same topic, the corresponding entry for that word in the input matrix V is decreased by

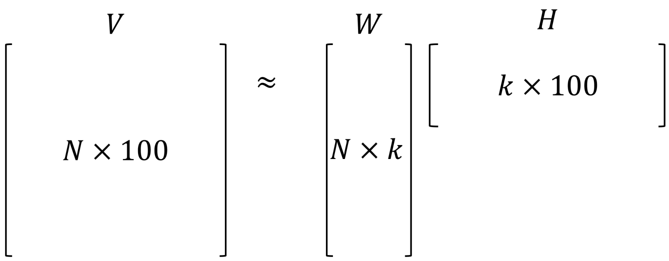

Figure 1: Visual representation of NMF (non-negative matrix factorization) algorithm, where N is the number of articles, k is the number of topics, and 100 is the number of words.

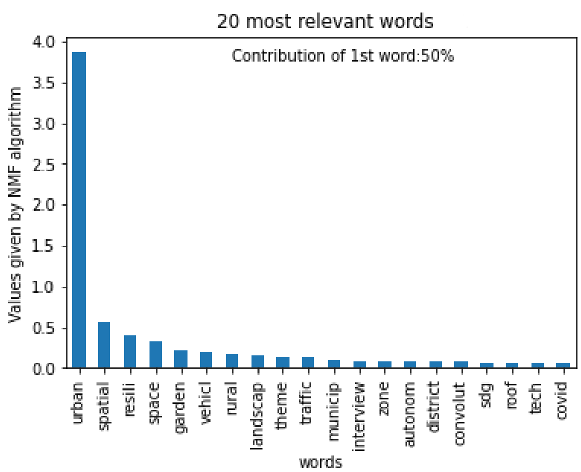

Figure 2: Bar chart representing the 20 words with the highest values for a row of the H matrix of the NMF algorithm before being modified (from the articles in Web of Science searches).

Algorithm 1 Modified NMF

$N \leftarrow$ number of articles
$V \leftarrow (N \times 100)$ binary matrix where $V_{ij} = 1$ if article i contains word j
$W, H \leftarrow$ NMF decomposition of V
$l \leftarrow$ list of the words where $max(H_i)/\sum H_i > 0.2$, for each topic i
while l is not empty **do**
  for every article a **do**
    for every word w in l **do**
      $V_{aw} \leftarrow$ decrease 1%
    end for
  end for
  $W, H \leftarrow$ NMF decomposition of V
  $l \leftarrow$ list of the words where $max(H_i)/\sum H_i > 0.2$, for each topic i
end while

Figure 3: Algorithm of the modified NMF, used to correct overfitting.

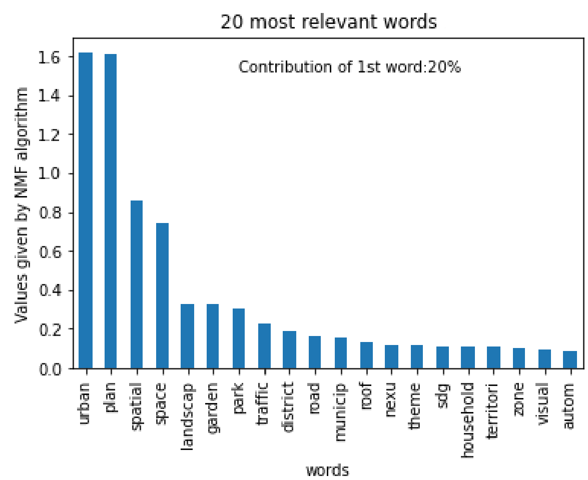

Figure 4: Bar chart representing the 20 words with the highest values for a row of the H matrix of the modified NMF algorithm (from the articles in Web of Science searches).

1%, as detailed in Figure 3. This value was chosen based on preliminary experiments, which indicated that a 1% decrease does not significantly alter the results, but performing this process iteratively is computationally efficient and does not require extensive processing time.

The application of this modified version of the NMF algorithm changed the values of the highest weights in matrix H. A typical example is shown in Figure 4. They show that the decomposition obtained with the modified NMF algorithm requires a larger number of words to define a topic. Furthermore, the semantic consistency of the several words defining each topic shows that this procedure produces much more insightful results.

D. Number of Topics

In order to choose k, the number of topics that each search contained, the squared Frobenius norm was calculated, according to Eq. (2) [12].

$$P_k = ||V - WH||_F^2 = \sum_{ij} |v_{ij} - (WH)_{ij}|^2 \qquad (2)$$

In this equation, W and H have k columns and lines, respectively. The squared Frobenius distance was calculated using the decompositions found with different values of k and using the modified NMF (Figure 3). The results are shown for the two article collections in Figures 5 and 6.

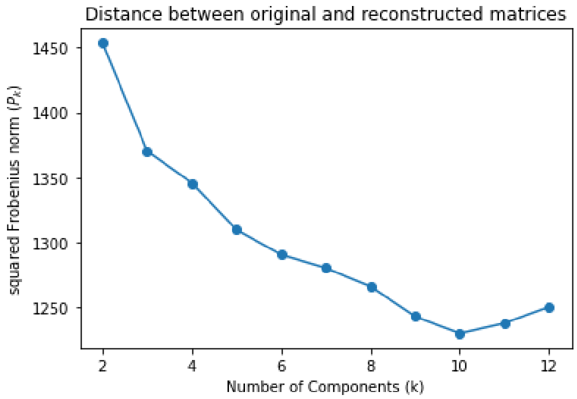

Figure 5: WoS squared Frobenius norm.

From Figure 5, it is possible to observe a minimum for $k = 10$. As such, it was considered that the Web of Science (WoS) research could be grouped into 10 different topics. Likewise, according to Figure 6, $k = 15$ topics were considered optimal for the PubMed search.

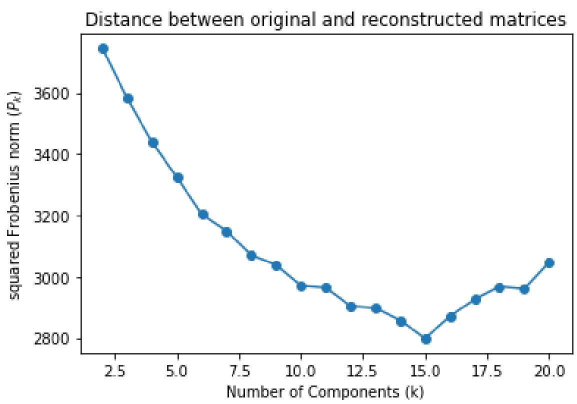

Figure 6: Pubmed squared Frobenius norm.

Given that $k = 15$ represents a significant number of topics, an exploration was conducted to assess whether the increase in the number of topics relatively to the WoS search, was indeed real. This examination aimed to determine if new topics were not generated spuriously. To facilitate this comparison, a distinct nomenclature was adopted for the two sets of topics: numerical labels from 1 to 10 were used for the decomposition with $k = 10$, while capital letters from A to O were employed for $k = 15$. This approach facilitates the clear identification of any correspondence between the two groups of topics.

III. Results

The first result is shown in Table I, where the 15 most common words (i.e., the words with the highest FR) on both searches are listed. These words, although related, are not coincident, showing that the two searches are complementary.

TABLE I: MOST COMMON STEM WORDS FOUND FOR THE TWO SEARCHES, SORTED FROM MOST TO LEAST COMMON

Search	15 Most Common Words
Web of Science: (smart cities or *smart city*) and (*food or nutrition*)	urban, iot, propos, servic, sensor, internet, infrastructur, plan, intellig, spatial, blockchain, mobil, scenario, machin, fresh
PubMed: nutritional psysiological phenomena and (*artificial intelligence* or *machine learning*)	diet, machin, dietari, intak, patient, algorithm, weight, intervent, accuraci, obes, diabet, gut, microbiota, glucos, score

Another interesting result concerns the number of articles published along the time for each search as shown in Fig.7. These results show a clear growth in activity in recent years, on both cases, demonstrating a growing significance of these topics within the scientific community.

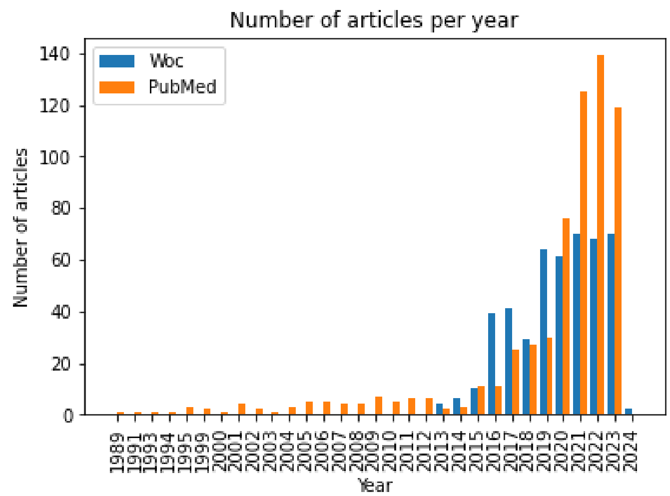

Figure 7: Temporal distribution of the number of articles collected with the two searches.

Moreover, after selecting the most frequent and meaningful 100 words for the analysis, some articles still lacked any of these words in their title or abstract. These articles had to be discarded, as they lacked information for this study. This happened for 23 (out of 464) articles in the WoS search and for 8 (out of 630) in the PubMed search.

Upon implementing the modified NMF algorithm, the rows of matrix H specify the words distinctive to each of the k topics, while matrix W holds the coefficients representing how each article breaks down into these various topics. This information is vital for connecting articles with topics and, as such, a straightforward strategy involves linking each article with the topic exhibiting the highest coefficient in matrix W.

However, choosing the maximum value dismisses the instances where the second, or even third, highest values of the W decomposition are very close to the maximum, meaning that that article has components of more than one topic. This would happen if an article would be at the interface of different topics. In order to define which articles could fall into this situation, it was established that an article i belongs to topic j if it satisfies the condition in Eq. (3):

$$\frac{max(W_i) - W_{i,j}}{max(W_i)} \leq 0.05 \quad (3)$$

Following this criteria, 29 articles were classified as belonging to 2 distinct topics and 1 article was placed into 3 topics in the WoS search. The distribution of the articles in the several topics is shown in Table II. A similar procedure was carried on for the articles in the PubMed search. In this case, the number of articles per topic is shown in Table III for the 622 articles included in the analysis. In this case, 28 articles were associated to 2 different topics simultaneously. These results show that all topics have a significant number of articles, even though some topics have more articles than others.

TABLE II: WOS: NUMBER OF ARTICLES PER TOPIC

	Topic									
	1	2	3	4	5	6	7	8	9	10
Calculation Method	**Number of Articles**									
Max Value	98	47	46	40	34	18	44	50	35	29
Equation 3	103	50	47	42	37	22	45	51	37	38

TABLE III: PUBMED: NUMBER OF ARTICLES PER TOPIC

	Topic														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Calculation Method	**Number of Articles**														
Max Value	114	83	29	28	33	34	48	23	38	23	47	45	30	19	28
Equation 3	114	91	32	29	33	36	51	25	40	24	50	45	31	20	29

In order to confirm that this association procedure indeed makes sense, an individual analysis of articles, as well as of the most important words characterizing each topic (see Figure 8, related to the WoS search, and Figure 10, related to the PubMed search) was undertaken. For example, in the WoS seach, [13] was classified in topics 1, 4 and 6, while based on the maximum value it would only be associated with topic 1. This research explores a method to establish a trusted, decentralized system for food distribution within the framework of a Smart City that integrates IoT, Blockchain technology, and city LoRa network, which is within the frame of topics 1, 4 and 6. In this case, looking at the most important words of topics 1 (servic, infract, rural), 4 (air, store, fresh) and 6 (blockchain, transpar, propos, traceabl) seen in Figure 8, it is possible to understand that this article belongs to topic 1 since it relates to food distribution services, to topic 4 since it is necessary to store the food in order to distribute it and to topic 6 due to the use of blockchain technology.

Another example is of the work of Ragab, Osama, and Ramzy [14], which uses AI and ML to simulate the impact of industries (among which the food industry) in smart cities, being within the framework of both topics 1 and 9. Furthermore, [15] explores the zero-emission strategy and implementation plans of Tokyo city, and its efforts to progress towards achieving net zero emissions by 2050, fitting within topics 1 and 7. Finally, [16] simulated different vegetation cover scenarios, in order to quantify their effects on the microclimate of a mixed-residential-industrial area, relating to both topics 5 and 7.

Similarly, an examination of selected articles belonging to more than one topic, alongside the key terms was made for articles in the Pubmed search. For instance, in [17], which investigates personalized nutrition employing artificial intelligence for patients with irritable bowel syndrome, topics B and I are pertinent. Topic B addresses patient diagnosis and clinical studies, while topic I concerns the utilization of artificial intelligence in such contexts. Similarly, [18] explored the interplay between oxidative damage, the redox status, and metabolic biomarkers during long-term fasting, revealing decreases in blood glucose, insulin, glycated hemoglobin, total cholesterol, low-density lipoprotein, and triglycerides, alongside an increase in the total cholesterol/high-density lipoprotein ratio. These findings align with topics J and O. Furthermore, [19] investigated the interactions between serum free fatty acids and fecal microbiota in obesity, employing a machine learning algorithm to classify subjects by BMI. This study intersects with topics A and F.

Another insightful representation of the distribution of articles per topics is provided in Figures 8, 9 and 10. In these figures, the temporal evolution of the number of articles in each topic is displayed. Furthermore, the most significant words for each topic are also presented, in decreasing order of importance, as mentioned before. These have been obtained from the entries of the H matrix with the highest values. The words considered are those with a weight, in the H matrix, of at least 40% of the word with the highest value for said topic.

To enhance the visualization clarity, a heatmap approach was utilized. For each topic, the year with the highest number of articles appears as deep red, while the year with the lowest count appears as deep blue, with the intensity of color corresponding to the volume of articles. Figures 9 and 10 differ on the k value used in the modified NMF ($k = 10$ and $k = 15$, respectively).

It is noteworthy to observe a global rise in publication activity across nearly all topics. However, due to the relatively small number of articles per topic, drawing further conclusions regarding trends is challenging. Moreover, it is important to acknowledge that although the latest results are recorded up to March 2024, ongoing updates to databases may still be occurring for articles published in 2023.

The heatmaps depicted in Figures 9 and 10 reveal that, at $k = 10$, certain topics appear to encompass articles spanning multiple subject areas. Conversely, with an increase to $k = 15$, new topics emerge (e.g., topics D, E, H, and J), stemming either from the splitting of previous topics or the aggregation of articles from multiple topics.

To support this hypothesis, the fraction P($i|j$), representing the probability of an article coming from topic i ($i \in 1, ..., 10$) for $k = 10$, knowing that was classified under topic j ($j \in A, ..., O$) for $k = 15$, was calculated for all topics. Large P($i|j$) values mean that topic j originated mainly from topic

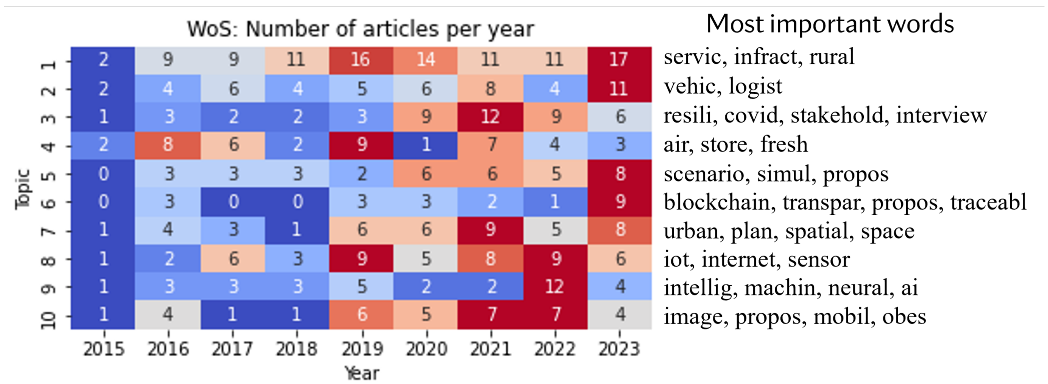

Figure 8: WoS: Temporal evolution of the articles in each topic, as well as the most relevant words.

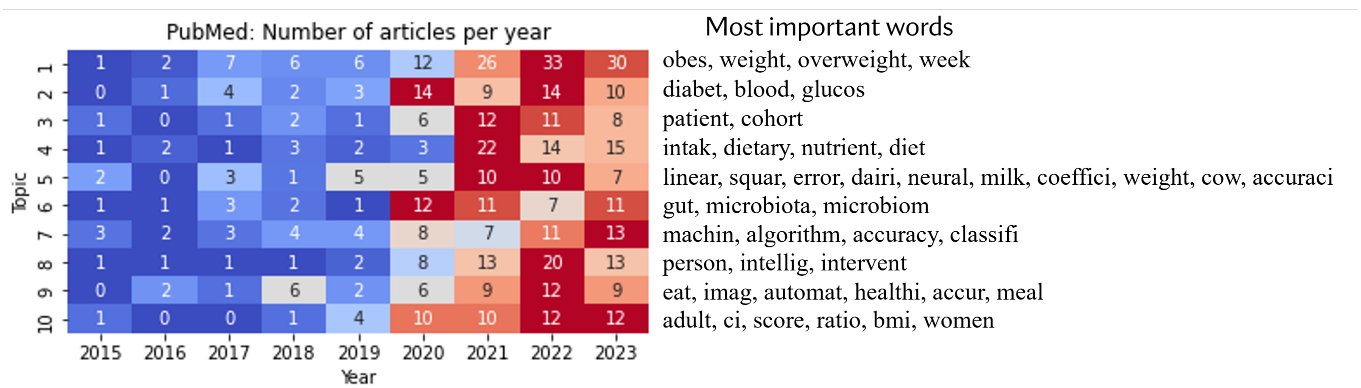

Figure 9: PubMed: Temporal evolution of the articles in each topic (out of 10), as well as the most relevant words.

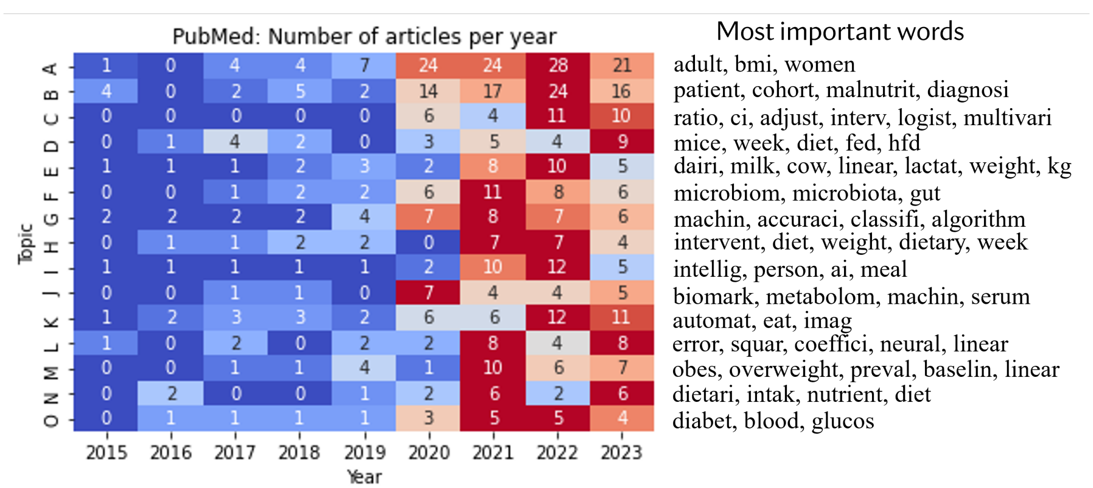

Figure 10: PubMed: Temporal evolution of the articles in each topic (out of 15), as well as the most relevant words.

i. The following results were obtained:

- A originates from $10(31\%)$ and $1(29\%)$
- B originates from $3(45\%)$
- C originates from $10(41\%)$
- D originates from $1(64\%)$
- E originates from $5(66\%)$
- F originates from $6(76\%)$
- G originates from $7(77\%)$
- H originates from $1(47\%)$ and $8(39\%)$
- I originates from $8(78\%)$
- J originates from $6(34\%)$ and $4(26\%)$
- K originates from $9(65\%)$
- L originates from $5(57\%)$
- M originates from $1(86\%)$
- N originates from $4(84\%)$
- O originates from $2(100\%)$

Based on these findings, it is clear that certain topics identified in the analysis with $k = 15$ were already delineated as topics within the $k = 10$ framework. This is particularly notable in instances such as direct transcriptions of O or N. Conversely, some topics identified with $k = 10$ encompassed two distinct themes when scrutinized with $k = 15$. Notably, topics L and E emerged as distinct entities, yet both were belonged to topic 5. This observation confirms that $k = 15$ is a better parameter choice for the number of topics.

Moreover, topics A, D, H, and M originated from a shared theme present in the $k = 10$ analysis, topic 1, which revolves around diet and weight management. However, these topics offer varied perspectives on this overarching theme. Notably, certain nuances within these topics become more pronounced with $k = 15$, as exemplified by the focus on mice experimental models within topic D.

IV. Discussion

In this section, the aim is to examine the ongoing research in the areas regarding the future of the food services sector, including potential benefits to human health, given the emergence of smart cities. As such, some of the articles with less overlap, i.e., with the highest ratio $\max(W_i)/\max(W_i \notin \max(W_i))$, will be further explored. In Tables IV and V, an analysis of the topics emerging from the two searches is outlined. Both searches were coincident on some topics - for example, on applications of artificial intelligence, topic 9 (WoS) and topics G and I (PubMed) - while others provide additional insights. This demonstrates that there is consistency between the methods, which are also complementary.

A thorough analysis of Tables IV and V provides important insights on how food services may evolve in smart cities. For instance, the topic 2 of WoS emphasizes the importance of transporting food to the customers efficiently, reducing transport emissions. Also, topic 5 of WoS shows the importance of planning and mitigating waste in the most sustainable way.

Topic 10 of WoS and topics K and N of PubMed highlight problems related to self-report food intake. For example, children and adolescents are unable to self-report food intake without caregivers [83] and even trained individuals have difficulties in estimating food portions accurately [84]. Solutions with sensors (body sensors, cameras, etc) and artificial intelligence could contribute to monitor and help control food intake, helping to develop better food prescriptions, better disease prediction, and improved prevention strategies.

Also, technologies such as blockchain (WoS topic 6) and internet of things (WoS topic 8), as well as machine learning and artificial intelligence (WoS topic 9 and PubMed topics G and I), will be used for a multitude of applications, from preserving privacy, to improving resource efficiency, image recognition, monitoring and personalizing nutrition.

Moreover, even though just topic E specifically focus on farms, in this case dairy farms, a lot of articles (see WoS topics 6, 8 and 9 in Table IV and topic L in Table V) are related to agriculture, emphasizing the importance of developing a direct connection between farms and consumers, to reduce food insecurity, which is especially relevant for fresh food options (a concern also highlighted in WoS topic 7) and to promote sustainable city expansion and social well-being. Therefore, food services creating a direct link between farms and the consumers' table could be perceived positively.

Furthermore, several diseases or health conditions could benefit from interaction with future food services. For instance, accurate food intake monitoring could help control obesity and diabetes (topic 10 of WoS and topics H, K, M, N and O of PubMed). Also, knowledge awareness for the interplay between nutrition and microbiota (topic F) makes it likely that dietary choices will become increasingly recognized for their pivotal role in intestinal modulation. This understanding underscores the importance of actively selecting ingredients that promote gut health, thereby contributing to the prevention of intestinal disorders and ultimately impacting immune function and overall well-being.

Additionally, regarding the PubMed search, while some topics followed a clear trend, others were more miscellaneous, namely topics C, G, I and L. For example, topic C just consists of articles where the abstract is mostly numerical and presents a lot of results, hence the ratio and confidence interval words. Topics G and I, as mentioned before, are related to machine learning and artificial intelligence in a broad sense, even though most articles have an AI/ML component. Lastly, topic L is related to prediction algorithms across different domains, with all articles sharing a common thread of statistical analysis in their abstracts. Metrics like mean square error and correlation coefficients are commonly reported, highlighting the emphasis on quantitative analysis rather than thematic content within this topic, akin to Topic C.

Furthermore, smart cities and farms can interact regularly to manage storage and delivery efficiently. In the future, food services may interact with consumers, suggesting meals based on information received from them, who are also engaging with health services. Consumers can share health data through tools and sensors, allowing health services to online monitor and recommend nutritional plans. The consumer/patient will have closer access to the healthcare system, thereby

TABLE IV: WEB OF SCIENCE SEARCH: CONTENT TABLE FOR THE TOPICS OBTAINED

Topic	Type of Solution	Example of Problem Addressed
1	Services	Studying the contribution of ecosystem services to farmers' livelihoods [20]
	Sustainable infrastructures	Reviewing the role of smart city innovation for sustainable infrastructure in Bahrain [21]
		Managing thermal energy in a medium aquaponics system, a sustainable alternative food production infrastructure, for biological sustainability, especially during cold weather [22]
2	Urban mobility and transportation of food	Energy efficiency analysis using unmanned aerial vehicle systems and electric scooters in the transport of takeaway food, which is a cheap, fast, and green transportation solution that fits into the zero-emission transport policy of people and goods [23]
		Comprehensive review and analysis of the latest trends in last-mile delivery solutions from both industry and academic perspective [24]
		Mechanism that exploits automated electric vehicles in future smart cities and regions to provide both people transport and fresh food distribution that minimizes empty miles of the vehicles (and thus enhances transport efficiency) while meeting the constraints on passenger transit time and food freshness [25]
3	Urban resilience	Urban resilience capacity and its relations with the economic, social and environmental well-being in smart cities in the state of São Paulo, particularly after the 2008 financial crisis [26]
		Examining government employees' experience and expectation of socioeconomic hardships during the COVID-19 pandemic [27]
4	Storage quality and conditions	Evaluating the physiological and chemical traits of three red and three green baby leaf lettuce during postharvest cold storage [28]
		Characterizing and sensory analyzing the volatile profile of two hybrids of "Radicchio di Chioggia" stored in air or passive modified atmosphere during 12 days of cold storage [29].
		Examining the phenolic profiles and changes in postharvest quality of radicchio leaves when freshly cut and when stored in unsealed bags or in passive modified atmosphere [30].
5	Planning and mitigating waste	Studying whether the installation of food waste disposers in private homes or separate collection and transport of organic waste to biogas plants is a more viable environmental and economic solution, being the latter the most sustainable [31]
	Simulations	Studying the effectiveness behavior change techniques by using computational models that describe human [32]
6	Blockchain and preserving privacy	Blockchain-based innovative framework for privacy-preserving and secure IoT data sharing in a smart city environment [33]
		Novel BIoT-based layered framework using EOSIO for effective food traceability in smart cities [34]
		Literature review about the application of blockchain in the agricultural sector, focusing on food traceability issues [35]
7	Urban planning	Effects of uncontrolled urban development and urban compactness on the quality of life [36]
	Willingness to buy green food products	Studying smart city millennials' willingness to pay a premium toward toxic-free food products [37]
		Identifying the predictors that influence young, educated customers' intention to purchase green food products in India by utilizing the extended theory of planned behavior, which includes elements such as environmental concern, perceived customer effectiveness, willingness to pay a premium, and product availability [38]
8	Implementation of internet of things in smart cities	Internet of Things technology for efficient farming processes, specifically collecting and processing soil nutrients and weather data of crop avocados in an orchard [39]
		Reviewing current literature related to IoT and big data-based food waste management models, algorithms, and technologies with the aim of improving resource efficiency [40]
9	Use of artificial intelligence in food and smart cities	Rice quality evaluation system based on computer vision and machine learning [41]
		Artificial intelligence technology to screen active components of natural products [42]
10	Food image recognition	Integrated image processing framework that extracts numerical nutrition information from the Thai GDA label images [43]
		Computer-aided technical solutions to enhance and improve the accuracy of current measurements of dietary intake [44]
	Medical imaging	Using magnetic resonance imaging processing for fat depot assessments and data analysis [45]
		Using magnetic resonance imaging to examine the differences between metabolically healthy and unhealthy overweight/obesity phenotypes on specific abdominal fat depots and explore whether cardiorespiratory fitness plays a major role in the differences between metabolic phenotypes among overweight/obese children [46]

TABLE V: PUBMED SEARCH: CONTENT TABLE FOR THE TOPICS OBTAINED

Topic	Type of Solution	Example of Problem Addressed
A	Effects of high BMI	Study on plasma folate in pregnant and lactating Chinese women, found that one of the factors contributing to lower folate concentration is higher BMI [47]
	Diets and optimal intakes of food or water in adults.	Prediction of optimal water intake in adults using machine learning [48]
		Evaluating the relationship between the diet quality index of young adults and cardiometabolic risk factors [49]
B	Identifying and/or studying malnutrition in patients	Phase angle as a tool for nutritional monitoring and management in patients with Chron's Disease, usually accompanied by malnutrition [50]
		Personalized nutrition for treating malnutrition in cancer patients using machine learning tools [51]
		Machine learning (multivariate regression and random forest) to predict in-hospital complications in elderly patients diagnosed with malnutrition [52]
C	Numerical results	Predicting hospital length of stay [53]
D	Experiences on mice and the effects of diet on mice	Studying the relationship between the components of cellular magnesium (Mg) home- ostasis and energy metabolism in cardiomyocytes by analysing two groups of mice: one fed a diet with normal Mg content and another fed a diet with low Mg content [54]
		Showing feeding is fragmented and divergent motivations for food consumption or environment exploration compete throughout the feeding process by delineating the behavioral repertoire of mice by developing a machine-learning-assisted behavior tracking system [55]
E	Dairy farms and the feeding of the cows	Using sensors that measure the behavior of cows intended to be included in pasture management [56]
		Real time continuous decision making using big data in dairy farms by using sensors and robotic systems that can collect, integrate, manage and analyze on and off farm data in real time [57]
		MIR spectroscopy to authenticate the milk source at both farm and processor levels for grass fed and non-grass fed milks [58]
F	Gut microbiota	Clarifying the effects of nutrients consumed on the entire gut microbiome by studying gut microbiota differences in identical twins [59]
		Comparing gut microbiota of Cameroonians to people from Philadelphia, finding that the Cameroon diet has more gut parasites, which increases gut microbial diversity [60]
		Finding optimal maturation trajectory of the gut ecosystem through machine learning models, to study the effects of diet on gut microbiome [61]
G	Machine learning models related to food and nutrition	Creating a network-based machine learning platform to identify putative food based cancer beating molecules and a "food map" with the anti-cancer potential of each ingredient [62]
		Using natural language processing and machine learning approaches to categorize food and predict nutrition quality [63]
H	Weight loss and dietary interventions	Artificial intelligence based virtual health assistant [64]
		Machine learning (WEKA decision trees) to predict dietary lapses during weight loss [65]
		Studying the efficiency of a diet-related mobile application based on artificial intelligence by studying the nutritional status of children post-cardiac surgery, giving to part of the study group this application and the other part just a pamphlet. This application was proved efficient [66]
I	Artificial intelligence models	A systematic literature review of precision nutrition [67]
		An educational approach based on intelligent systems and its application in nursing education [68]
J	Metabolomic biomarkers and patterns	Prognostic value of the human milk metabolome and exposome in children with the risk of neurodevelopmental delay (NDD) using a predictive classifier [69]
		Comparing metabolite profiles of habitual diet in serum and urine, using metabolomics for identifying objective dietary biomarkers, and creating serum and urine multiple-metabolite models to predict food intake [70]
K	Diet monitoring	Current technological approaches to monitoring energy intake [71]
		Tool that uses new digital photographing technology to reduce measurement error associated with a food record and the recording burden for respondents by photographing the food before eating [72]
		Reviewing two of the most relevant and recent researches on automatic diet monitoring: one based on image analysis and the other on wearable sensors [73]
L	Prediction algorithms	Machine-learning techniques to predict the daily nutritional needs of pregnant pigs using solely sensor data, according to various configurations of digital farms [74]
		A novel method for predicting the absorption of the drugs from their structures, offering understanding into the structural characteristics linked to drug absorption [75]
M	Predicting Obesity	Using information about the diet and physical activity of the residents of a neighborhood to improve the estimate of neighborhood-level obesity prevalence and help identify the neighborhoods that are more likely to suffer from obesity [76]
		Random Forest and Gradient Boosting Machine models to predict the body mass index of children and, consequently, prevent children obesity [77]
	Identifying diseases in already obese individuals	Machine learning to generate predictive models for knee osteoarthritis incidence in overweight and obese women [78]
N	Nutrient aspect of diet monitoring	Machine learning to explore how reducing the number of questions affects the predicted nutrient values and diet quality score [79]
		Validating the accuracy of an internet-based app against the Nutrition Data System for Research (NDSR), assessing the dietary intake of essential macro-and micro-nutrients for precision nutrition [80]
O	Blood glucose regulation and diabetes management	Artificial intelligence to predict blood glucose level changes resulting from regimen disturbances and recommend regimen changes for compensation for type 1 diabetes [81]
		Comparing machine learning-based prediction models (i.e., Glmnet, RF, XGBoost, LightGBM) to commonly used regression models for prediction of undiagnosed type 2 diabetes mellitus [82]

reducing barriers for health professionals, which can then provide tailored nutritional information, offer practical tools for implementing long-term changes, and monitor progress effectively.

Connecting agriculture and food services in smart cities can create a shorter supply chain, enabling continuous feedback from food services that could influence food supply. This is particularly important, as typical restaurants waste up to 10 percent of their purchased food before it reaches consumers [85]. Smart food services could reduce this waste by optimizing stock and accessing fresh ingredients sent daily from farms, using an automatic stock management system that connects restaurant storage areas with farms, determining when and how much of each ingredient is needed.

Additionally, data on food/nutrient intake is important to inform nutritional policies and monitor individual programs, that can modulate food services. There are several methods to access individual dietary intake and the most frequently used is 24-h recall method. However, there are some difficulties to validate the data, mainly there are no clear evidence that children and adolescents are capable to self-report food intake without caregivers help [83]. There are additional problems that includes difficulties to memorize the 24h food intake and even trained individuals have difficulties to estimate the food portion accurately [84]. So, artificial intelligence will help to quantify food intake with more accuracy. Also, with the development of a customized kitchen capable of precisely controlling food portions, supported by a comprehensive database, which can automatically generate nutritional information including calories, macro-nutrients, and micronutrients, accurate monitoring of food intake becomes feasible.

People have the need to connect, and the smart city is a place were everything is connected. Ultimately, connectivity is the key to optimize resources and reduce water and food waste. Through a system of on-demand production, ingredients are not wasted as production, storage, and cooking align with the required quantity. Consequently, this is a scenario where food production, encompassing agriculture, seamlessly integrates with restaurants and households via an automated storage management system. This tool anticipates the need for replenishment and regulates production quantities in real-time.

Finally, automation and connectivity are likely to change human habits in future smart cities. A number of companies are pioneering the development of fully automated food services (see Table VI). These will decrease meal preparation costs and, consequently, it is likely that consumers will adopt these services on a daily basis. It is important to note that the companies listed in Table VI were identified through a separate investigation focused on current market leaders in automated food services, and are not directly related to the systematic literature review. In any case, many of the technologies discussed here will be easily put in play.

TABLE VI: LIST OF AUTOMATED FOOD SERVICE COMPANIES

Name	Location	Cuisine type	Autonomy
Sweetgreen (formerly Spyce)	USA	Bowls	fully
Kitchen Robotics	Florida, USA	Several	partial
Circus (formerly Aitme)	Berlin, Germany	Several	fully
Chef Jasper YPC Tech	Montreal, Canada	Several	fully
Miso Robotics	Pasadena, California	Fast food	partial
Hyper Robotics	Israel	Pizza	fully
Da Vinci Kitchen	Leipzig, Germany	Pasta	fully
Moley Robotics	London, UK	Several	fully
Blendid	Sunnyvale, California	Smoothies	fully
Nala Robotics	Illinois, USA	Several	partial or fully
Dexai Robotics	Boston, USA	Salads	fully
Picnic Pizza	Seattle, USA	Pizza	partial
Café X	San Francisco, USA	Coffee	fully
Roboeatz	Ontario, Canada	Several	fully
Mezli	San Mateo, California, USA	Bowls	fully
I-Robo	Tokyo, Japan	Bowls	partial
Hale	Singapore	Smoothies	fully
Cibotica	Canada	Salads and Bowls	partial
BreadBot	Washington, USA	Bread	fully
Wish&Cook	Portugal	Several	fully
Eatch	Netherlands	Several	fully

V. Conclusion

The imminent advent of the fifth industrial revolution promises transformative changes, particularly in the interface between industries and consumers, with profound implications for the food services sector. A systematic literature review was conducted employing a modified Non-negative Matrix Factorization algorithm to quantitatively identify emerging research trends and existing work at the intersection of smart cities, food, nutrition, and artificial intelligence.

The analysis reveals an increasing interest in this interdisciplinary domain within the scientific community, unveiling diverse subtopics for exploration. Through this lens, smart cities emerge as pivotal facilitators of seamless connections among consumers, farms, restaurants, and health systems. Leveraging automated management systems triggered by meal orders, these urban hubs hold promise for waste reduction, optimized environmental logistics, and the customization of consumer preferences, culminating in personalized health recommendations.

This literature review provides insight into how future smart cities, through the establishment of a network made of various participants interconnected with feedback loops, will facilitate the development of innovative food services. Key factors poised to shape the future of food services in smart cities are identified, including urban mobility, transportation logistics, diet monitoring, and waste reduction strategies. The integration of cutting-edge technologies such as blockchain, Internet of Things (IoT), and artificial intelligence and machine learning emerges as indispensable for streamlining processes and fostering interconnectedness.

ACKOWLEDGMENT

This research received funding from European Regional Development Fund (ERDF), Centro2020 and Agência Nacional de Inovação through the projet "Cook4Me", CENTRO-01-0247-FEDER-070247. A L Ferreira and F. Abreu also acknowledge funding from i3N project–UID/CTM/50025.

REFERENCES

[1] A. Lucena, A. S. Freitas, A. L. Ferreira, and F. Abreu, "Inspiration from Systematic Literature Reviews to Predict the Future of Food Services in Smart Cities," in *SMART 2023, The Twelfth International Conference on Smart Cities, Systems, Devices and Technologies*, ThinkMind, IARIA, Jun. 2023.

[2] M. Xu, J. M. David, S. H. Kim, *et al.*, "The Fourth Industrial Revolution: Opportunities and Challenges," *International Journal of Financial Research*, vol. 9, no. 2, pp. 90–95, 2018.

[3] A. Hassoun, A. Aıt-Kaddour, A. M. Abu-Mahfouz, *et al.*, "The Fourth Industrial Revolution in the Food Industry—Part I: Industry 4.0 Technologies," *Critical Reviews in Food Science and Nutrition*, pp. 1–17, 2022.

[4] A. Hassoun, A. E.-D. Bekhit, A. R. Jambrak, *et al.*, "The Fourth Industrial Revolution in the Food Industry—Part II: Emerging Food Trends," *Critical Reviews in Food Science and Nutrition*, pp. 1–31, 2022.

[5] P. Pathak, P. R. Pal, M. Shrivastava, and P. Ora, "Fifth Revolution: Applied AI & Human Intelligence with Cyber Physical Systems," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 3, pp. 23–27, 2019.

[6] S. M. Noble, M. Mende, D. Grewal, and A. Parasuraman, "The Fifth Industrial Revolution: How Harmonious Human–Machine Collaboration is Triggering a Retail and Service [R] Evolution," *Journal of Retailing*, vol. 98, no. 2, pp. 199–208, 2022.

[7] N. L. of Medicine, *PubMed*, https://pubmed.ncbi.nlm.nih.gov/.

[8] C. A. PLC, *Web of Science*, https://www.webofscience.com/wos/woscc/basic-search.

[9] G. S. University, *PubMed - Searching Medical Literature: Using Medical Subject Headings (MeSH)*, 2022. [Online]. Available: https://research.library.gsu.edu/c.php?g=115556%7B%5C&%7Dp=753156.

[10] A. I. for AI, *Semantic Scholar: AI-powered research tool for scientific literature*, https://www.semanticscholar.org/.

[11] J. Eggert and E. Korner, "Sparse Coding and NMF," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, IEEE, vol. 4, 2004, pp. 2529–2533.

[12] C. Ma, Y. Kamp, and L. F. Willems, "A Frobenius Norm Approach to Glottal Closure Detection from the Speech Signal," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 258–265, 1994.

[13] U. Tokkozhina, B. M. Mataloto, A. L. Martins, and J. C. Ferreira, "Decentralizing Online Food Delivery Services: A Blockchain and IoT Model for Smart Cities," *Mobile Networks and Applications*, pp. 1–11, 2023.

[14] A. Ragab, A. Osama, and A. Ramzy, "Simulation of the Environmental Impact of Industries in Smart Cities," *Ain Shams Engineering Journal*, vol. 14, no. 6, p. 102 103, 2023.

[15] C. K. Toh, "Tokyo's City Sustainability: Strategy and Plans for Net Zero Emissions by 2050," *IET Smart Cities*, vol. 4, no. 2, pp. 81–91, 2022.

[16] S. Sabrin, M. Karimi, and R. Nazari, "The Cooling Potential of Various Vegetation Covers in a Heat-Stressed Underserved Community in the Deep South: Birmingham, Alabama," *Urban Climate*, vol. 51, p. 101 623, 2023.

[17] T. Karakan, A. Gundogdu, H. Alagözlü, *et al.*, "Artificial Intelligence-Based Personalized Diet: A Pilot Clinical Study for Irritable Bowel Syndrome," *Gut Microbes*, vol. 14, no. 1, p. 2 138 672, 2022.

[18] F. Grundler, R. Mesnage, N. Goutzourelas, *et al.*, "Interplay Between Oxidative Damage, the Redox Status, and Metabolic Biomarkers During Long-Term Fasting," *Food and Chemical Toxicology*, vol. 145, p. 111 701, 2020.

[19] T. Fernández-Navarro, I. Díaz, I. Gutiérrez-Díaz, *et al.*, "Exploring the Interactions Between Serum Free Fatty Acids and Fecal Microbiota in Obesity Through a Machine Learning Algorithm," *Food Research International*, vol. 121, pp. 533–541, 2019.

[20] L. Zhou, D. Guan, L. Sun, *et al.*, "What Is the Relationship Between Ecosystem Services and Farmers' Livelihoods? Based on Measuring the Contribution of Ecosystem Services to Farmers' Livelihoods," *Frontiers in Ecology and Evolution*, vol. 11, p. 1 106 167, 2023.

[21] R. Mamlook, O. Al-Jayyousi, S. Alabyooki, and S. A. A. Attallah, "Smart City and Technology Transfer: Towards Sustainable Infrastructure in Bahrain," in *Research & Innovation Forum 2019: Technology, Innovation, Education, and their Social Impact 1*, Springer, 2019, pp. 255–268.

[22] C. Van Beukering, P. Hertzog, and A. J. Swart, "Investigating Thermal Supplementation of an Aquaponics System Under Severe Climate Conditions," *World Journal of Engineering*, vol. 18, no. 4, pp. 639–644, 2021.

[23] A. Kwasiborska, A. Stelmach, and I. Jabłońska, "Quantitative and Comparative Analysis of Energy Consumption in Urban Logistics Using Unmanned Aerial Vehicles and Selected Means of Transport," *Energies*, vol. 16, no. 18, p. 6467, 2023.

[24] W. A. Mohammad, Y. Nazih Diab, A. Elomri, and C. Triki, "Innovative Solutions in Last Mile Delivery: Concepts, Practices, Challenges, and Future Directions," in *Supply Chain Forum: An International Journal*, Taylor & Francis, vol. 24, 2023, pp. 151–169.

[25] A. Pal and K. Kant, "Smartporter: A Combined Perishable Food and People Transport Architecture in Smart Urban Areas," in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, 2016, pp. 1–8.

[26] C. A. da Silva, E. A. dos Santos, S. M. Maier, and F. S. da Rosa, "Urban Resilience and Sustainable Development Policies: An Analysis of Smart Cities in the State of São Paulo," *Revista de Gestão*, vol. 27, no. 1, pp. 61–78, 2020.

[27] J. Park and Y. Ahn, "Government Employees' Experience and Expectation of COVID-19 Hardships: The Moderating Role of Gender and Race in the United States," *The American Review of Public Administration*, vol. 52, no. 1, pp. 15–35, 2022.

[28] A. Fadda, B. Pace, A. Angioni, A. Barberis, and M. Cefola, "Suitability for Ready-to-Eat Processing and Preservation of Six Green and Red Baby Leaves Cultivars and Evaluation of Their Antioxidant Value during Storage and after the Expiration Date," *Journal of Food Processing and Preservation*, vol. 40, no. 3, pp. 550–558, 2016.

[29] R. Cozzolino, A. Martignetti, M. P. Pellicano, *et al.*, "Characterisation of Volatile Profile and Sensory Analysis of Fresh-Cut "Radicchio di Chioggia" Stored in Air or Modified Atmosphere," *Food Chemistry*, vol. 192, pp. 603–611, 2016.

[30] M. Cefola, V. Carbone, P. Minasi, and B. Pace, "Phenolic Profiles and Postharvest Quality Changes of Fresh-Cut Radicchio (Cichorium intybus L.): Nutrient Value in Fresh vs. Stored Leaves," *Journal of Food Composition and Analysis*, vol. 51, pp. 76–84, 2016.

[31] M. Thomsen, D. Romeo, D. Caro, M. Seghetta, and R.-G. Cong, "Environmental-Economic Analysis of Integrated Organic Waste and Wastewater Management Systems: A Case Study from Aarhus City (Denmark)," *Sustainability*, vol. 10, no. 10, p. 3742, 2018.

[32] N. Ullah, M. Klein, and J. Treur, "Food Desires, Negative Emotions, and Behaviour Change Techniques: A Computational Analysis," *Smart Cities*, vol. 4, no. 2, pp. 938–951, 2021.

[33] I. Makhdoom, I. Zhou, M. Abolhasan, J. Lipman, and W. Ni, "PrivySharing: A Blockchain-Based Framework for Privacy-Preserving and Secure Data Sharing in Smart Cities," *Computers & Security*, vol. 88, p. 101 653, 2020.

[34] A. K. Tripathi, K. Akul Krishnan, and A. C. Pandey, "A Novel Blockchain and Internet of Things-Based Food Traceability System for Smart Cities," *Wireless Personal Communications*, vol. 129, no. 3, pp. 2157–2180, 2023.

[35] G. Mirabelli and V. Solina, "Blockchain and Agricultural Supply Chains Traceability: Research Trends and Future Challenges," *Procedia Manufacturing*, vol. 42, pp. 414–421, 2020.

[36] M. Saniroychansyah, A. Farmawati, D. S. Anindyah, and L. Atianta, "Urban Compactness Effects on the Distributions of Healthy Houses

in Yogyakarta City," *Procedia-Social and Behavioral Sciences*, vol. 227, pp. 168–173, 2016.

[37] M. Trivedi, H. P. Bulsara, and Y. Shukla, "Why Millennials in Smart Cities Are Willing to Pay a Premium for Toxic-Free Food Products: Social Media Perspective," *British Food Journal*, 2023.

[38] H. P. Bulsara and M. Trivedi, "Exploring the Role of Availability and Willingness to Pay Premium in Influencing Smart City Customers' Purchase Intentions for Green Food Products," *Ecology of Food and Nutrition*, pp. 1–23, 2023.

[39] P. Moreno-Bernal, P. Arizmendi-Peralta, J. A. Hernández-Aguilar, J. del Carmen Peralta-Abarca, and J. G. Velásquez-Aguilar, "IoT Platform for Monitoring Nutritional and Weather Conditions of Avocado Production," in *Ibero-American Congress of Smart Cities*, Springer, 2022, pp. 95–109.

[40] S. Ahmadzadeh, T. Ajmal, R. Ramanathan, and Y. Duan, "A Comprehensive Review on Food Waste Reduction Based on IoT and Big Data Technologies," *Sustainability*, vol. 15, no. 4, p. 3482, 2023.

[41] H. Zia, H. S. Fatima, M. Khurram, I. U. Hassan, and M. Ghazal, "Rapid Testing System for Rice Quality Control through Comprehensive Feature and Kernel-Type Detection," *Foods*, vol. 11, no. 18, p. 2723, 2022.

[42] Q. Cao, "Study on the Screening of Active Components of Natural Products Based on Artificial Intelligence," in *2016 INTERNATIONAL CONFERENCE ON SMART CITY AND SYSTEMS ENGINEERING (ICSCSE)*, International Conference on Smart City and Systems Engineering (ICSCSE), Zhangjiajie, PEOPLES R CHINA, NOV 25-26, 2016, Hunan City Univ; Hunan Univ; Changsha Univ Sci & Technol, Commun Res Inst; Cent S Univ; Res Inst New Energy & Energy Saving & Emiss Reduct; Guangdong Acad Bldg Res, Res Inst Road & Bridge, 2016, pp. 606–609, ISBN: 978-1-5090-5530-2. DOI: 10.1109/ICSCSE.2016.124.

[43] N. Ratisoontorn and N. Chatwattanasiri, "Integrated Image Processing Framework to Determine Nutrient Quantities from Guideline Daily Amount (GDA) Label," in *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, IEEE, 2021, pp. 558–564.

[44] C. Liu, Y. Cao, Y. Luo, *et al.*, "Deepfood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment," in *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*, Springer, 2016, pp. 37–48.

[45] C. Cadenas-Sanchez, R. Cabeza, F. Idoate, *et al.*, "Effects of a Family-Based Lifestyle Intervention Plus Supervised Exercise Training on Abdominal Fat Depots in Children with Overweight or Obesity: A Secondary Analysis of a Nonrandomized Clinical Trial," *JAMA Network Open*, vol. 5, no. 11, e2243864–e2243864, 2022.

[46] C. Cadenas-Sanchez, M. Medrano, A. Villanueva, *et al.*, "Differences in Specific Abdominal Fat Depots Between Metabolically Healthy and Unhealthy Children with Overweight/Obesity: The Role of Cardiorespiratory Fitness," *Scandinavian Journal of Medicine & Science in Sports*, 2023.

[47] Y.-B. Zhou, K.-Y. Si, H.-T. Li, *et al.*, "Trends and Influencing Factors of Plasma Folate Levels in Chinese Women at Mid-Pregnancy, Late Pregnancy and Lactation Periods," *British Journal of Nutrition*, vol. 126, no. 6, pp. 885–891, 2021.

[48] A. Dolci, T. Vanhaecke, J. Qiu, *et al.*, "Personalized Prediction of Optimal Water Intake in Adult Population by Blended Use of Machine Learning and Clinical Data," *Scientific Reports*, vol. 12, no. 1, p. 19 692, 2022.

[49] R. Xu, B. E. Blanchard, J. M. McCaffrey, *et al.*, "Food Liking-Based Diet Quality Indexes (dqi) Generated by Conceptual and Machine Learning Explained Variability in Cardiometabolic Risk Factors in Young Adults," *Nutrients*, vol. 12, no. 4, p. 882, 2020.

[50] Z. Peng, D. Xu, Y. Li, Y. Peng, and X. Liu, "Phase Angle as a Comprehensive Tool for Nutritional Monitoring and Management in Patients with Crohn's Disease," *Nutrients*, vol. 14, no. 11, p. 2260, 2022.

[51] O. Raphaeli and P. Singer, *Towards Personalized Nutritional Treatment for Malnutrition using Machine Learning-Based Screening Tools*, 2021.

[52] S.-S. Ren, M.-W. Zhu, K.-W. Zhang, *et al.*, "Machine Learning-Based Prediction of In-Hospital Complications in Elderly Patients Using GLIM-, SGA-, and ESPEN 2015-Diagnosed Malnutrition as a Factor," *Nutrients*, vol. 14, no. 15, p. 3035, 2022.

[53] N. Kiss, M. Hiesmayr, I. Sulz, *et al.*, "Predicting Hospital Length of Stay at Admission Using Global and Country-Specific Competing Risk Analysis of Structural, Patient, and Nutrition-Related Data from nutritionDay 2007–2015," *Nutrients*, vol. 13, no. 11, p. 4111, 2021.

[54] Z. Tatarkova, J. H. de Baaij, M. Grendar, *et al.*, "Dietary Mg2+ Intake and the Na+/Mg2+ Exchanger SLC41A1 Influence Components of Mitochondrial Energetics in Murine Cardiomyocytes," *International Journal of Molecular Sciences*, vol. 21, no. 21, p. 8221, 2020.

[55] Q. Liu, X. Yang, M. Luo, *et al.*, "An Iterative Neural Processing Sequence Orchestrates Feeding," *Neuron*, vol. 111, no. 10, pp. 1651–1665, 2023.

[56] A. Z. Shafiullah, J. Werner, E. Kennedy, *et al.*, "Machine Learning Based Prediction of Insufficient Herbage Allowance with Automated Feeding Behaviour and Activity Data," *Sensors*, vol. 19, no. 20, p. 4479, 2019.

[57] V. E. Cabrera, J. A. Barrientos-Blanco, H. Delgado, and L. Fadul-Pacheco, "Symposium Review: Real-Time Continuous Decision Making Using Big Data on Dairy Farms," *Journal of Dairy Science*, vol. 103, no. 4, pp. 3856–3866, 2020.

[58] M. Frizzarin, T. O'Callaghan, T. B. Murphy, D. Hennessy, and A. Casa, "Application of Machine-Learning Methods to Milk Mid-Infrared Spectra for Discrimination of Cow Milk from Pasture or Total Mixed Ration Diets," *Journal of Dairy Science*, vol. 104, no. 12, pp. 12 394–12 402, 2021.

[59] N. Matsumoto, J. Park, R. Tomizawa, *et al.*, "Relationship Between Nutrient Intake and Human Gut Microbiota in Monozygotic Twins," *Medicina*, vol. 57, no. 3, p. 275, 2021.

[60] M. A. Rubel, A. Abbas, L. J. Taylor, *et al.*, "Lifestyle and the Presence of Helminths Is Associated with Gut Microbiome Composition in Cameroonians," *Genome Biology*, vol. 21, no. 1, pp. 1–32, 2020.

[61] C. I. Le Roy, "Defining and Directing Gut Microbiota Maturation in Early Life: A Nutrition Perspective," *Intersections of Nutrition: Retracing Yesterday, Redefining Tomorrow*, vol. 97, pp. 11–19, 2023.

[62] K. Veselkov, G. Gonzalez, S. Aljifri, *et al.*, "HyperFoods: Machine Intelligent Mapping of Cancer-Beating Molecules in Foods," *Scientific Reports*, vol. 9, no. 1, 2019.

[63] G. Hu, M. Ahmed, and M. R. L'Abbé, "Natural Language Processing and Machine Learning Approaches for Food Categorization and Nutrition Quality Prediction Compared with Traditional Methods," *The American Journal of Clinical Nutrition*, vol. 117, no. 3, pp. 553–563, 2023.

[64] C. R. Davis, K. J. Murphy, R. G. Curtis, and C. A. Maher, "A Process Evaluation Examining the Performance, Adherence, and Acceptability of a Physical Activity and Diet Artificial Intelligence Virtual Health Assistant," *International Journal of Environmental Research and Public Health*, vol. 17, no. 23, 2020.

[65] S. P. Goldstein, F. Zhang, J. G. Thomas, *et al.*, "Application of Machine Learning to Predict Dietary Lapses During Weight Loss," *Journal of Diabetes Science and Technology*, vol. 12, no. 5, pp. 1045–1052, 2018.

[66] M. Zahid and U. Sughra, "Development of a Protocol for Conducting a Randomized Control Trial on Effects of Artificial Intelligence on Nutritional Status of Children Post Cardiac Surgery," *JPMA*, vol. 72, no. 908, 2022.

[67] D. Kirk, C. Catal, and B. Tekinerdogan, "Precision Nutrition: A Systematic Literature Review," *Computers in Biology and Medicine*, vol. 133, p. 104 365, 2021.

[68] P. Kokol, V. Brumec, A. Habjanič, *et al.*, "Intelligent Systems for Nursing Education," in *MEDINFO 2001*, IOS Press, 2001, pp. 1047–1050.

[69] K. Li, K. Bertrand, J. C. Naviaux, *et al.*, "Metabolomic and Exposomic Biomarkers of Risk of Future Neurodevelopmental Delay in Human Milk," *Pediatric Research*, vol. 93, no. 6, pp. 1710–1720, 2023.

[70] M. C. Playdon, J. N. Sampson, A. J. Cross, *et al.*, "Comparing Metabolite Profiles of Habitual Diet in Serum and Urine," *The American Journal of Clinical Nutrition*, vol. 104, no. 3, pp. 776–789, 2016.

[71] N. Alshurafa, A. W. Lin, F. Zhu, *et al.*, "Counting Bites with Bits: Expert Workshop Addressing Calorie and Macronutrient Intake

Monitoring," *Journal of Medical Internet Research*, vol. 21, no. 12, e14904, 2019.

[72] R. Weiss, P. J. Stumbo, and A. Divakaran, "Automatic Food Documentation and Volume Computation Using Digital Imaging and Electronic Transmission," *Journal of the American Dietetic Association*, vol. 110, no. 1, p. 42, 2010.

[73] H. Hassannejad, G. Matrella, P. Ciampolini, *et al.*, "Automatic Diet Monitoring: A Review of Computer Vision and Wearable Sensor-Based Methods," *International Journal of Food Sciences and Nutrition*, vol. 68, no. 6, pp. 656–670, 2017.

[74] M. Durand, C. Largouët, L. B. de Beaufort, J. Y. Dourmad, and C. Gaillard, "Prediction of the Daily Nutrient Requirements of Gestating Sows Based on Sensor Data and Machine-Learning Algorithms," *Journal of Animal Science*, vol. 101, skad337, 2023.

[75] H. Liu, R. Hu, R. Zhang, *et al.*, "The Prediction of Human Oral Absorption for Diffusion Rate-Limited Drugs Based on Heuristic Method and Support Vector Machine," *Journal of Computer-Aided Molecular Design*, vol. 19, pp. 33–46, 2005.

[76] R. Z. Zhou, Y. Hu, J. N. Tirabassi, Y. Ma, and Z. Xu, "Deriving Neighborhood-Level Diet and Physical Activity Measurements from Anonymized Mobile Phone Location Data for Enhancing Obesity Estimation," *International Journal of Health Geographics*, vol. 21, no. 1, pp. 1–18, 2022.

[77] H. Marcos-Pasero, G. Colmenarejo, E. Aguilar-Aguilar, *et al.*, "Ranking of a Wide Multidomain Set of Predictor Variables of Children Obesity by Machine Learning Variable Importance Techniques," *Scientific Reports*, vol. 11, no. 1, p. 1910, 2021.

[78] N. Lazzarini, J. Runhaar, A. Bay-Jensen, *et al.*, "A Machine Learning Approach for the Identification of New Biomarkers for Knee Osteoarthritis Development in Overweight and Obese Women," *Osteoarthritis and Cartilage*, vol. 25, no. 12, pp. 2014–2021, 2017.

[79] N. Reščič, T. Eftimov, B. Koroušić Seljak, and M. Luštrek, "Optimising an FFQ Using a Machine Learning Pipeline to Teach an Efficient Nutrient Intake Predictive Model," *Nutrients*, vol. 12, no. 12, p. 3789, 2020.

[80] Y.-L. Yang, H.-L. Yang, J. D. Kusuma, and S.-Y. P. K. Shiao, "Validating Accuracy of an Internet-Based Application against USDA Computerized Nutrition Data System for Research on Essential Nutrients among Social-Ethnic Diets for the E-Health Era," *Nutrients*, vol. 14, no. 15, 2022.

[81] E. Otto, C. Semotok, J. Andrysek, and O. Basir, "An Intelligent Diabetes Software Prototype: Predicting Blood Glucose Levels and Recommending Regimen Changes," *Diabetes Technology & Therapeutics*, vol. 2, no. 4, pp. 569–576, 2000.

[82] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early Detection of Type 2 Diabetes Mellitus Using Machine Learning-Based Prediction Models," *Scientific Reports*, vol. 10, no. 1, p. 11 981, 2020.

[83] M. Livingstone and P. Robson, "Measurement of Dietary Intake in Children," *Proceedings of the Nutrition Society*, vol. 59, no. 2, pp. 279–293, 2000.

[84] D. Rhyner et al., "Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals with Type 1 Diabetes Mellitus: A Comparative Study," *Journal of Medical Internet Research*, vol. 18, no. 5, e101, 2016.

[85] C. Baldwin, N. Wilberforce, and A. Kapur, "Restaurant and Food Service Life Cycle Assessment and Development of a Sustainability Standard," *The International Journal of Life Cycle Assessment*, vol. 16, pp. 40–49, 2011.

Coherent Multi-disciplinary and Fact-based Contextualisation: Comprehensive Knowledge Complements and Component Frameworks Employed in Prehistorical Archaeology and Historico-cultural Contexts

Claus-Peter Rückemann
Unabhängiges Deutsches Institut für
Multi-disziplinäre Forschung (DIMF), Germany
Universität Münster, Germany
Leibniz Universität Hannover, Germany
Email: ruckema@uni-muenster.de

Birgit Frida Stefanie Gersbeck
Unabhängiges Deutsches Institut für
Multi-disziplinäre Forschung (DIMF), Germany
Email: info@udimf.de

***Abstract*—This extended paper presents the new results of the methodological fact-based knowledge contextualisation, discovery, and parallelisation of workflow logic of prehistorical archaeology discipline's contextualisation, based on coherent multi-disciplinary conceptual knowledge. This research targets coherent multi-disciplinary and fact-based contextualisation, which can provide important new solutions to the challenges of knowledge discovery in many disciplines. The paper includes the comprehensive knowledge complements and component frameworks employed in prehistorical archaeology and historico-cultural contexts. The practical example implementations based on both the knowledge frameworks and the components framework comprise the contextualisation and canvas conversion benchmarks and enable sustainable and efficient realisations. The goal is the creation of efficient, flexible, and sustainable contextualisation workflows, also providing efficient parallelised frame conversion. Implementations and realisations are enabled by the latest versions of the prehistory-protohistory and archaeology conceptual knowledge reference implementation and the component reference implementations framework. The paper provides the results on archaeological / prehistorical facts, universal contexts, and logical and formal entities, factual, conceptual, and procedural complements, components, and results required for exemplary practical hard criteria and fact-based contextualisation by the disciplines and even for consequent creative historico-cultural context exploitation. Future research will address the creation and further development of a conceptual knowledge reference implementation and a component reference framework for coherent multi-disciplinary conceptual contextualisation, enabling multi-disciplinary equal footing with contributions from all scientific disciplines for example for prehistorical archaeology knowledge integration, contextualisation and analysis with prehistorical and archaeological knowledge resources.**

***Keywords*–Fact-based Contextualisation; Benchmarking; Prehistory and Archaeology; Historico-cultural Interpretation; CKRI and CRI Framework.**

I. Introduction

This research targets coherent multi-disciplinary and fact-based contextualisation, which can provide important new solutions to the challenges of knowledge discovery in many disciplines. This paper is an extended and updated presentation of new research results, based on the research, publication, and presentation at the DataSys / INFOCOMP 2023 co-located group of conferences in Nice, France [1].

The following sections present the comprehensive knowledge complements and component frameworks employed in prehistorical archaeology and historico-cultural contexts.

The examples show advanced contextualisation reference implementation frameworks and coherent multi-disciplinary conceptual knowledge-spatial context discovery in practice. This goes along with a wider range of new survey result groups of archaeological settlement infrastructures and insights based on practical conceptual facets used for on demand contextualisation and symbolic representation [2].

The theoretical and practical developments of this ongoing research are one group of six impactful research endeavors selected worldwide for the 2023 Hyperion Research special report on recent High-Performance Computing (HPC) centric AI success stories [3] (AI, "Artificial Intelligence"). For the inclusion in the special report, Hyperion names the epistemologically relevant methodological fundament The Conceptual Knowledge Reference Implementation (CKRI) and the The Component Reference Implementations framework (CRI), providing sustainable standard component groups for implementation.

The entirety enables a consistent multi-disciplinary conceptual knowledge contextualisation, especially universal knowledge facets, method integration, flexible workflow definition, and parallelisation on HPC resources.

Due to recent interests and inquiries for practical cases of extended multi-disciplinary contextualisation matrices employing the frameworks for knowledge complements and components, we take the chance that this extended paper concentrates on an extension, on both the practical contextualisation matrix and the components, including a major example of sustainable component parallelisation and its benchmark results.

The need to implement workflows of many disciplines beyond 'manual operation' has been continuously increasing over the last decades. In practice, we can also see a strong motivation for sustainable Knowledge Resources (KR) creation and development and efficient employment of resources, e.g., with high performance computation and storage.

After reading, writing, and arithmetic are established and accepted as general competences, the capabilities of achieving efficient analysis and contextualisation solutions are becoming 'state-of-the-art' increasingly important personal competences in the sciences. Efficiently and sustainably organising and de-isolating knowledge complements, the results of research, within a discipline is at least as important as its short term analysis. These organisation processes cannot refer only to the data if they should be useful for reuse of knowledge and insight. Therefore, such organisation can in no ways be seen technically or being task of third parties without risk of losing the competence on fact-based methods, insight, and interpretation. Additionally, organisation of its knowledge complements is a core scientific matter of any discipline and closely associated with the methods employed, with the ongoing analysis processes, and with the further interpretation potential. In accordance with best practice, any scientists dealing with methodological workflows in a discipline, e.g., when applying a method, should know and practise themselves the way steps can be created, organised, and implemented, e.g., the algorithms, symbolic representation, and the structure and computation related characteristics. This practice is especially relevant with all knowledge complements or in other words the non-technical aspects of workflows in prehistorical archaeology. Scientific work, including state-of-the-art practices in archaeological disciplines and humanities, comprise of a number of essential principles, including further continuous employment of valuation methods for new factual knowledge and insights, re-contextualisation and resources development, and consequent fact-based contextualisation, analysis, and interpretation. When done properly, the tasks including contextualisation allow to practise equal footing with contributing scientific disciplines. Numerous surveys and studies were conducted for archaeological and prehistorical cases and multi-disciplinary contexts during the last decades, e.g., specific object groups' contextualisation [4] and discovery [5] and providing factual knowledge for interpretation, including historico-cultural contexts. The Prehistory and Archaeology Knowledge Archive (PAKA) is continuously collecting [6] new knowledge and insight. This research delivers the respective blueprints resulting from previously unpublished contexts and workflows and efficient workflow implementations proven sustainable over many years and widely reusable.

The rest of this paper is organised as follows. Section II presents the fundaments and state-of-the-art methodological implementations and realisations employed. Section III presents the results of the contexts and workflow logic for processes in prehistorical archaeology, factual/conceptual and procedural complements. Section IV delivers the discipline's results, efficiency results, and discussion for the presented contexts and workflow logic cases. Section V summarises lessons learned, conclusions, and future work.

II. Fundaments and Previous Work

Two major practical reference implementations were deployed for full implementations, realisations, and continuous further developments: the latest versions of the prehistory-protohistory and archaeology Conceptual Knowledge Reference Implementation (CKRI) [4], [7] and the Component Reference Implementations (CRI) framework [8], [9] for conceptual knowledge-based context integration, complements processing, and geoscientific visualisation. CKRI provides the universal knowledge framework, including multi-disciplinary contexts of natural sciences and humanities [10]. CRI provides the required component groups and components for the implementation and realisation of all the procedural modules.

Both, the ongoing research on the epistemological [4] and the components base [8] and its implementations were recently published and demonstrated at the Lawrence Livermore National Laboratory (LLNL) [11].

The reference implementations are based on the fundamental methodology of knowledge complements [12], considering that many facets of knowledge, including prehistory, need to be continuously acquired and reviewed [13]. Creating contextualisation requires to coherently integrate multi-disciplinary knowledge and to enable symbolic representations. Realisations need to integrate a wide range of components as required from participating disciplines, e.g., for dynamical processing, geoprocessing, spatial contextualisation. Prehistoric object groups and contexts are taken from the latest edition of PAKA, which is in continuous development for more than three decades [14], and from The Natural Sciences KR (NatSciKR), all released by DIMF [6]. The PAKA and The NatSciKR support Factual, Conceptual, Procedural, Metacognitive, Structural (FCPMS) knowledge complements [11] and enable seamless coherent multi-disciplinary conceptual knowledge integration and 'systematical' and methodological approaches based on CKRI for creation and development of workflow procedures. CKRI references are illustrated for demonstration via the multi-lingual Universal Decimal Classification (UDC) summary [15] released by the UDC Consortium under Creative Commons licence [16].

It is important to say that the reader cannot expect to understand the implementations in detail beyond the focus of coherent multi-disciplinary contextualisation of knowledge. We also have to use the precise specification 'coherent multi-disciplinary contextualisation' even as it may appear lengthy to today´s reader. It is as well important to emphasise that the sort order of the terms is significant. Informatics and technology is considered only a minor part of holistic information science. Information sciences can very much benefit from Aristotle's fundaments and a knowledge-centric approach [17] but for building holistic and sustainable solutions, supporting a modern definition of knowledge [18], they need to go beyond the available technology-based approaches and hypothesis [19] as analysed in Platon's Phaidon. The references of the employed components or individual implementations provide several ten thousand pages of documentation and thousands of tiny working examples on plain realisation issues.

For sustainable understanding and implementation by researchers new to these topics, it is recommended to include information science fundaments, e.g., epistemology and logic, during all stages of educational processes of all disciplines. Comprehensive studies of the works of the classical Organon, Categories [20], On Interpretation [21], Prior Analytics [22], Posterior Analytics [23], Topics [24], and On Sophistical Refutations [25] and understanding of logical fallacies [26] can be beneficial as starting points. These works deal with many fundaments required for understanding the being of knowledge, e.g., logic, classification, language, and the fundaments of analysis, and create a base for sustainable implementations and realisations. In most cases, it is also recommended to invest in understanding the discussion of unintelligence [27].

III. Discipline's Contexts and Workflow Logic

Prehistorical archaeology discipline's resulting contexts and workflow logic are often matter of multi-disciplinary long-term research, which requires universal context identification and assignment to contributing scientific disciplines.

A. Resulting Factual and Conceptual Complements Blueprint

The discipline's factual and conceptual knowledge complements and major logical and formal entities resulting from the long-term surveys and practical implementations are given in Table I. Employed resources are High Resolution (HR) Digital Elevation Model (DEM) data, e.g., (Space) Shuttle Radar Topography Mission (SRTM) data [28], updates [29], and further satellite data.

Common DEM can be supplemented by local Light Detection And Ranging (LiDAR) data for special features and resolutions. DEM data for spatial contexts is used via Network Common Data Form (NetCDF) [30], developed by the University Corporation for Atmospheric Research (UCAR/Unidata), [31], National Center for Atmospheric Research (NCAR). [32]

KR and complement implementations in contributing contexts and disciplines are PAKA and NatSciKR [6] accompanied by HR Digital Chart of the World (DCW) [33], and Global Self-consistent Hierarchical High-resolution Geography (GSHHG) [34].

Knowledge context, e.g., targeting symbolism and analysis have been integrated from a wide range of cases and disciplines [35] [36] [37] [38] [39] [40] [41] [42] [43] spanning scenarios over the continents and including object security and verification [44].

The symbolic representation of the contextualisation can be done with a wide range of methods, algorithms, and available components, e.g., via LX Professional Scientific Content-Context-Suite (LX PSCC Suite) [45] deploying the Generic Mapping Tools (GMT) and integrated modules [46] for visualisation. The GMT [47] suite application components are used for handling the spatial data (longitude, latitude, elevation), applying the related criteria, and for the visualisation. For sustainability we also consequently employ xyz files in GMT, e.g., Point of Interest (PoI) and Point of Discovery (PoD) contexts. Signatures and Colour Palettes (cpt) can also be flexibly integrated via GMT. For support of many algorithms and features, suites of full on-site on-premise environments and online environments were created, The LX Professional Scientific Content-Context-Suite (PSCC) [45] integrating The LX Professional Scientific Creation Academic Suite (PSCA) [48] supporting the TEX Live [49] and tool components. A good starting point for new projects is to install and configure the latter online [50], available under MIT licence [51], adding additional modules as required, building individual suites for multi-disciplinary use cases.

Mostly all contexts and object groups are in continuous development, based on their structural implementations. Practically all contexts are dealt with employing the CKRI and its facets and operation facilities. Many properties of the contexts, e.g., chorological and chronological properties, can be addressed using international standards, e.g., for georeferencing and time, addressing the Chorological Domain (ChDo) and Chronological Domain (CrDo). The consequent knowledge approach enables a wide range of workflow creation and analysis, in the scenarios discussed here ranging from fact-based contextualisation to consequent fact-based historico-cultural interpretation. The results allow even further consequent creative historico-cultural exploitation.

B. Resulting Procedural Complements Blueprint

The discipline's resulting procedural complements (the knowledge complements) and the corresponding workflow implementation resulting from the long-term surveys and practical implementations are given in Table II.

The implementations are designed for end-user deployment by members of every responsible discipline dealing with their major logical and formal entities. The matrix shows context / object groups, required logical and formal workflow entities (major processing groups pre, main, post), examples of their symbolic representation, structure and procedure implementations.

The table confirms that all contexts and object groups are in continuous further development, including the implementations of knowledge complements, e.g., factual, conceptual, procedural, and structural, which is a major achievement for scientific best practice and sustainability. The characteristics include the contexts addressed with CKRI and georeferencing, as well as the potential of mostly all contexts can be deployed in workflow parallelisation.

The table especially lists excerpts of embarrassingly (E) and loosely (L) parallelisation features. Both of these types are considered common features with parallelisation. What is special here is that most features can be implemented in either ways due to the component framework so that realisations can cope with many conditions for different environments.

All the CRI components of the workflow blueprint allow very high flexibility for fact-based methods and context integration of scientific, fact-based symbolic representation, e.g., the symbolic representation of archaeological, prehistorical contexts requires the employment of different geographic projections, e.g., geospherical orthographic, isometrical, and equal area. Projections can be flexibly implemented via GMT [47] and via PROJ [52].

Besides the implemented components we already named: The workflow allows processing usable for most disciplines, Area of Interest (AoI) calculations, regular expression patterns for context structures, e.g., via Perl Compatible Regular Expressions (PCRE) [53]. Attributations not applicable (n.a.) are marked accordingly. Workflow output, e.g., frames and visualisation can be created for many common structures, e.g., Joint Photographic Experts Group (JPG), Portable Network Graphics (PNG), and Portable Document Format (PDF),

Table I. Prehistorical Archaeology Discipline's Contexts and Logical / Formal Entities: Resulting Factual / Conceptual Contextualisation Matrix Implemented for Major Complements and Components (Excerpt).

Context / Discipline / Object Group	*Logical / Formal Entities*	*Symb. Repr. (Example)*	*Structure Impl. (Example)*	*In Dev.*	*CKRI*	*Georef.*	*Parallelisation E*	*L*
			Factual / Conceptual Domain (Focus Complements: FCPMS)					
Hybrid	(Spatial) structure	Signature / cpt	netCDF, GMT, LX PSCC	✓	✓	✓	✓	✓
		HR DEM	netCDF, GMT, LX PSCC	✓	✓	✓	✓	✓
		SRTM	netCDF, GMT, LX PSCC	✓	✓	✓	✓	✓
		LiDAR	netCDF, GMT, LX PSCC	✓	✓	✓	✓	✓
Point	Singular structure	Signature / Symbol	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Prehistorical archaeology		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Settlements		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Ritual places		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Notable objects		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Geophysics		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Impact craters		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Planetology		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Plate tectonics features		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Plate tectonic boundary types		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Geology		NatSciKR, PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Mineral resources		NatSciKR, PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Pedology		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Soil characteristics		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Bog and peat characteristics		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Volcanology		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Volcanological features		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Strato volcanoes, maars, …		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Speleology		NatSciKR, PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Caves		NatSciKR, PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Oceanography		NatSciKR, GSHHG	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Bathymetry features		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Hydrology		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Rivers		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Lakes		NatSciKR, DCW	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Mobility, transport		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Pre-modern holloways		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Pre-modern trackways		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Pre-modern bog trackways		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Linguistics		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Open field names		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Geography		NatSciKR	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Topography features		NatSciKR	xyz, GMT, LXPSCC	✓	✓	✓	✓	✓
Heights		NatSciKR	xyz, GMT, LXPSCC	✓	✓	✓	✓	✓
Humanities, politics features		NatSciKR, DCW	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Humanities, administrative		NatSciKR, DCW	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
…	…	…	…	…	…	…	…	…
Line / Polygon	Linear structure	Signature	xyz, GMT, LXPSCC	✓	✓	✓	✓	✓
Prehistorical archaeology		PAKA	xyz, GMT, LXPSCC	✓	✓	✓	✓	✓
…	…	…	…	…	…	…	…	…
Polygon	Areal structure	Signature	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Prehistorical archaeology		PAKA	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Bathymetry features		GSHHG, DEM	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
Administrative features		DCW	xyz, GMT, LX PSCC	✓	✓	✓	✓	✓
…	…	…	…	…	…	…	…	…

as well as Motion/Moving Pictures Expert Group, version 4 (MP4). Transformation can also be done for Keyhole Markup Language generation and interactive clipping zooming [54] [55]. Further preferred workflow components, automation, and features are provided by FFmpeg [56] [57], ImageMagick/-GraphicsMagick [58], GIMP [59], SubRip Subtile File (SRT), and VLC media player [60]. For efficiency of data handling, contextualisation workflows can integrate compression modules supporting multi-threading, e.g., pbzip2, using several streams concatenatable into the resulting output [61].

Multi-dimensional or sequences of view, e.g., focus dependent views for knowledge dimensional computation per object, are implemented via OpenMP [62] and specifications [63]. Job parallel procedures, e.g., for knowledge objects and resources

TABLE II. BLUEPRINT OF PREHISTORICAL ARCHAEOLOGY DISCIPLINE'S WORKFLOW LOGIC: RESULTING PROCEDURAL CONTEXTUALISATION MATRIX, FROM FC (TABLE I), IMPLEMENTED FOR MAJOR COMPLEMENTS AND COMPONENTS, INCLUDING PARALLEL FRAME CONVERSION (EXCERPT).

Context / Discipline / Object Group	*Logical / Formal Entities*	*Symb. Repr. (Example)*	*Struc. / Proc. Impl. (Example Complement / Environment)*	*In Dev.*	*CKRI*	*Georef.*	*Parallel. E*	*Parallel. L*
			Procedural Domain (Focus Complements: FCPMS)					
Selection, preparation (KR)	Pre-processing	Pre-routines	CKRI, PCRE, … / LX PSCC	✓	✓	(✓)	(✓)	(✓)
Context resources	Pre-processing	Pre-routines	netCDF, CKRI, PCRE, … / LX PSCC	✓	✓	(✓)	(✓)	(✓)
Sequence	Pre-proc., timing structure	Parameter	[FCPMS] / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Procedure modules	Main processing	Main-routines	[FCPMS)], …	✓	✓	(✓)	(✓)	(✓)
Contextualisation Scenario	Integration	Hybrid	… / GMT, LX PSCC …	✓	✓	(✓)	(✓)	(✓)
Observer path	Path / project	Line	xyz / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Observer track	Track / project	Line	xyz / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
AoI	Selection, cut	Area	netCDF, xyz / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Sampling	Resampling	[Raster]	netCDF / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Canvas mapping	Basemap	[Mapping]	netCDF / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Gridding	Grid operations, …	[Grid]	netCDF / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Illumination	Height	Singular	netCDF / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Math operations	Calculation	[Algorithm]	… / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Triangulation	Calculation	[Algorithm]	… / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Regression	Calculation	[Algorithm]	… / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Colour	Colourisation	[Sequence]	cpt / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Filtering	Selection, select	[Decimation]	… / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Movie module	Iteration, …	Parameter	… / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Limits to cores	[Integer]	Parameter	… / GMT, LX PSCC	n.a.	n.a.	n.a.	n.a.	n.a.
Canvas	[Integer×Integer]	Parameter	… / GMT, LX PSCC	n.a.	n.a.	n.a.	n.a.	n.a.
Frames	[Integer]	Parameter	… / GMT, LX PSCC	n.a.	n.a.	n.a.	n.a.	n.a.
Display rate	[Integer]	Parameter	… / GMT, LX PSCC	n.a.	n.a.	n.a.	n.a.	n.a.
Video product selection	[Type]	Parameter	… / GMT, LX PSCC	n.a.	n.a.	n.a.	n.a.	n.a.
Proc. of knowledge compl.	Calculation	[Algorithm]	CKRI, PCRE, … / LX PSCC	✓	✓	(✓)	(✓)	(✓)
Spatial proc. of preh. ctxts.	Selection, calculation	[Algorithm]	CKRI, PCRE / GMT, …, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Events	Symbolic, functional	Symb. repr.	CKRI, PCRE / GMT, …, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Arbitrary symbols	Symbolic, functional context	Symb. repr.	[vector graphics] / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Degenerated ellipses	Azimuthal	Area	xyz / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Range	Azimuthal	Area	xyz / GMT, LX PSCC	✓	✓	(✓)	(✓)	(✓)
Projection	Geospherical, orthographic	[Algorithm]	… / GMT, PROJ, LX PSCC	✓	✓	(✓)	(✓)	(✓)
…	…	…	…	…	…	…	…	…
Resource usage (main proc.)	Main proc., parallelisation	Frame, view	… / OpenMP, GNU parallel, LX PSCC	✓	✓	n.a.	✓	✓
	On-scratch processing	Various	… / OpenMP, GNU parallel, LX PSCC	✓	✓	n.a.	✓	✓
	Model reduction frame, anim.	Various	… / GMT, (LX PSCC)	✓	✓	n.a.	✓	✓
	Live frame control	Various	JPG, PNG, PDF / (LX PSCC)	✓	✓	n.a.	✓	✓
Transform., symbolic repr.	Post-processing, batch	Post-routines	Scales, KML, … / LX PSCC	✓	✓	n.a.	(✓)	(✓)
Visualisation, analysis	Post-processing, interactive	Image, Video	PNG, MP4, … / LX PSCC	✓	✓	n.a.	(✓)	(✓)

localities, are supported by respective modular solutions [64]. The implementations under Linux in general showed very efficient, both for OpenMP and GNU Parallel (GNU, Gnu's Not Unix).

C. Supporting Varieties of Workflows

As recommended in the blueprint section, several configurations and developments were added for multi-disciplinary contextualisation scenarios employed for prehistorical archaeology and historico-cultural contexts.

Regarding the large scale and long-term nature of this research, collaborative services, e.g., The Archaeological Data Collector (ADC) [65], were developed and established as interfaces deployed for the support of practical workflows in prehistorical and archaeological surveys and excavations and the further and ongoing development of resources.

In this scenario, ADC is developed in order to provide online interfaces to all participating parties, supporting distributed workflow components, e.g., for collection, standardisation, conversion, verification, integration, and analysis of contributions. The components fully support all relevant sustainable structures, formats, and codes, e.g., American Standard Code for Information Interchange (ASCII), UCS (Universal Character Set) / UTF (UCS Transformation Format), and TEX [49] representations.

The services support structured data and formatted data and are used for archaeological survey contributions, data collection, and for ongoing data integration and analysis. For example, context structures of the contextualisation matrix, which comprise of a large number of disciplines contributions and knowledge complements and can be fully integrated with the PSCC [45] and PSCA [48] suites.

IV. Discipline's Results and Workflow Efficiency

A. Discipline's Workflow: Parallelisation and Results

Table III shows the scalability of the example workflow procedure for parallelised processing parts (pre, timing; main, parallelisation; post, batch) of the coherent multi-disciplinary conceptual knowledge. The results are referring to a scenario of a set of 1440 frames created in parallel for 4 k canvas size for a 60 s sequence with a rate of 24 FPS (Frames Per Second).

Table III. Scalability of Discipline's Workflow (Example Runs, Parallelised Processing KR and Context Resources).

Threads (Cores)	*Wall Time*				
	Pre, Timing	*Main, Parallel*	*Post, Batch*	$\Sigma Pre, Main, Post$	
1	1145 s	2581175 s	84972 s	2667292 s	$\approx 741\,h$
18	526 s	143668 s	4759 s	143668 s	$\approx 40\,h$
36	262 s	71833 s	2386 s	74481 s	$\approx 21\,h$

The architecture chosen for this realisation is an efficient 36-core-based Central Processing Unit (CPU) (Intel® Xeon® CPU E5-2695 v4 at 2.10 GHz), which is taking into account that we commonly use 36 cores for many basic global approaches, e.g., considering 360 degrees of a global model. Results for trends on other architectures with same numbers of respective cores should be highly comparable.

Precondition for parallelisation is sufficient memory for parallel use of integrated resources. Considering the employed resources, especially SRTM / NetCDF and KR, 128 GB RAM (Random Access Memory) for 36 parallel processes is comfortable when data limits are cut to the limits required for the algorithms with the range of a few hundred kilometres area per object entity. The parallel instances are allowed for 90 GB HDD (Hard Disk Drive) space and separate 50 GB SSD (Solid State Disk) space for highly volatile data of parallel instances.

Wall and compute times, especially of multi-dimensional workflow results, can greatly be reduced from the integrated parallelisation, which makes the procedural solution highly scalable. The wall times for thread numbers confirm the high scalability when implementations of the workflow are using higher numbers of threads.

Many practical workflows may contain some parts which cannot be reasonably parallelised. This is especially true for scientific tasks with a certain complexity. Anyhow, the percentage of non parallelised parts is very low with CKRI and the CRI framework. However, individual instances may show non-linear characteristics due to instance content and references, e.g., different satellite data, different data types, and different knowledge complements. For large sets, hundreds up to thousands of CPU cores were employed, so parallelised wall times per object can be very reasonably reduced from days to hours or even minutes, e.g., for warning and tracking systems.

The following results from the above discipline workflow show an excerpt of eight frames from a large frame sequence for calculated Areas of Interest (AoI) contexts in top views (Figure 1). The foci are prehistoric settlement infrastructures in factual and historico-cultural chorological and chronological contexts with a volcanological features group (maars, Holocene-historical) and satellite data based on the coherent conceptual knowledge integration and discovery.

Ellipsoid is World Geodetic System 84 (WGS-84). Projection for frames is Lambert Azimuthal Equal Area. The resolution is drastically reduced for use in this publication. Each frame is calculated to represent the contextualisation of knowledge complements under chorological and chronological conditions for 150 km radii. Generated representations include integrated CKRI references, projection of topographic and bathymetric results, and further knowledge for respective areas, based on the coherent conceptual knowledge. The frame sequence of symbolic representations enable to contextualise named factual data (CKRI: UDC:551.2…,551.21,550.3,(23);"62" and UDC:167/168…;51… referring to CKRI: UDC:711.…,692,903,902 for 150 km radii) [4], demonstrated via UDC [15].

Major multi-disciplinary results are the shown insights regarding the details of prehistoric settlement infrastructures / Holocene maars for which we find larger numbers of prehistorical settlements were set and used in the volcanic regions Eifel (DE) and Auvergne (FR) areas than in the other areas, all of which can be precisely assigned and further contextualised.

The workflow results are computed for eight equidistant steps for on-track equal-distance AoI locations (yellow symbols), including prehistoric settlement infrastructures (blue rectangular symbols) and volcanological groups (colored volcano symbols), Holocene-historical maars.

Especially, the connection between the dormant West Eifel Volcanic Field maars and the Auvergne (dormant Monchatre maar) shows comparable prehistoric settlement densities for the volcanic regions. Ongoing analysis and discussion of the multitude of resulting historico-cultural meanings will be given in more extensive consecutive publications.

B. Discipline's Workflow: Frame Conversion Benchmark

The number of parallel cores used for the making of individual frames can be efficiently controlled. The parallel processing itself does not depend on OpenMP. Table IV gives the dimensions of canvas sizes for an excerpt of common formats, represented by pixel (px) scales.

Table IV. Canvas Sizes and Formats Used in Practical Case Scenario Implementations (Table II, excerpt).

Canvas Size (px)	*Format*	
Format 16:9 (e.g., 24×13.5 cm)		
7680 × 4320	UHD-2	8 k
3840 × 2160	UHD	4 k
1920 × 1080	HD	
Format 4:3 (e.g., 24×18 cm)		
1600 × 1200	UXGA	
1400 × 1050	SXGA+	
1024 × 768	XGA	

The given formats High Definition (HD), Ultra High Definition (UHD), Ultra Extended Graphics Array (UXGA),

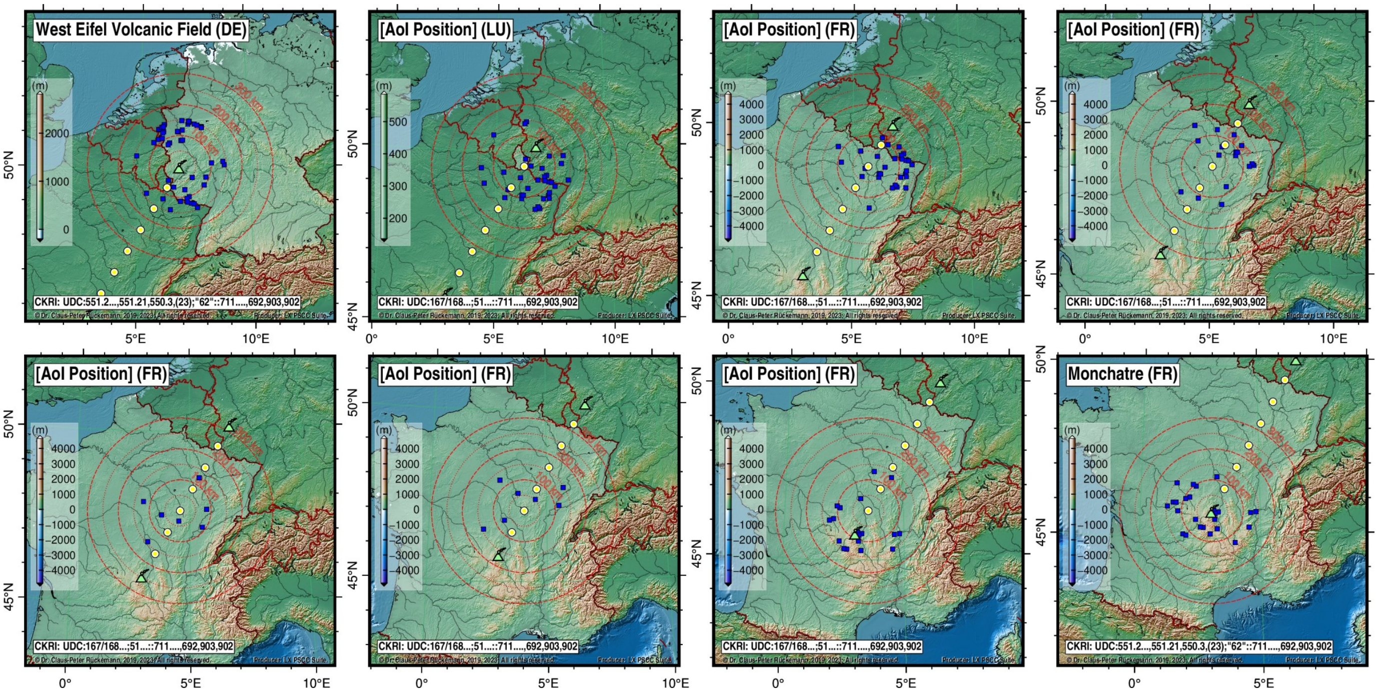

Figure 1. Discipline workflow results of prehistoric settlement infrastructures in factual and historico-cultural chorological and chronological contexts with a volcanological features group (maars, Holocene-historical) and satellite data based on the coherent conceptual knowledge integration and discovery (excerpt).

Extended Graphics Array (XGA), and Super XGA Plus (SXGA+) are commonly used in resources development and practical high resolution workflows.

The conversion of frames can be done in parallel using GraphicsMagick [58]. GraphicsMagick includes Gnu's Not Unix (GNU) libgomp [66] of the GNU Offloading and Multi-Processing Project (GOMP). All benchmarks were done on a recent Linux multi-core system. Table V shows the frame conversion benchmark results for different canvas sizes as used in the parallel implementations of practical case scenarios.

The results compare number of threads, iterations, user time, total time, iterations per second, iterations per CPU, speedup, and Karp-Flatt result. The conversion uses a common 128×128 granite texture pattern (px) iteration for standardisation. The benchmark uses the Karp-Flatt metric [67], which is a measure of code parallelisation in parallel processor systems [68].

The resulting implementation is very scalable and can use practical workflow parallelisations from small canvas sizes up to defined sizes even beyond UHD-2. Sizes of UHD are very appropriate for many HR scenarios with commonly available technical infrastructures while being relatively efficient with resources.

V. Conclusion

This extended paper presented the results achieved for the methodological discovery and parallelisation of workflow logic of contextualisation in prehistorical archaeology targeting historico-cultural contexts and based on coherent multi-disciplinary conceptual knowledge. The implemented workflows employed the latest versions of the prehistory-protohistory and archaeology CKRI and the CRI framework. In the center of this research, contextualisation is the target. Contextualisation is not an option. Contextualisation can be dealt with to any depth and width. Therefore, a realisation can take days and even decades of work with ongoing development, which is considered a major benefit for serious knowledge based work. The implemented and realised contextualisation workflows proved efficient, flexible, and sustainable. The presented contexts, entities, and workflow implementations provide solid fact-based fundaments for contextualisation and consequent fact-based historico-cultural interpretation, procedures, which should be deployed by members of the contributing disciplines.

Examples of further realisation are in-time integration of dynamically generated context instances, weather contexts, earthquakes, tsunamis, and wild fires. Full updates can, for example, be shown in about 15–30 minutes per frame view. First response version with reduced context data can be generated in seconds.

Ongoing, the reference implementations and procedures will be extended for generation of symbolic representation for advanced multi-dimensional knowledge models. Future research will address the creation and further development of the prehistory-protohistory and archaeology CKRI and the CRI framework for coherent multi-disciplinary conceptual contextualisation, enabling multi-disciplinary equal footing with contributions from all scientific disciplines, e.g., natural sciences, soil science, and linguistics, especially supporting new advanced methods in prehistorical archaeology for knowledge integration, contextualisation, and analysis.

TABLE V. FRAME CONVERSION BENCHMARK RESULTS FOR CANVAS SIZES USED IN PARALLEL IMPLEMENTATIONS OF PRACTICAL MULTI-DISCIPLINARY CASE SCENARIO WORKFLOWS OF ARCHAEOLOGICAL / PREHISTORICAL CONTEXTUALISATION (TABLE II, EXCERPT).

Threads	*Iterations*	*User Time*	*Elapsed Time*	*Iterations / s*	*Iterations / CPU*	*Speedup*	*Karp-Flatt*
7680 px × 4320 px (UHD-2)							
1	2	10.56 s	10.563899 s	0.189	0.189	1.00	1.000
4	8	42.24 s	11.161860 s	0.717	0.189	3.79	0.019
8	14	74.15 s	10.476729 s	1.336	0.189	7.06	0.019
12	19	100.79 s	10.110634 s	1.879	0.189	9.93	0.019
16	24	127.42 s	10.179369 s	2.358	0.188	12.45	0.019
18	26	138.39 s	10.094287 s	2.576	0.188	13.60	0.019
20	28	149.43 s	10.087584 s	2.776	0.187	14.66	0.019
24	32	170.80 s	10.140287 s	3.156	0.187	16.67	0.019
28	35	187.10 s	10.011514 s	3.496	0.187	18.47	0.019
32	39	208.90 s	10.251268 s	3.804	0.187	20.09	0.019
36	41	220.25 s	10.067891 s	4.072	0.186	21.51	0.019
3840 px × 2160 px (UHD)							
1	8	10.62 s	10.625725 s	0.753	0.753	1.00	1.000
4	29	38.55 s	10.156136 s	2.855	0.752	3.79	0.018
8	54	72.25 s	10.110667 s	5.341	0.747	7.09	0.018
12	76	102.13 s	10.066204 s	7.550	0.744	10.03	0.018
16	94	128.20 s	10.001879 s	9.398	0.733	12.48	0.019
18	104	141.39 s	10.075150 s	10.322	0.736	13.71	0.018
20	113	153.91 s	10.070419 s	11.221	0.734	14.90	0.018
24	128	176.02 s	10.071374 s	12.709	0.727	16.88	0.018
28	141	197.50 s	10.059111 s	14.017	0.714	18.62	0.019
32	155	216.18 s	10.010670 s	15.483	0.717	20.57	0.018
36	166	233.24 s	10.056526 s	16.507	0.712	21.92	0.018
1024 px × 768 px (XGA)							
1	82	10.08 s	10.078310 s	8.136	8.135	1.00	1.000
4	313	40.03 s	10.009561 s	31.270	7.819	3.84	0.014
8	601	79.95 s	10.007529 s	60.055	7.517	7.38	0.012
12	864	119.93 s	10.000504 s	86.396	7.204	10.62	0.012
16	1079	159.91 s	10.006024 s	107.835	6.748	13.25	0.014
18	1191	179.99 s	10.007169 s	119.015	6.617	14.63	0.014
20	1274	199.89 s	10.003267 s	127.358	6.374	15.65	0.015
24	1517	239.81 s	10.002562 s	151.661	6.326	18.64	0.013
28	1674	279.97 s	10.005288 s	167.312	5.979	20.56	0.013
32	1788	319.93 s	10.001453 s	178.774	5.589	21.97	0.015
36	1856	358.52 s	10.001333 s	185.575	5.177	22.81	0.017

ACKNOWLEDGEMENTS

This paper was deeply gratefully completed in memoriam to our most esteemed friend and scientific fellow colleague, Dipl.-Biol. Birgit Frida Stefanie Gersbeck ((Gersbeck-Schierholz)) (Aug. 14., 1960 – Feb. 2, 2024), who directly contributed to this long-term research. Birgit was a member of the permanent commission and the board of trustees, DIMF, IARIA board member, IARIA Fellow, scientific peer reviewer, and former scientific member of the Leibniz Universität Hannover, Germany. Birgit had dedicated a large part of her scientific career to multi-disciplinary research and providing service to the international scientific community for more than thirty years, having actively contributed to research, results, and success of scientific events and collaborations. Birgit will be unforgotten in our hearts and minds forever!

This ongoing research is supported by scientific organisations and individuals. We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), for partially funding this research, implementation, case studies, and publication under grants D2022F1P05308, D2020F1P05228, and D2022F2P05355 and to its senior scientific members and members of the permanent commission of the science council, especially to Dr. Friedrich Hülsmann, DIMF, for fruitful discussion, inspiration, and practical multi-disciplinary contextualisation and case studies. We are grateful to Dipl.-Geogr. Burkhard Hentzschel and Dipl.-Ing. Eckhard Dunkhorst, Minden, Germany, for prolific discussion and exchange of practical spatial, Unmanned Aerial Vehicles (UAV), context scenarios, and measurement support. We are grateful to Dipl.-Ing. Hans-Günther Müller, Göttingen, Germany, for providing specialised, manufactured high end computation, storage, and visualisation solutions. We are grateful to The Science and High Performance Supercomputing Centre (SHPSC) for long-term support. / DIMF-PIID-DF98_007; URL: https://scienceparagon.de/cpr .

References

[1] C.-P. Rückemann, "Prehistorical Archaeology Discipline's Contextualisation Facts and Workflow Logic: Complements-Components Blueprints for the Creation of Efficient Coherent Multi-disciplinary Conceptual Knowledge-based Discovery," in Proceedings of The Thirteenth International Conference on Advanced Communications and Computation (INFOCOMP 2023), June 26 – 30, 2023, Nice, France. XPS Press, Wilmington, Delaware, USA, 2023, pp. 7–12, ISSN: 2308-3484, ISBN: 978-1-68558-073-5, URL: https://www.thinkmind.org/index.php?view=article&articleid=infocomp_2023_1_20_60015 [accessed: 2024-05-26].

[2] C.-P. Rückemann, "Coherent Knowledge Structures and Fusion Practice for Contextualisation Insight in Prehistory and Protohistory," in The Eleventh Symposium on Advanced Computation and Information in Natural and Applied Sciences, Proceedings of The 19th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 20–26, 2021, Rhodes, Greece, Proceedings of the American Institute of Physics (AIP), AIP Conference Proceedings 2849, Volume 2849, no. 1. AIP Press, American Institute of Physics, Melville, New York, USA, 2023, pp. 040 001 – 040 004, ISBN: 978-0-7354-4589-5, ISSN: 0094-243X, DOI: 10.1063/5.0162058.

[3] T. Sorensen and E. Joseph, "Recent HPC-centric AI Success Stories," Hyperion Research, Special Report, #HR12.0027.05.02.2023, May 2023, URL: https://hyperionresearch.com/wp-content/uploads/2023/05/Hyperion-Research-HPC-AI-Success-Story-Report-May-2023.pdf [accessed: 2024-05-26] .

[4] C.-P. Rückemann, "Faceting the Holocene-prehistoric Inventory of Volcanological Features Groups Towards Sustainable Multi-disciplinary Context Integration in Prehistory and Archaeology Based on the Methodology of Coherent Conceptual Knowledge Contextualisation," International Journal on Advances in Intelligent Systems, vol. 15, no. 3&4, 2022, pp. 115–129, ISSN: 1942-2679, LCCN: 2008212456 (Library of Congress), URL: http://www.iariajournals.org/intelligent_systems [accessed: 2024-05-26].

[5] C.-P. Rückemann, "Advanced Contextualisation Reference Implementation Frameworks in Practice: Coherent Multi-disciplinary Conceptual Knowledge-Spatial Context Discovery Results from the Holocene-prehistoric Volcanological Features and Archaeological Settlement Infrastructure Surveys," in Proceedings of The Fifteenth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2023), April 24 – 28, 2023, Venice, Italy. XPS Press, Wilmington, Delaware, USA, 2023, pp. 98–103, ISSN: 2308-393X, ISBN-13: 978-1-68558-079-7, URL: https://www.thinkmind.org/articles/geoprocessing_2023_2_150_30100.pdf [accessed: 2024-05-26].

[6] "The Prehistory and Archaeology Knowledge Archive (PAKA) Licence," 2024, (release 2024), Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[7] C.-P. Rückemann, "Towards Conceptual Knowledge Reference Implementations for Context Integration and Contextualisation of Prehistory's and Natural Sciences' Multi-disciplinary Contexts," International Journal on Advances in Systems and Measurements, vol. 14, no. 1&2, 2021, pp. 113–124, ISSN: 1942-261x, LCCN: 2008212470 (Library of Congress), URL: http://www.iariajournals.org/systems_and_measurements [accessed: 2024-05-26].

[8] C.-P. Rückemann, "Component Framework Implementation and Realisation for Development and Deployment of a Coherent Multi-disciplinary Conceptual Knowledge-based Holocene-prehistoric Inventory of Volcanological Features Groups and Faceting," International Journal on Advances in Intelligent Systems, vol. 15, no. 3&4, 2022, pp. 103–114, ISSN: 1942-2679, LCCN: 2008212456 (Library of Congress), URL: http://www.iariajournals.org/intelligent_systems [accessed: 2024-05-26].

[9] C.-P. Rückemann, "Towards a Component Reference Implementations Frame for Achieving Multi-disciplinary Coherent Conceptual and Chorological Contextualisation in Prehistory and Prehistoric Archaeology," International Journal on Advances in Systems and Measurements, vol. 14, no. 1&2, 2021, pp. 103–112, ISSN: 1942-261x, LCCN: 2008212470 (Library of Congress), URL: http://www.iariajournals.org/systems_and_measurements [accessed: 2024-05-26].

[10] C.-P. Rückemann, "The Information Science Paragon: Allow Knowledge to Prevail, from Prehistory to Future – Approaches to Universality, Consistency, and Long-term Sustainability," The International Journal "Information Models and Analyses" (IJ IMA), vol. 9, no. 3, 2020, pp. 203–226, Markov, K. (ed.), ISSN: 1314-6416 (print), Submitted accepted article: November 18, 2020, Publication date: August 17, 2021, URL: http://www.foibg.com/ijima/vol09/ijima09-03-p01.pdf [accessed: 2024-05-26].

[11] C.-P. Rückemann, "Coherent Knowledge Solutions From Prehistory to Future – Towards Coherent Multi-disciplinary Knowledge Reference Implementation Blueprints for Industrial Learning: Insight from Consistent Coherent Conceptual Integration of Prehistory, Archaeology, Natural Sciences, and Humanities," ML4I – Machine Learning for Industry Forum 2021; High-Performance Computing Innovation Center (HPCIC) and Data Science Institute (DSI), at Lawrence Livermore National Laboratory (LLNL), Aug. 10–12, 2021, Livermore, U.S.A., (Invited Speech), URL: http://www.llnl.gov [accessed: 2024-05-26], 2021.

[12] C.-P. Rückemann, "From Knowledge and Meaning Towards Knowledge Pattern Matching: Processing and Developing Knowledge Objects Targeting Geoscientific Context and Georeferencing," in Proc. GEOProcessing 2020, November 21–25, 2020, Valencia, Spain, 2020, pp. 36–41, ISSN: 2308-393X, ISBN-13: 978-1-61208-762-7.

[13] R. Gleser, Zu den erkenntnistheoretischen Grundlagen der Prähistorischen Archäologie. Leiden, 2023, (title in English: On the Epistemological Fundaments of Prehistorical Archaeology), in: M. Renger, S.-M. Rothermund, S. Schreiber, and A. Veling (Eds.), Theorie, Archäologie, Reflexion. Kontroversen und Ansätze im deutschsprachigen Diskurs, (in print).

[14] C.-P. Rückemann, "Information Science and Inter-disciplinary Long-term Strategies – Key to Insight, Consistency, and Sustainability: Conceptual Knowledge Reference Methodology Spanning Prehistory, Archaeology, Natural Sciences, and Humanities," International Tutorial, DataSys Congress 2020, Sept. 27 – Oct. 1, 2020, Lisbon, Portugal, 2020, URL: http://www.iaria.org/conferences2020/ProgramINFOCOMP20.html [accessed: 2024-05-26].

[15] "Multilingual Universal Decimal Classification Summary," 2012, UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), URL: http://www.udcc.org/udcsummary/php/index.php [accessed: 2024-05-26].

[16] "Creative Commons Attribution Share Alike 3.0 license," 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2024-05-26], (first release 2009, subsequent update 2012).

[17] L. W. Anderson and D. R. Krathwohl, Eds., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon, Boston, MA (Pearson Education Group), USA, 2001, ISBN: 978-0801319037.

[18] C.-P. Rückemann, F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski, Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, Sept. 23, 2015, The Fifth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 13th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–29, 2015, Rhodes, Greece, 2015, pp. 1–7, DOI: 10.15488/3409.

[19] Plato, Phaedo, 2008, (Written 360 B.C.E.), Translated by Benjamin Jowett, Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Plato/phaedo.html [accessed: 2024-05-26].

[20] Aristotle, Categories, 2008, (Written 350 B.C.E.), Translated by E. M. Edghill Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/categories.html [accessed: 2024-05-26].

[21] Aristotle, On Interpretation, 2008, Peri Hermeneias, (Written 350 B.C.E.), Translated by E. M. Edghill Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/interpretation.html [accessed: 2024-05-26].

[22] Aristotle, Prior Analytics, 2008, (Written 350 B.C.E.), Translated by A. J. Jenkinson Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/prior.html [accessed: 2024-05-26].

[23] Aristotle, Posterior Analytics, 2008, (Written 350 B.C.E.), Translated by G. R. G. Mure Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/posterior.html [accessed: 2024-05-26].

[24] Aristotle, Topics, 2008, (Written 350 B.C.E.), Translated by W. A. Pickard-Cambridge, Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/topics.mb.txt [accessed: 2024-05-26].

[25] Aristotle, On Sophistical Refutations, 2008, (Written 350 B.C.E.), Translated by W. A. Pickard-Cambridge Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/sophist_refut.html [accessed: 2024-05-26].

[26] B. Bennett, Logically Fallacious: The Ultimate Collection of Over 300 Logical Fallacies, 1st ed. eBookIt, 2012, ISBN: 978-1-4566-0737-1.

[27] M. Broussard, Artificial Unintelligence, How Computers Misunderstand the World. The MIT Press, Cambridge - Massachusetts, London - England, 2018, ISBN: 978-0-262-03800-3.

[28] C. L. Olson, J. J. Becker, and D. T. Sandwell, "SRTM15_PLUS: Data fusion of Shuttle Radar Topography Mission (SRTM) land topography with measured and estimated seafloor topography," (NCEI Accession 0150537), National Centers for Environmental Information (NCEI), NOAA, 2016.

[29] B. Tozer, D. T. Sandwell, W. H. F. Smith, C. Olson, J. R. Beale, and P. Wessel, "Global Bathymetry and Topography at 15 Arc Sec: SRTM15+," Earth and Space Science, vol. 6, no. 10, Oct. 2019, pp. 1847–1864, ISSN: 2333-5084, DOI: 10.1029/2019EA000658.

[30] "Network Common Data Form (NetCDF)," 2024, DOI: 10.5065/D6H70CW6, URL: http://www.unidata.ucar.edu/software/netcdf/ [accessed: 2024-05-26].

[31] "University Corporation for Atmospheric Research (UCAR)," 2024, URL: https://ncar.ucar.edu/ [accessed: 2024-05-26].

[32] "National Center for Atmospheric Research (NCAR)," 2024, URL: https://www.ucar.edu/ [accessed: 2024-05-26].

[33] P. Wessel, "DCW for GMT 6 or later," 2022, URL: http://www.soest.hawaii.edu/pwessel/dcw/ [accessed: 2024-05-26].

[34] P. Wessel, "GSHHG," 2017, URL: http://www.soest.hawaii.edu/pwessel/gshhg/ [accessed: 2024-05-26].

[35] C.-P. Rückemann and B. F. S. Gersbeck-Schierholz, "Walking With Gods, Scholars, and Caesars in Rodini Park," KiMrise, Knowledge in Motion, September 27, 2013, Circle Summit Workgroup Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Rodini Park, Rhodos, Greece, 2013.

[36] B. F. S. Gersbeck-Schierholz and C.-P. Rückemann, "Ageless Elements of Nature - Environments of the Immortals: Rodini Park," KiMrise, Knowledge in Motion, September 27, 2013, Circle Summit Workgroup Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Rodini Park, Rhodos, Greece, 2013, URL: https://scienceparagon.de/bgs/bgs_rhodes13_imm.html [accessed: 2024-05-26], URL: https://beckilabs.de/bgs_rhodes13_imm.html [accessed: 2024-05-26], URL: http://www.user.uni-hannover.de/zzzzgers/bgs_rhodes13_imm.html [accessed: 2021-04-30] (Orig.-URL, orphaned link), (Media Presentation).

[37] B. F. S. Gersbeck-Schierholz and C.-P. Rückemann, "Nature in the Light of Gods," KiMrise, Knowledge in Motion, September 27, 2015, Circle Summit Workgroup Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Rodini Park, Rhodos, Greece, 2015, URL: https://scienceparagon.de/bgs/bgs_rhyta.html [accessed: 2024-05-26], URL: https://beckilabs.de/bgs_rhodes15_nature.html [accessed: 2024-05-26], URL: http://www.user.uni-hannover.de/zzzzgers/bgs_rhodes15_nature.html [accessed: 2021-04-30] (Orig.-URL, orphaned link), (Media Presentation).

[38] B. F. S. Gersbeck-Schierholz and C.-P. Rückemann, "The Plethora of Rhyta: Art and Use," KiM On-site Summit, Knowledge in Motion, February 25, 2019, On-site Summit Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", National Archaeological Museum, Athens, Greece, 2019, URL: https://scienceparagon.de/bgs/bgs_rhyta.html [accessed: 2024-05-26], URL: https://beckilabs.de/bgs_rhyta.html [accessed: 2024-05-26], URL: http://www.user.uni-hannover.de/zzzzgers/bgs_rhyta.html [accessed: 2021-04-30] (Orig.-URL, orphaned link), (Media Presentation).

[39] B. F. S. Gersbeck-Schierholz and C.-P. Rückemann, "Life in Sulfuric Environment," KiM On-site Summit, Knowledge in Motion, September 24, 2019, On-site Summit Meeting, "Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF)", Nisyros, Greece, 2019, URL: https://scienceparagon.de/bgs/bgs_nisyros.html [accessed: 2024-05-26], URL: https://beckilabs.de/bgs_nisyros.html [accessed: 2024-05-26], URL: http://www.user.uni-hannover.de/zzzzgers/bgs_nisyros.html [accessed: 2021-04-30] (Orig.-URL, orphaned link), (Media Presentation).

[40] C.-P. Rückemann, "Methodology and Integrated Knowledge for Complex Knowledge Mining: Natural Sciences and Archaeology Case Study Results," in Proc. of The 9th Int. Conf. on Adv. Geo. Inf. Syst., Appl., and Serv. (GEOProcessing 2017), March 19–23, 2017, Nice, France. XPS, 2017, pages 103–109, ISSN: 2308-393X, ISBN: 978-1-61208-539-5, URL: http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2017_7_10_30036 [accessed: 2024-05-26].

[41] C.-P. Rückemann, "Best Practice 'Golden Spike' Result: Radiocarbon Dating Practice," in Proceedings of The 20th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19–25, 2022, Heraklion, Crete, Greece . AIP Press, American Institute of Physics, Melville, New York, USA, Apr. 2024, pp. 030 001–1, in: [69], New Methods, Results, and Insights Achieved at The Twelfth Symposium on Advanced Computation and Information in Natural and Applied Sciences, Simos, T. E., G., Tsitouras, C. (eds.), ISBN-13: 978-9925-7552-3-3, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), URL: http://proceedings.aip.org/ [accessed: 2024-05-26], (to appear).

[42] C.-P. Rückemann, "From Knowledge and Meaning Towards Knowledge Pattern Matching: Processing and Developing Knowledge Objects Targeting Geoscientific Context and Georeferencing," in Proceedings of The Twelfth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2020), (FIRST DATES: March 22–26, 2020, Valencia, Spain), (SECOND DATES: June 22–25, 2020, Valencia, Spain), November 21–25, 2020, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2020, rückemann, C.-P. and Doytsher, Y. and Ritz, T. (eds.), pages 16–21, 3rd edition, December 12, 2020, ISSN: 2308-393X, ISBN-13: 978-1-61208-762-7, ISBN-13: 978-1-61208-098-7 (Electronic Proceedings), URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2020_1_30_30035 [accessed: 2024-05-26], URL: http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2020_1_30_30035 [accessed: 2024-05-26].

[43] C.-P. Rückemann, "The Abydos Ships Method: New-insight-driven Numeral Systems Approach for Coherent Multi-disciplinary Analysis and Contextualisation," in The Twelfth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), Proceedings of The 20th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19–25, 2022, Heraklion, Crete, Greece. AIP Press, American Institute of Physics, Melville, New York, USA, 2024, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), URL: http://icnaam.org/ [accessed: 2024-05-26], (to appear).

[44] C.-P. Rückemann and B. F. S. Gersbeck-Schierholz, "Object Security and Verification for Integrated Information and Computing Systems," in Proceedings of the Fifth International Conference on Digital Society (ICDS 2011), Proceedings of the International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS 2011), February 23–28, 2011, Gosier, Guadeloupe, France / DigitalWorld 2011. XPS, 2011, pp. 1–6, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=cyberlaws_2011_1_10_70008 [accessed: 2024-05-26].

[45] "LX Professional Scientific Content-Context-Suite (LX PSCC Suite) Licence," 2024, (release 2024), Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[46] P. Wessel, W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe, "The Generic Mapping Tools (GMT)," 2020, URL: http://www.generic-mapping-tools.org/ [accessed: 2024-05-26].

[47] "GMT - Generic Mapping Tools," 2024, URL: http://www.generic-mapping-tools.org/ [accessed: 2024-05-26].

[48] "The LX Professional Scientific Creation Academic Suite (LX PSCA Suite) Licence," 2024, (release 2024), Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[49] The TEX Live Team, TEX Live, 2023, URL: https://tug.org/texlive/ [accessed: 2024-05-26].

[50] D. Carlisle, "The TeXLive.net Server," 2023, (The TeXLive.net server was written by David Carlisle and hosted on a virtual machine provided by Stefan Kottwitz, with financial support from DANTE).

[51] "MIT License," 2024, SPDX License List, spdx.org, SPDX Working Group, URL: https://spdx.org/licenses/MIT.html [accessed: 2024-05-26].

[52] PROJ contributors, PROJ Coordinate Transformation Software Library, Open Source Geospatial Foundation, 2024, URL: https://proj.org/ [accessed: 2024-05-26].

[53] "Perl Compatible Regular Expressions (PCRE)," 2024, URL: https://www.pcre.org/ [accessed: 2024-05-26].

[54] "Google," 2024, URL: https://google.com/ [accessed: 2024-05-26].

[55] "OGC Approves KML as Open Standard," 2008, OpenGeospatial.org, 2008-04-14, Press Release 857, URL: http://www.opengeospatial.org/pressroom/pressreleases/857 [accessed: 2024-05-26].

[56] "FFmpeg," 2024, URL: https://ffmpeg.org/ [accessed: 2024-05-26].

[57] S. Tomar, "Converting video formats with FFmpeg," Linux Journal, vol. 2006, no. 146, 2006, p. 10.

[58] "GraphicsMagick Image Processing System," 2024, URL: http://www.graphicsmagick.org/ [accessed: 2024-05-26].

[59] "GNU Image Manipulation Program," 2024, URL: http://www.gimp.org/ [accessed: 2024-05-26].

[60] VideoLan, "VLC media player," 2024, URL: https://www.videolan.org/vlc/index.html [accessed: 2024-05-26].

[61] "Parallel BZIP2 (PBZIP2)," 2024, URL: http://compression.ca/pbzip2/ [accessed: 2024-05-26].

[62] L. Dagum and R. Menon, "OpenMP: an industry standard API for shared-memory programming," Computational Science & Engineering, (IEEE), vol. 5, no. 1, 1998, pp. 46–55.

[63] OpenMP Architecture Review Board, "OpenMP API 5.1 Specification," Nov. 2020, URL: https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-1.pdf [accessed: 2024-05-26].

[64] "GNU Parallel," 2024, URL: http://www.gnu.org/s/parallel [accessed: 2024-05-26].

[65] "The Archaeological Data Collector (ADC) Licence," 2024, (release 2024), Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[66] "GNU libgomp – GNU Offloading and Multi-Processing Project (GOMP)," 2024, URL: https://gcc.gnu.org/projects/gomp/ [accessed: 2024-05-26].

[67] A. H. Karp and H. P. Flatt, "Measuring Parallel Processor Performance," in Communications of the ACM, 33, vol. 5, 1990, pp. 539–543, DOI: 10.1145/78607.78614.

[68] M. J. Quinn, Parallel Programming in C with MPI and OpenMP. McGraw-Hill, 2017, ISBN: 978-0-07-058201-9.

[69] C.-P. Rückemann, "New Methods, Results, and Insights Achieved at The Twelfth Symposium on Advanced Computation and Information in Natural and Applied Sciences," in Proceedings of The 20th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19–25, 2022, Heraklion, Crete, Greece, Proceedings of the American Institute of Physics (AIP) . AIP Press, American Institute of Physics, Melville, New York, USA, Apr. 2024, pp. 030 001–1, ISBN-13: 978-9925-7552-3-3, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), URL: http://proceedings.aip.org/ [accessed: 2024-05-26], (to appear).

An Empirical Study of Machine Learning for Course Failure Prediction: A Case Study in Numerical Methods

Isaac Caicedo-Castro
Socrates Research Team
Research Team: Development, Education, and Healthcare
Faculty of Engineering
University of Córdoba
Carrera 6 No. 76-103, 230002, Montería, Colombia
ORCID: 0000-0002-7567-3774
e-mail: isacaic@correo.unicordoba.edu.co

***Abstract*—In this paper, we address the problem of predicting whether a student might fail a course before it starts, based on their academic history. This study is centered on predicting failure in the numerical methods course, which is part of the curriculum for the bachelor's degree in systems engineering at the University of Córdoba in Colombia. To tackle this problem, we adopt classification methods from supervised machine learning. To this end, we utilize a dataset initially collected in [1] and subsequently expanded in [2]. This dataset is used to fit and validate the machine learning methods employed in this study. Our work contributes to improving the quality of the forecasting task compared to prior research [1], [2]. This improvement has been achieved by modifying the vector representation of the student's academic history, considering only the student's performance in mathematics, as evidenced in the admission test called Saber 11 and prerequisite courses. The results of the experimental validation reveal that the method based on Gaussian processes with the Radial Basis Function achieves mean values of accuracy, precision, recall, and harmonic mean of 83%, 80.67%, 77%, and 76.70%, respectively. This method has outperformed the others studied in this work. Moreover, the prediction outcome of Gaussian processes is the probability that a given student will fail the course, which is convenient for designing an intervention plan to help them succeed. Therefore, the conclusion of this study is twofold. Firstly, Gaussian processes are the best choice to implement an intelligent system for the prediction task studied herein. Secondly, this study finds a clear correlation between the probability of succeeding in the numerical methods course and the student's competencies in mathematics obtained before enrolling in this course. This suggests that good training in mathematics courses is required to succeed in the numerical methods course.**

***Index Terms*—Machine learning; educational data mining; classification algorithm.**

I. Introduction

The aim of this study is to leverage machine learning accuracy to design an intelligent system for predicting student failure in the numerical methods course before it even starts. The input variables are derived from the student's performance in prerequisite courses that are assumed necessary for succeeding in the numerical methods course within the bachelor's degree program in systems engineering at the University of Córdoba, Colombia.

Herein and in several literature references, failing a course is often referred to as either dropping out or not passing the course successfully. Identifying students at risk of failing a specific course is crucial, as it enables stakeholders such as lecturers, students, academic policymakers, and others to take necessary precautions to prevent failure. This proactive approach helps students avoid psychological stress, frustration, and financial loss.

Prior research has explored various machine learning approaches to identify students at risk of course failure. These methods include classification methods, such as artificial neural networks or multilayer perceptron [1]–[7], support vector machines [1]–[4], [8], quantum-enhanced support vector machine [9], logistic regression [1], [4], [8], [10], decision trees [1], [2], [4], [8], [10], [11], ensemble methods with different classification methods [3], [6], random forest [1], [2], [4]–[6], gradient boosting [5], extreme gradient boosting (XGBoost) [1], [2], [5], [6], variants of gradient boosting [5], [8], such as CatBoost [12] and LightGBM [13], and Gaussian processes for classification [1], [2].

Much of the previously referenced literature has centered on online courses [3], [4], [6], [7], [10], covering various topics such as computer networking and web design [3], mathematics [10], and STEM (science, technology, engineering, and mathematics) in general [7]. It is noteworthy that the primary goal of these research endeavors is not to predict the risk of failure before students begin their courses; instead, they are focused on forecasting risk during the course development phase. This forecasting relies on students' activities, including the number of course views, content downloads, and grades achieved in assignments, tests, quizzes, projects, and other assessments.

The objectives pursued in [1], [2], [9] align with the goals of this study, operating within the same context of predicting student failure in the numerical methods course based on performance in prerequisite courses within the undergraduate program in systems engineering at the University of Córdoba in Colombia. However, it is worth noting that [1] utilized a smaller dataset compared to the one employed in this study, which corresponds to the dataset collected in [2] and used in [9].

In [1], the student's performance in prerequisite courses, including mathematics, physics, and computer programming, as well as their outcomes in the admission test, are used as input variables for predicting whether they might fail the numerical method course. In contrast, in [2], it is observed that excluding the student's outcomes in the admission test leads to increased accuracy, precision, and recall in the prediction. Furthermore, in [9], quantum machine learning is adopted; however, this approach does not outperforms the performance of Gaussian processes for classification as utilized in [2].

A 10-Fold Cross-Validation conducted in [2] indicates that the Gaussian process with the Matern kernel achieves mean values of accuracy, precision, recall, and harmonic mean of 80.45%, 83.33%, 66.5%, and 72.52%, respectively. In contrast, our study has found far better results in terms of accuracy and harmonic mean by reducing the input variables to only include the student's performance in the prerequisite mathematics course and the outcomes of the admission test in the same subject. With this input configuration, the Gaussian process with the Radial Basis Function kernel yields mean values of accuracy, precision, recall, and harmonic mean of 83%, 80.67%, 77%, and 76.79%, respectively.

The remainder of this paper is organized as follows: Section II formalizes the problem, presents key assumptions, outlines the representation of the input space, and introduces the target variable. This section also presents the machine learning methods adopted in this study, along with the validation method. Section III provides details on the dataset features, programming language, software library, and computing environment used during the validation process. In Section IV, the results of the experiments are presented, followed by a discussion in Section V. Finally, Section VI concludes the discussion of the results, highlights the novelty of this study, and suggests directions for future research.

II. Methods

The problem addressed in this study is to identify regular patterns between failing the numerical methods course and student's performance in prerequisite courses, as well as their performance in the admission test. To cope with this problem, a quantitative approach is adopted. This approach involves quantifying student's performance through their grades in prerequisite courses, their scores on the admission test, and considering the frequency of course enrollments due to previous failures.

In this study, we use a dataset initially collected in [14], which was subsequently expanded with additional instances in [2] and further utilized in [9]. The dataset describes input variables for each prerequisite course as follows: the variables $x_{i,j}$ and $x_{i,j+2}$ represent the highest and lowest final grades obtained by the ith student in the jth course, respectively, while $x_{i,j+1}$ denotes the number of semesters the ith student has enrolled in the jth course. If the ith student does not fail the jth course upon the first enrollment, $x_{i,j}$ and $x_{i,j+2}$ will have the same value, and $x_{i,j+1}$ will be equal to one. Additionally, in this study, $x_{i,1}$ represents the score achieved by the ith student in the subject of mathematics in the admission test.

This dataset has been collected through a survey conducted among students of systems engineering at the University of Córdoba in Colombia. To safeguard the privacy of the participants, the dataset has been anonymized, retaining only students' grades and admission scores. Personal information, including identification numbers, names, gender, and economic stratum, has been omitted.

In this study, it is assumed that only mathematics courses are prerequisite for success in the numerical methods course. Therefore, the prerequisite courses considered are linear algebra, calculus I, II, and III. Conversely, in previous studies [2], [9], [14], physics and computer programming courses are also considered prerequisites, although the outcomes of the admission test are not utilized in [2], [9], while all admission test outcomes are used in [14]. As a result, the input space in our study has 13 dimensions (three per prerequisite course and the mathematics subject score of the admission test), compared to 33 dimensions in [2], [9], and 38 dimensions in [14].

Mathematically, the ith student is represented by a D-dimensional vector $\mathbf{x}_i \in \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^D$ and $D = 13$, accounting for four prerequisite courses and one additional dimension for the mathematics subject score in the admission test. Each component of the vector $\mathbf{x}_i$ corresponds to a specific input variable, detailed as follows:

- x_{i1}: The score obtained by the student in the admission test for mathematics, ranging from 0 to 100 ($x_{i1} \in \mathbb{Z}$, $0 \leq x_{i1} \leq 100$).
- x_{i2} to $x_{i,13}$: Grades and enrollment information for the prerequisite courses, with x_{ij}, $x_{i,j+2}$, and $x_{i,j+1}$ representing the highest final grade, lowest final grade, and number of semesters enrolled for each course j, respectively. Here, j takes values of 2, 5, 8, and 11, corresponding to calculus I, II, III, and linear algebra courses.
- In Colombian universities, students are graded in the range from 0 up to 5 (see student's code of the University of Córdoba [15]), i.e., $x_{ij} \in \mathbb{R}$ and $0 \leq x_{ij} \leq 5$ for $j = 2, 4, 5, 7, 8, 10, 11, 13$.
- The number of semesters enrolled for each course j is a natural number or zero if a the ith student has never enrolled it, i.e., $x_{ij} \in \mathbb{N} \cup \{0\}$ for $j = 3, 6, 9, 12$.

On the other hand, y_i represents the target variable associated with the ith student, where $y_i = 1$ if the student failed the numerical methods course, and $y_i = 0$ otherwise (i.e., $y_i \in \{0, 1\}$). Thus, the $\mathcal{D}$ denotes the dataset defined as $\mathcal{D} = \{(x_i, y_i) | x_i \in \mathcal{X}, y_i \in \{0, 1\}, \text{for all } i = 1, \ldots, n\}$, where n represents the size of the dataset.

Given the dataset $\mathcal{D}$ described earlier, the problem in this study is to determine the function g, which maps the input variables to the target variable, i.e., $g : \mathcal{X} \to \{0, 1\}$. Once this function is established, it may be used to predict whether a new student, represented by the vector $x' \in \mathcal{X}$, might fail the numerical methods course. Specifically, if $g(x') = 1$, the function predicts that the student might fail the course; otherwise, if $g(x') = 0$, the prediction is that the student will not fail. This problem falls under supervised learning and is addressed using classification methods.

Before training various classifiers, the dataset undergoes preprocessing. This involves centering each input variable by removing the mean and scaling to unit variance. It is important to note that the dataset contains no missing data, eliminating the need for any imputation methods.

After preparing the dataset, K-fold cross-validation is employed to evaluate each classification method. This process involves splitting the dataset $\mathcal{D}$ into K equal parts, denoted as $\mathcal{D}_i$, where i ranges from 1 to K. During each iteration, one of the K parts is set aside as the validation set, denoted as $\mathcal{V}_k = \mathcal{D}_j$, while the remaining $K-1$ parts are used for training the classifier, represented as $\mathcal{T}_k = \cup_{i=1, i\neq j}^{K} \mathcal{D}_j$. This partitioning is carried out as follows:

$$\begin{aligned} \mathcal{V}_1 &= \mathcal{D}_1, & \mathcal{T}_1 &= \mathcal{D}_2 \cup \mathcal{D}_3 \cup \cdots \cup \mathcal{D}_K \\ \mathcal{V}_2 &= \mathcal{D}_2, & \mathcal{T}_2 &= \mathcal{D}_1 \cup \mathcal{D}_3 \cup \cdots \cup \mathcal{D}_K \\ &\vdots & &\vdots \\ \mathcal{V}_K &= \mathcal{D}_K, & \mathcal{T}_K &= \mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_{K-1} \end{aligned}$$

In this study, various classification methods were validated, including logistic regression. The latter assumes the existence of a hyperplane that separates vectors into two classes in a multidimensional real-valued space. This assumption is reasonable given that the input space is a multidimensional real-valued vector space (i.e., $D = 13$).

Nevertheless, other classification methods more suitable for nonlinear classification problems are also utilized in this study, under the assumption that there might be a more effective input representation in a higher-dimensional space. Probabilistic methods such as Gaussian processes (GPs) have been adopted. Based on Bayesian inference, GPs assume that the probability distribution of the target variable is drawn from a Gaussian or normal distribution, hence the name of the method [16], [17]. The main advantage of this method is its ability to incorporate prior knowledge about the problem, contributing to improved forecasting accuracy, even with a small training dataset, as is the case in the context of this study.

In this study, we used several kernels (a.k.a., covariant functions) with GPs. For instance, the Radial Basis Function kernel, which is defined as follows:

$$k_G(\mathbf{x}_i, \mathbf{x}_j) = \gamma \exp\left(-\frac{||\mathbf{x}_j - \mathbf{x}_i||^2}{2\sigma^2} \right), \tag{1}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^D$ are two D-dimensional vectors in real-valued space, and the hyperparameters $\gamma, \sigma \in \mathbb{R}$ are real numbers that corresponds to the weight and length scale of the kernel, respectively.

In addition, we used the Matern kernel, which is defined as follows:

$$k_M(\mathbf{x}_i, \mathbf{x}_j) = \frac{\gamma(\sqrt{2\nu}||\mathbf{x}_j - \mathbf{x}_i||^2)^\nu}{\Gamma(\nu)2^{\nu-1}\sigma^\nu} K_\nu\left(\frac{\sqrt{2\nu}||\mathbf{x}_j - \mathbf{x}_i||^2}{\sigma} \right), \tag{2}$$

where $K_\nu(\cdot)$ and $\Gamma(\cdot)$ are the modified Bessel function and the gamma function, respectively. The hyperparameter $\nu \in \mathbb{R}$ controls the smoothness of the kernel function.

Moreover, a rational quadratic kernel is utilized, which defined as follows:

$$k_r(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \frac{||\mathbf{x}_j - \mathbf{x}_i||^2}{2\alpha\sigma^2} \right)^{-\alpha}, \tag{3}$$

where σ is used for the same purpose as in (1), while $\alpha \in \mathbb{R}$ is the scale mixture parameter, such that $\alpha > 0$.

Furthermore, Matern kernel and radial basis function are combined by summing both as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \gamma_G k_G(\mathbf{x}_i, \mathbf{x}_j) + \gamma_M k_M(\mathbf{x}_i, \mathbf{x}_j), \tag{4}$$

where γ_G and γ_M are the weights assigned to the kernels.

On the other hand, the classification method based on support vector machines (SVMs) is considered one of the most theoretically motivated and successful in modern machine learning practices ([18], p. 79). SVMs are based on convex optimization, allowing for the identification of a global maximum solution, which is their main advantage. However, SVMs are not well-suited for interpretation in data mining; nevertheless, they excel in training accurate machine learning systems. For a detailed description of this method, refer to [19].

Both SVMs and logistic regression are linear classification methods, operating under the assumption that the input vector space can be separated by a linear decision boundary or, in the case of multidimensional input spaces, by a hyperplane. However, when this assumption is not met, SVMs may be used alongside kernel methods to handle nonlinear decision boundaries (see [19] for further details). In this study, we utilize the radial basis function kernel, similar to the one presented in (1), defined as follows:

$$k_G(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_j - \mathbf{x}_i||^2), \tag{5}$$

where γ controls the radius of this spherical kernel, whose center is $\mathbf{x}_j$. Additionally, polynomial and Sigmoid kernels are used, which defined in (6) and (7), respectively. In (6), $d \in \mathbb{N}$ is the degree of the kernel, and $\gamma \in \mathbb{R}$ is the coefficient in (7).

$$k_p(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d \tag{6}$$

$$k_s(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\langle \mathbf{x}_i, \mathbf{x}_j \rangle) \tag{7}$$

Although support vector machines (SVMs) are considered one of the most successful methods in modern machine learning, multilayer perceptrons (MLPs) and their variants, namely artificial neural networks, have emerged as the most successful in deep learning and big data applications, particularly in tasks such as speech recognition, computer vision, and natural language processing ([20], p. 3). In this study, MLPs are trained using the back-propagation algorithm with cross-entropy error minimization [21], along with the optimization algorithm known as Adam [22]. Specifically, MLPs with one and five hidden layers are adopted in this work.

MLPs offer a significant advantage as they can approximate any function for both classification and regression tasks, making them universal approximators. However, they also have a major drawback: the objective function, typically based on cross-entropy error, is not convex. Consequently, the synaptic weights obtained during training might not converge to the optimal solution due to the presence of multiple local minima in the objective function. The solution heavily relies on the random initialization of synaptic weights. Additionally, MLPs require tuning more hyperparameters compared to other learning methods such as SVMs or naive Bayes, which presents another disadvantage.

Among the methods mentioned above, logistic regression stands out for its interpretability. However, for the remaining methods, interpretability is a challenge. To address this, decision trees are adopted, as they are classification algorithms commonly used in data mining and knowledge discovery. During decision tree training, a tree is constructed using the dataset as input, where each internal node represents a test on an independent variable, each branch represents the result of the test, and leaves represent predicted classes. The tree is built recursively, starting with the entire dataset as the root node. At each iteration, the fitting algorithm selects the next attribute that best separates the data into different classes. The fitting algorithm may halt based on various criteria, such as when all training data have been classified or when the classifier's accuracy or performance can no longer be improved.

Decision trees are constructed using heuristic algorithms, often employing greedy strategies. At each node, these algorithms may identify several local optimal solutions, leading to no guarantee that the learning process will converge to the most optimal solution. This issue is not unique to decision trees but is also present in other algorithms, such as multilayer perceptrons. However, it remains a primary drawback of decision trees, as small variations in the training dataset can cause significant changes in the tree structure.

The method of decision trees is introduced in 1984, in [23] is delved into its details. To improve the performance of decision trees, ensemble methods based on multiple decision trees have been developed. These methods include Adaboost (adaptive boosting) [24], random forest [25], and extreme gradient boosting, which is also known as XGBoost [26].

Ultimately, so far there is no analytical method to definitively determine the best machine learning approach, as demonstrated by the No Free Lunch Theorem [27]. According to this theorem, the predictive quality of machine learning methods hinges on the unknown distribution of the dataset enshrined to fit them. Consequently, experimental validation becomes the only mean of identifying the most effective method for addressing the problem studied in this work. The following section details the experimental setup adopted to empirically validate these methods.

III. Experimental Setting

To fit and validate the machine learning methods outlined in the previous section, we employed a dataset containing 103 examples. Each example consists of 38 independent variables along with its corresponding dependent or target variable. However, in this study, only 13 out of the 38 independent variables are utilized, as explained earlier. Notably, the dataset is the same used in [2] to compare the results of both studies.

Figure 1 illustrates the proportion of positive and negative instances in the dataset. Positive instances represent examples where students failed the numerical methods course, while negative instances denote cases where students passed. The pie chart depicts that the dataset is reasonably balanced, with a slightly higher number of negative examples due to more students successfully passing the course.

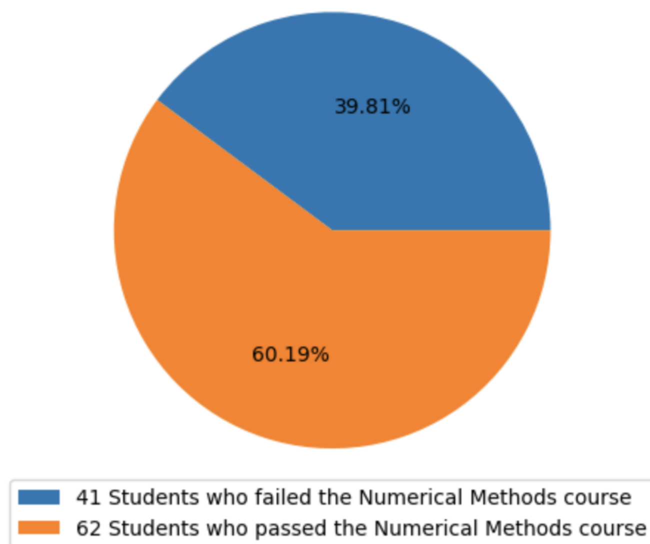

Fig. 1. Distribution of student outcomes in the numerical methods course dataset. The figure illustrates that 41 out of 103 students who participated in the study failed the numerical methods course (39.81% of the surveyed students), while 62 out of 103 students passed the course (60.19% of the sample).

The dataset has been collected through a survey conducted on students enrolled in courses from the fifth to ninth semester of the bachelor's degree program in systems engineering at the University of Córdoba, Colombia. Due to changes in the curriculum structure in 2018, data collection before that year was not feasible, resulting in the dataset's limited size.

As explained in the previous section, the students' outcomes from the Saber 11 test are included in the dataset used in this study. Figure 2 shows that students who failed the numerical methods course scored lower in the mathematics section of the Saber 11 test than those who succeeded in the course. Indeed, the notches of the boxplots in the figure do not overlap, indicating that the median score of students who succeeded in the course is significantly higher than that of students who failed.

The number of times each student enrolls in a prerequisite course is one of the variables in the dataset. If students succeed in the prerequisite course on the first enrollment, this value is equal to one; otherwise, it is greater. Students who have failed the numerical methods course tend to enroll in prerequisite courses more times than those who succeed in the numerical methods course (see Figure 3).

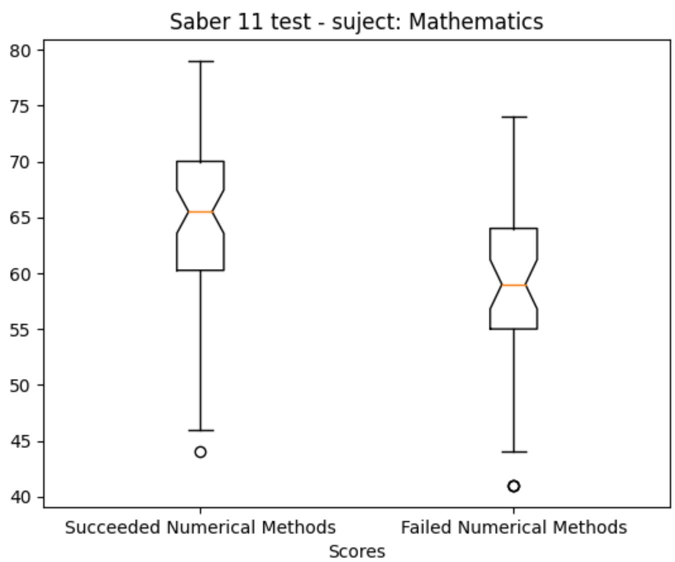

Fig. 2. Boxplots showing the scores obtained by the students in the mathematics subject of the Saber 11 test, categorized based on success or failure in the numerical methods course.

Enrolling in a prerequisite course several times means that a given student has failed it with the same frequency. However, some students who have passed prerequisite courses on their first try have failed the numerical methods course as well. This explains why variables related to the number of enrollments are not sufficient to predict success in the numerical methods course, as evidenced in Figure 3. The boxplots in this figure illustrate that the median of each of these aforementioned variables is equal to one for students who have failed numerical methods.

The lowest grades attained in prerequisite courses are lower for students who have failed the numerical methods course compared to those who have succeeded in it (see Figure 4). The notches of the boxplots do not overlap for most prerequisite courses, except for calculus III. This suggests that performance in calculus III might not significantly contribute to accurate predictions, while the median of the lowest grades in other courses varies depending on whether students have succeeded or failed the numerical methods course.

This observation aligns with the histogram of the lowest grades achieved in calculus III shown in Figure 5, where students who failed this prerequisite course also failed numerical methods, whereas several students who succeeded in calculus III on the first enrollment still failed numerical methods. Recall that if a student succeeds in a prerequisite course on the first enrollment, their lowest grade is at least 3; otherwise, the lowest grade is lower than this value (see Section II).

Similarly, the highest grades in prerequisite courses for students who succeeded in the numerical methods course are better than the highest grades in prerequisite courses for those who failed the numerical methods, as illustrated in Figures 6 and 7.

Therefore, the statistics indicate that students who failed the numerical methods course perform less effectively in prerequisite courses compared to those who succeeded. Indeed, this observation constitutes a key assumption of our study.

Besides, the validation of each classification method with the aforementioned dataset is conducted using the Python programming language and the open-source library Scikit-learn [28]. Scikit-learn provides comprehensive support for various machine learning tasks, including supervised and unsupervised methods. The validation tests are implemented in notebooks within the Google Colaboratory platform [29].

Furthermore, grid search and K-fold Cross-Validation (K-Fold CV) are both utilized in combination to explore various hyperparameter value combinations and tune the hyperparameters of each model.

Finally, during K-Fold CV, a value of $K = 10$ is chosen, although $K = 30$ is also common. However, the larger K is, the smaller the validation set becomes, this potentially limits the ability to test hypotheses concerning the performance of the methods. To assess the validation outcomes, a paired t-test is employed. The results of the validation are presented in the next section, while an analysis of their significant differences, based on paired t-test, is discussed in Section V.

IV. Results

The results obtained from the Ten-Fold Cross-Validation (10-Fold CV) reveal that the Gaussian process (GP) with the Radial Basis Function (RBF) kernel attained the highest accuracy, recall, and harmonic mean (F_1). While the GP with the RBF kernel occupies third place in terms of precision, the method called Support Vector machines (SVM) with the same kernel (RBF) ranks among the top three most accurate methods, alongside the GP with the rational quadratic kernel. Details of these results can be found in Table I.

The GP with the RBF kernel has the best trade-off between precision and recall. This is evident in its F_1 score, which is a desirable feature for an intelligent system predicting student failure in the numerical methods course. For instance, while SVMs with the RBF kernel achieved the highest precision, they had lower recall compared to the GP with the same kernel. This means that a system based on SVMs is more likely to miss students at risk of failing the course compared to one based on GPs.

The results regarding the F_1 metrics are aligned with the confusion matrix shown in Table II, where GP with the RBF kernel predicted that 8 out of 62 students would fail the numerical methods course although they never did, resulting in 8 false positive examples. Moreover, the GP predicted that 10 out of 41 students would not fail the course, although they did, resulting in 10 false negative instances. According to the same confusion matrix, 85 out of 103 students are classified properly during the validation of GP.

TABLE I
TEN-FOLD CROSS-VALIDATION RESULTS

Machine learning method	*Mean Accuracy (%)*	*p-value*	*Mean Precision (%)*	*p-value*	*Mean Recall (%)*	*p-value*	*Mean* $\mathbf{F_1}$ *(%)*	*p-value*
Gaussian process with the radial basis function kernel	**83.00**		80.67		**77.00**		**76.79**	
Gaussian process with the Matern kernel	81.00	0.80	81.67	0.92	69.50	0.55	73.49	0.77
Gaussian process with a sum of radial basis function and Matern kernel	79.00	0.61	76.00	0.63	72.00	0.69	71.62	0.64
Gaussian process with the dot product kernel	76.00	0.33	70.67	0.28	72.00	0.69	68.51	0.42
Gaussian process with the rational quadratic kernel	80.09	0.69	77.33	0.79	67.50	0.54	69.27	0.54
Support vector machines with the radial basis function kernel	80.09	0.69	**85.00**	0.74	55.00	0.13	63.48	0.31
Support vector machines with the sigmoid kernel	78.00	0.48	75.83	0.60	67.00	0.41	69.25	0.46
Support vector machines with the polynomial kernel (degree = 3)	78.00	0.49	79.17	0.91	52.50	0.09	60.55	0.21
Decision tree with gini index	72.91	0.18	75.00	0.59	59.50	0.16	61.90	0.17
Decision tree with entropy index	66.18	**0.04**†	59.67	0.07	54.50	0.07	55.91	0.06
XGBoost	75	0.27	76.83	0.76	52.50	0.06	59.60	0.14
Adaboost with the entropy index	62.27	**0.04**†	56.33	**0.05**†	52.50	0.06	53.41	0.06
Random forest with the entropy index	72.09	0.16	72.50	0.46	52.00	**0.05**†	58.29	0.10
Logistic regression	69.82	0.07	70.00	0.53	27.50	**0.0006**†	37.90	**0.005**†
Multilayer perceptron with a single hidden layer	69.09	**0.04**†	62.76	0.06	69	0.56	61.01	0.12
Multilayer perceptron with five hidden layers	75.09	0.32	65.33	0.21	67.50	0.51	64.33	0.33

†Paired t-test reveals the difference between means is statistically significant

TABLE II
CONFUSION MATRIX ILLUSTRATING THE PERFORMANCE OF GAUSSIAN PROCESS CLASSIFICATION USING THE RADIAL BASIS FUNCTION KERNEL.

	Forecasted class		
True class	*Student might not fail*	*Student might fail*	*Total*
Student did not fail	54	8	62
Student failed	10	31	41
Total	64	39	103

It is noteworthy that the accuracy of the GP with the RBF kernel aligns with the area under the Receiver Operating Characteristics (ROC) curve, as illustrated in Figure 8. With an area of 0.81, this result indicates that the classification method performs much better than random guessing. The dataset's richness contributes to accurately predicting the probability of course failure, demonstrating the classifier's robust discriminatory power, which is well-suited for this predictive task.

The optimal hyperparameter settings for each classifier is determined through 10-Fold CV and grid search. To facilitate the reproducibility of the results in future research, the hyperparameter settings corresponding to the outcomes presented in Table I are as follows:

- Gaussian processes for classification:
  - Radial Basis Function kernel: $\gamma = 0.125$, $\sigma = 0.5$.
  - Matern kernel: $\gamma = 2.44 \times 10^{-4}$, $\sigma = 0.5, \nu = 1.3$.
  - Combination of Radial Basis Function and Matern kernel: $\gamma_G = 0.125$, $\gamma_M = 2.44 \times 10^{-4}$
  - Rational Quadratic kernel: $\gamma = 32$, $\sigma = 0.25$
- Support Vector Machines:

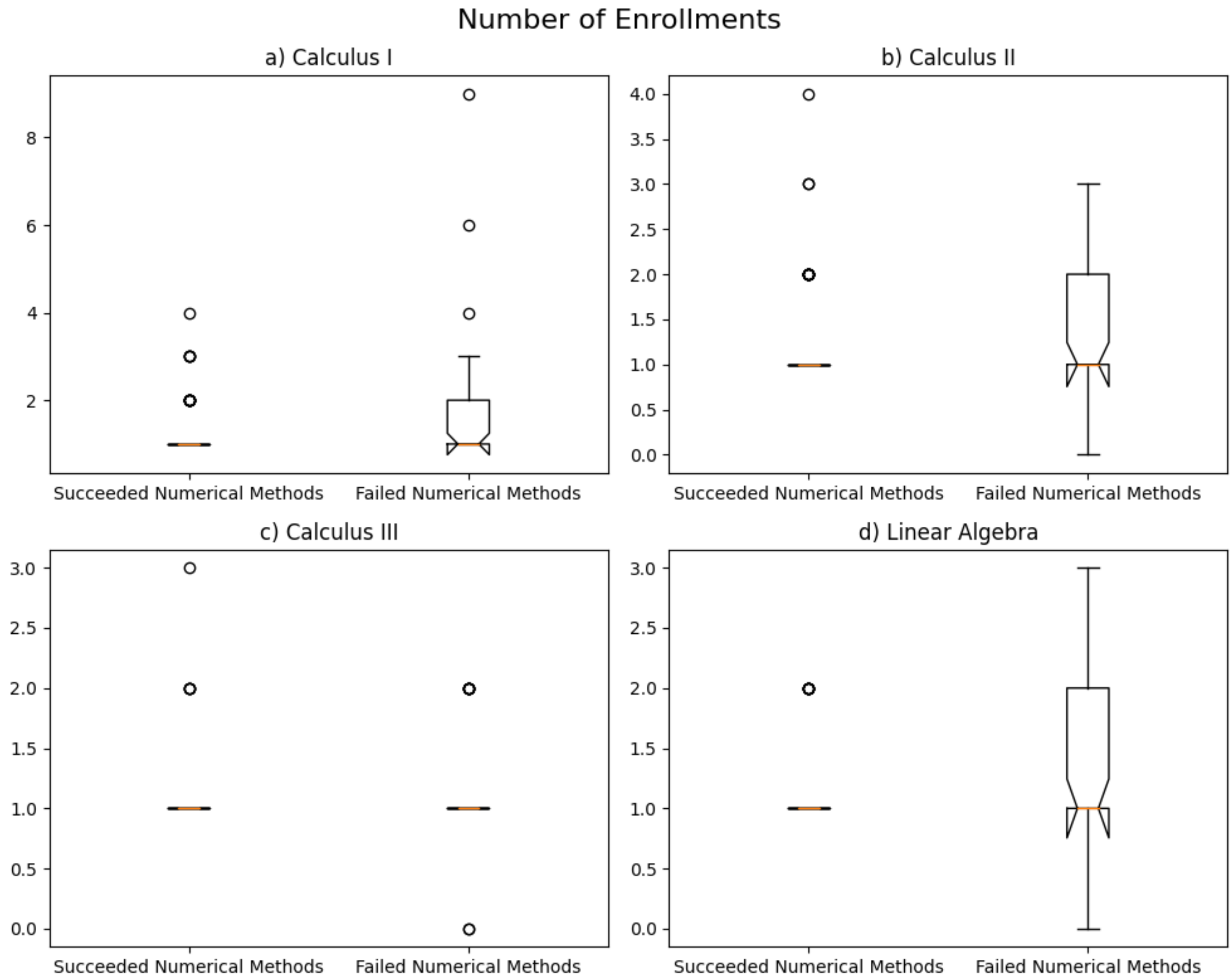

Fig. 3. Boxplots depicting the number of enrollments in each prerequisite course, categorized based on success or failure in the numerical methods course.

  - Radial Basis Function kernel: $C = 1$, $\gamma = 0.25$
  - Polynomial kernel: Regularization parameter $C = 8$, $d = 3$
  - Sigmoid kernel: Regularization parameter $C = 8192$, $\gamma = 3.05 \times 10^{-5}$
- Logistic regression: Regularization parameter $C = 0.01$
- Decison trees: Gini and Entropy indexes.
- XGBoost: Learning rate equal to 0.0625, maximum depth of 5 levels, 80 estimators, and Entropy index.
- Adaboost: Learning rate equal to 0.124, 50 estimators , and Entropy index.
- Random forest: 15 trees, minimum 2 samples per leaf, minimum 4 samples per split, maximum depth of 8 levels.

V. Discussion

Based on the validation results (see Table I), Gaussian processes with the Radial Basis Function kernel (GPRBF) emerged as the top-performing machine learning method in this study. This outcome is due to the fact that there is no hyperplane decision boundary that separates the original input space between the two classes (i.e., students at risk of failing and those not at risk). There is a regular pattern in the academic history of students who fail numerical methods, but no single variable is sufficient to accurately predict their likelihood of failure. For instance, some students who have failed the numerical methods course succeeded in prerequisite courses on their first enrollment, as shown in Figure 4. Therefore, a nonlinear method such as GPRBF is well-suited to the problem addressed in this study.

Besides, paired t-tests are conducted on the means of each metric obtained during validation, revealing the following insights:

- The mean accuracy of GPRBF is far greater than one attained through extreme gradient boosting (or XGBoost), Adaboost, and Multilayer perceptron with a single hidden layer.

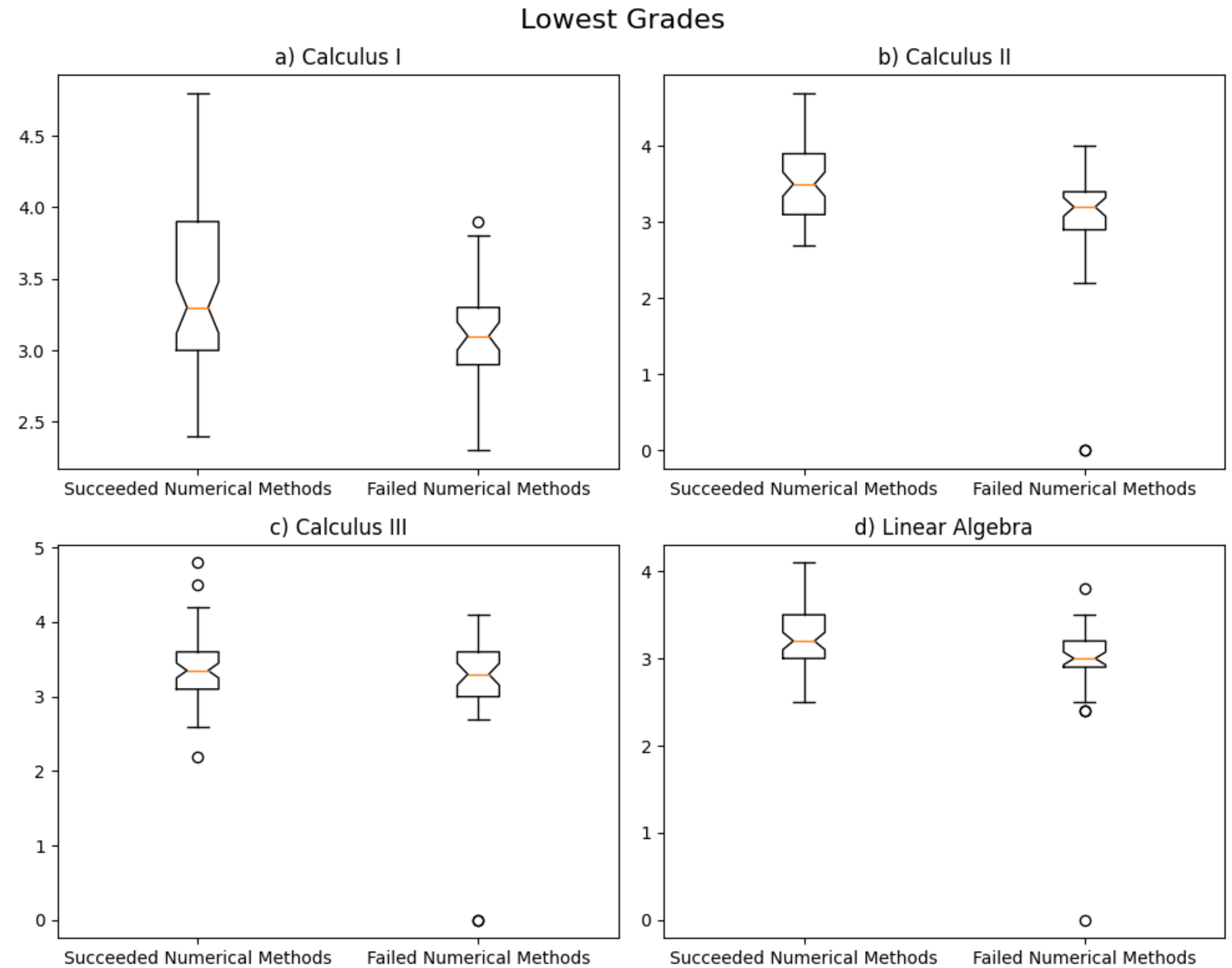

Fig. 4. Boxplots illustrating the lowest grades attained in each prerequisite course, categorized based on success or failure in the numerical methods course.

- A notable difference in mean precision is observed between GPRBF and Adaboost, indicating that the former method is considerably more precise.
- The means of precision between GPRBF and Support Vector Machines with the Radial Basis Function kernel (SVMRBF) do not show significant differences.

Moreover, classification methods with a probabilistic nature, such as Gaussian processes (GP), logistic regression, and Multilayer Perceptron, offer the advantage of providing the user with information about the probability of failing the numerical methods course. For instance, it becomes evident that precautions may be necessary to prevent failure for a specific student, especially when their probability of failing is as high as 80%, compared to another student whose probability is approximately 58%. This underscores the suitability of GP for implementing intelligent systems aimed at the predictive task addressed in this study.

Students identified as being at high risk of failing the course might benefit from support services [30]. These services may include access to course advisors, psychologists, learning and writing advisors, counselors, librarians, disability specialists, and so forth. By directing these resources toward students facing a significant risk of failure, it is possible to optimize cost-effectiveness and ensure targeted support where it is most needed.

Implementing pedagogical contracts between lecturers and students presents another intervention method to support students at high risk of failing the course. These contracts include personalized agreements tailored to each at-risk student's unique needs, strengths, weaknesses, and learning style. They establish clear goals aligned with the student's capabilities and offer flexibility to adapt to individual circumstances throughout the course. By continuously adjusting the contract, educators can provide targeted support to facilitate the student's progress and improve their chances of success.

Thus, students with a moderate probability of failing the

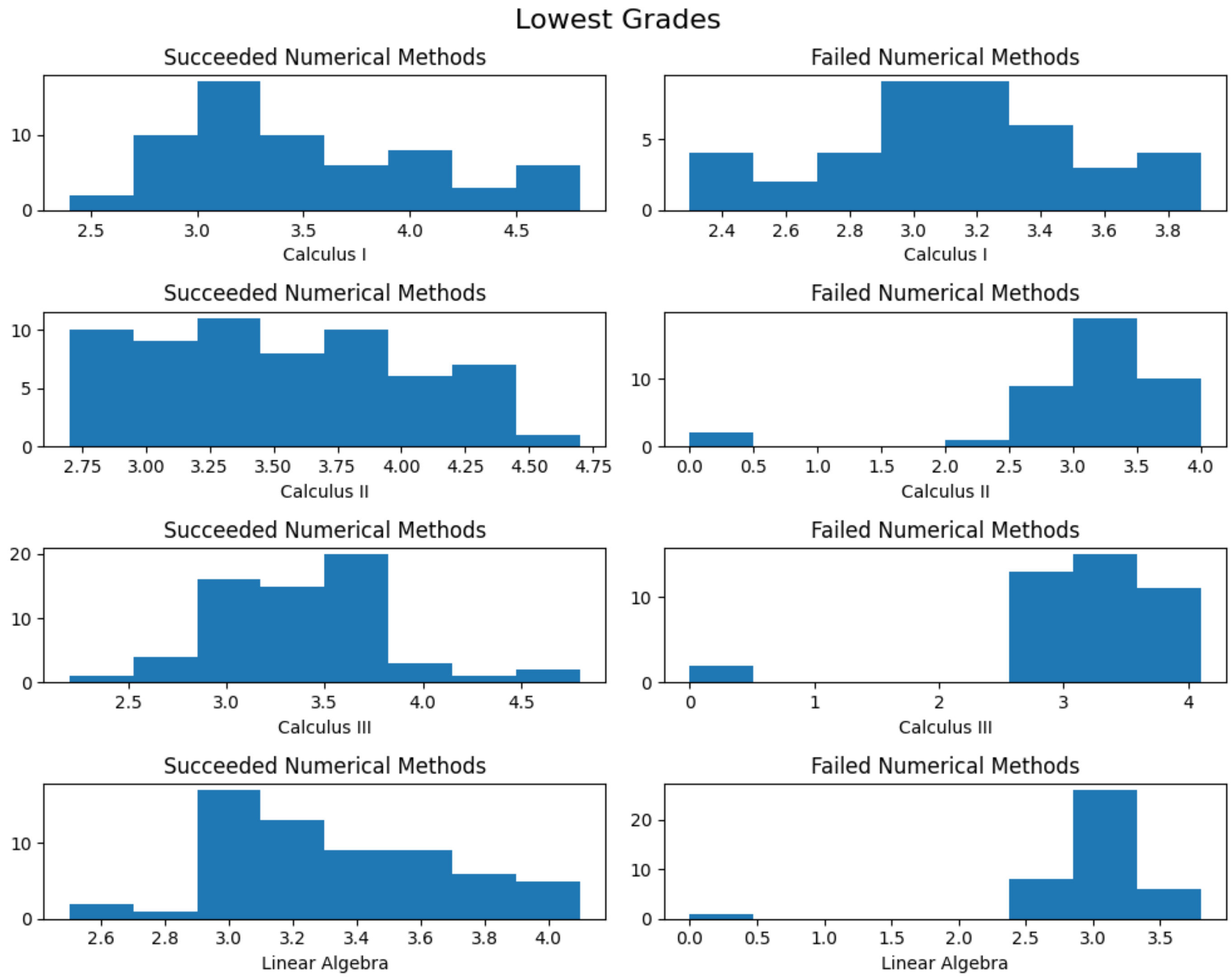

Fig. 5. Histograms displaying the lowest grades obtained in each prerequisite course, categorized based on success or failure in the numerical methods course.

course may benefit from less intensive intervention strategies. For example, providing a variety of instructional approaches, materials, and activities can help to engage these students and address their learning needs effectively. It is also essential to monitor their progress closely and refer them to available support services within the university if necessary, ensuring they receive the assistance required to succeed.

Furthermore, lecturers may leverage probability information to implement differentiated instruction strategies tailored to individual student needs. By understanding the probability of each student failing the course, lecturers may identify areas of weakness and adjust their teaching approach accordingly. For instance, conducting pre-tests and quizzes allows lecturers to determine student comprehension levels and tailor instruction to address specific misconceptions or gaps in understanding. Timely feedback and targeted remediation further support student learning by providing opportunities for reinforcement and mastery of prerequisite competencies. Research in educational psychology has shown that personalized learning approaches may lead to improved student outcomes and engagement [31]. Therefore, by incorporating probability-based insights into instructional planning, instructors can create a more inclusive and effective learning environment for all students.

Utilizing the probability of failing as a metric for student differentiation opens up avenues for fostering collaborative learning environments within the classroom. By identifying at-risk students based on their probability scores, lecturers may orchestrate peer-to-peer instructional sessions and problem-solving activities tailored to address the specific needs of these individuals. This approach facilitates the formation of balanced study groups or teams, where students proficient in prerequisite competences and skills might provide mentorship and support to their at-risk peers. Through collaborative engagement, at-risk students not only receive targeted assistance but also benefit from exposure to diverse perspectives and collective problem-solving, enhancing their overall learning outcomes.

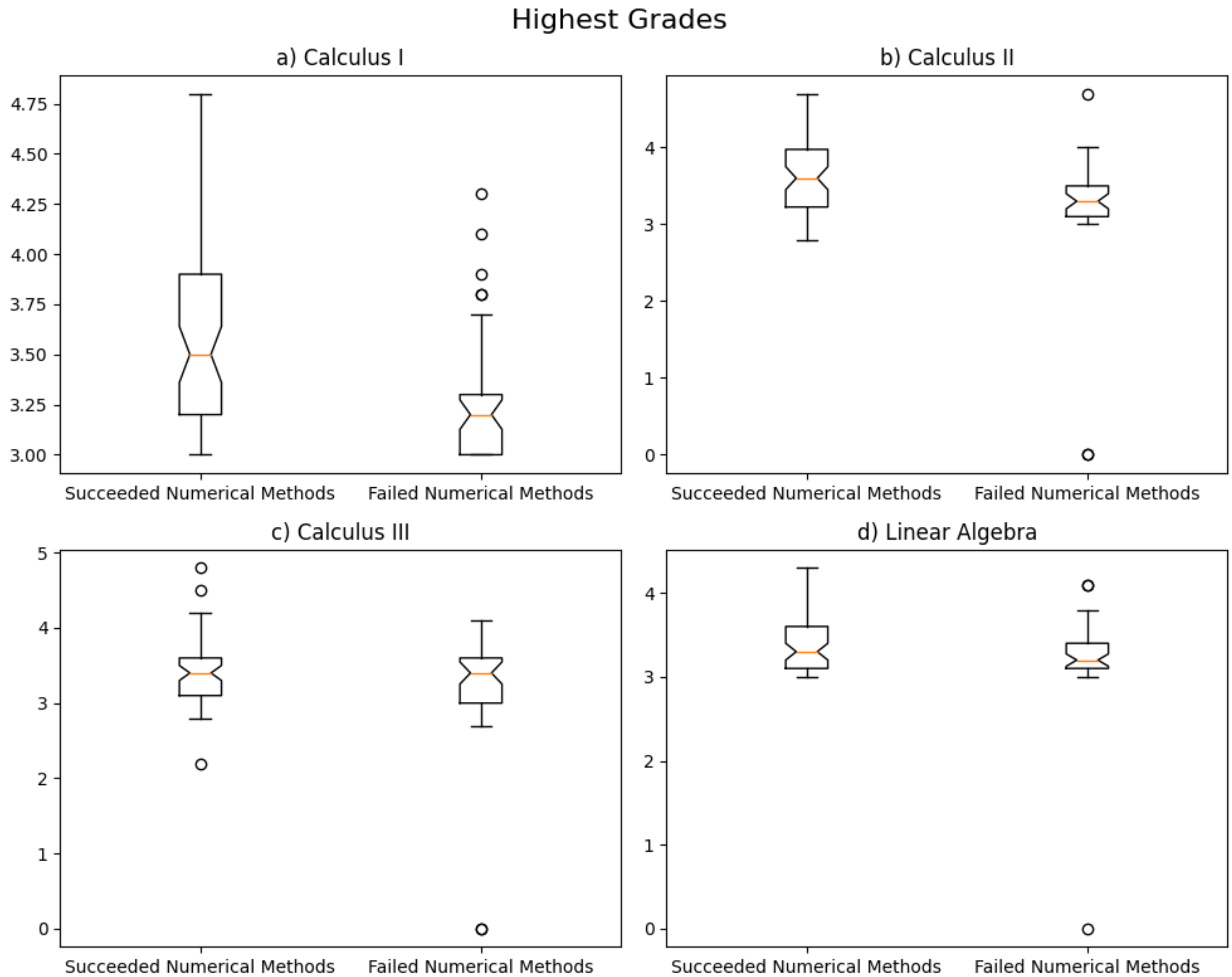

Fig. 6. Boxplots showing the highest grades attained in each prerequisite course, classified based on success or failure in the numerical methods course.

Additionally, at-risk students can benefit from instruction in metacognitive strategies, including techniques for goal-setting, self-monitoring, and reflection. By equipping the student with these cognitive tools, they can enhance their study habits and develop into more strategic and self-regulated learners. These strategies empower the student to take ownership of their learning process, identify areas for improvement, and implement targeted interventions to address challenges they might encounter.

In the context of this study, reporting the probability of failing offers distinct advantages for policymakers and other stakeholders, enabling them to design more effective intervention plans than merely identifying at-risk students in advance. This aspect positions GPRBF as superior to SVMRBF, as the latter does not inherently provide probability information with its predictions. However, this limitation can be addressed by employing Platt scaling [32], which estimates probabilities from the decision values of SVMRBF. Notably, this capability is internally implemented in the Scikit-Learn library.

Incorporating the student's performance in prerequisite mathematics courses, along with their scores on the admission test in this subject, yields improved predictive quality compared to using all scores from the admission test or performance in prerequisite courses related to computer programming or general science (e.g., physics) as input variables. This suggests a significant relationship between the student's proficiency in mathematics and their probability of failing the numerical methods course. It implies that the effectiveness of training in prerequisite mathematics courses such as calculus and linear algebra directly impacts performance in numerical methods.

This relationship between prerequisite mathematics courses and the numerical methods course requires a thorough review of the content covered in prerequisite courses. This review aims to identify key concepts, competencies, abilities, and techniques essential for success in the numerical methods

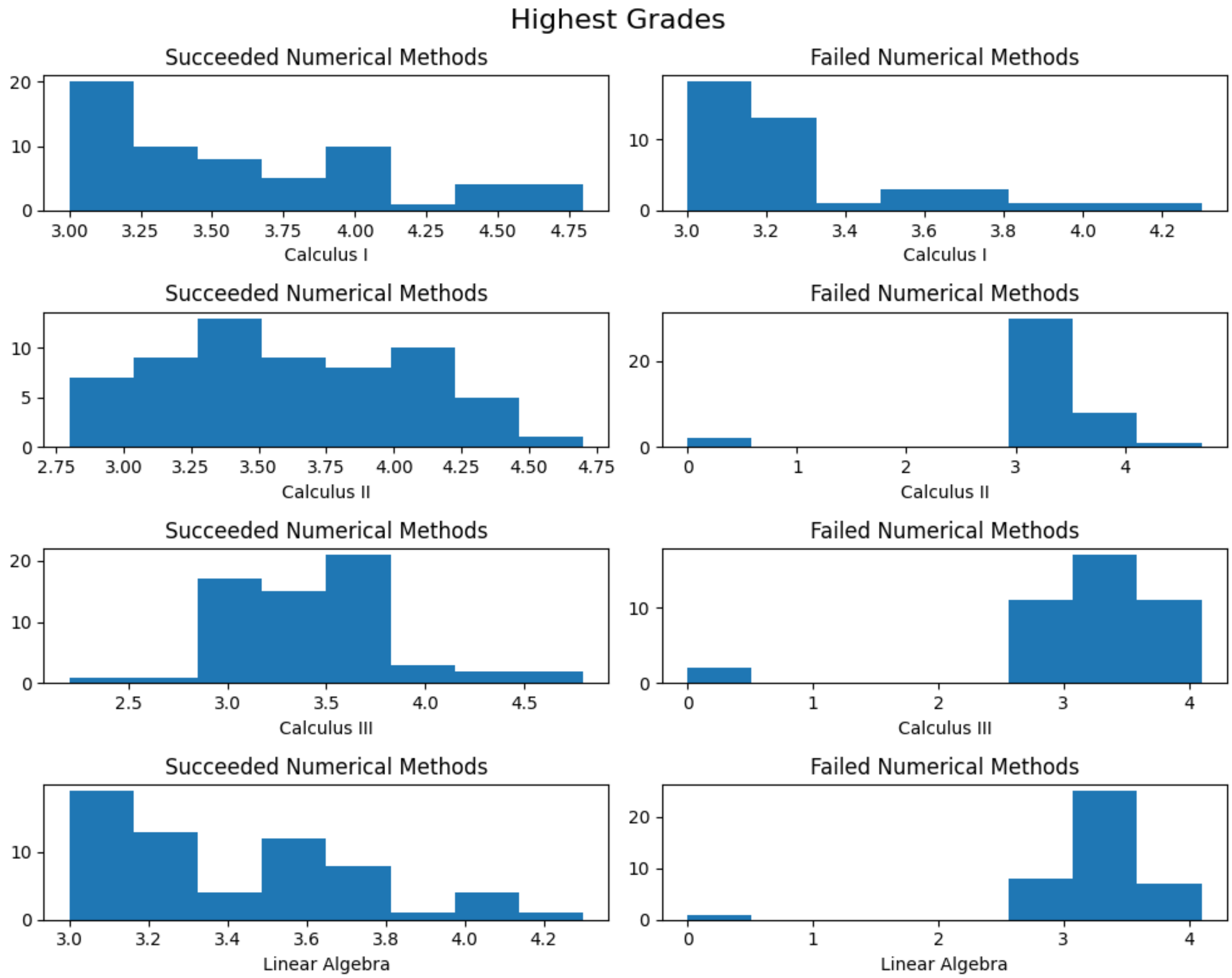

Fig. 7. Histograms illustrating the highest grades achieved in each prerequisite course, categorized based on success or failure in the numerical methods course.

course.

Besides, this process entails identifying common topics, principles, and mathematical techniques that are relevant to both the prerequisite courses and the numerical methods course. This may encompass subjects such as, e.g., differentiation, integration, matrix operations, and so forth. Additionally, it may be beneficial to include in the curriculum, preliminary courses in the first semesters, which introduce foundational mathematical concepts to facilitate the transition from high school to university. Subjects such as set theory, number theory, basic algebra, and analytical geometry might be reviewed to ensure students are adequately prepared.

Finally, this process culminates in mapping the learning outcomes between the numerical methods course and prerequisite courses. This involves articulating the learning outcomes of the numerical methods course and specifying the competences, skills, and knowledge that students are expected to attain. Subsequently, these learning outcomes are aligned with the concepts and goals outlined in the prerequisite mathematics courses.

VI. Conclusions

In conclusion, Gaussian processes for classification with the Radial Basis Function kernel emerged as the top-performing method for the predictive task at hand. Significantly outperforming XGBoost, Adaboost, and Multilayer perceptrons with a single hidden layer, this approach demonstrates superior predictive accuracy. Leveraging machine learning methods, the study forecasts student failure in the numerical methods course based on their performance in prerequisite mathematics courses and admission test scores in the same subject.

The Gaussian processes classification method offers distinct advantages due to its probabilistic nature, providing predictions in the form of probabilities for failing the course. This

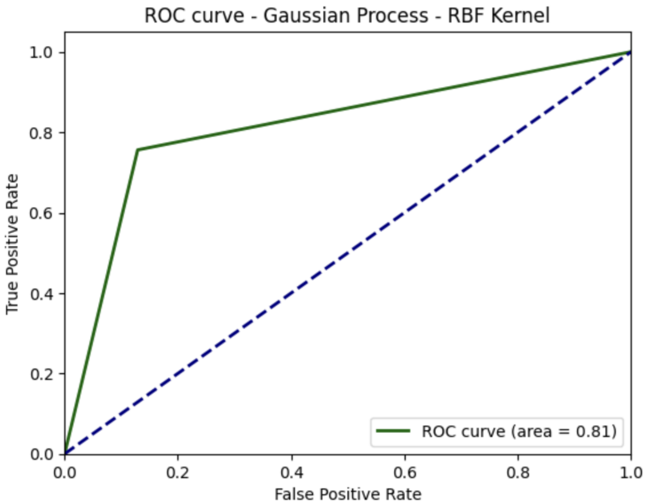

Fig. 8. Receiver operating characteristics (ROC) curve for the Gaussian process with the Radial Basis Function kernel. The diagonal dashed line represents the performance of random guessing, where the classifier performs better than random chance.

feature enables the design of tailored intervention plans for at-risk students while optimizing resource allocation and time efficiency. Future research endeavors could focus on validating intervention plans aimed at supporting students identified as at-risk.

The primary contribution of this work lies in enhancing the prediction quality compared to prior research [1], [2]. This improvement was achieved by modifying the vector representation used in previous studies. Specifically, the focus was narrowed to the student's performance in mathematics, excluding consideration of other subjects such as computer programming and natural science (physics). This underscores the importance of a strong foundation in mathematics for success in the numerical methods course, a finding not previously emphasized in the literature. Nevertheless, further research is recommended to investigate the causal relationship between a student's mathematical skills acquired prior to enrolling in the numerical methods course and their probability of passing it.

Furthermore, the machine learning methodology employed in this study could be expanded to explore the correlation between prerequisite courses across various bachelor's degree programs. This extension has the potential to facilitate curriculum design by mapping the corresponding learning outcomes of prerequisite courses more efficiently.

In future research, we aim to adopt ensemble methods to improve the predictive performance of Gaussian processes. By combining multiple models, such as those with different hyperparameters, kernels, or subsets of the training dataset, we might reduce variance and potentially enhance predictive accuracy. Ensemble techniques may involve averaging predictions of multiple models or using a regression method to weigh their predictions. Although ensemble methods offer the potential for significant improvement in prediction quality, they also introduce computational complexity and require careful hyperparameter tuning.

Acknowledgment

The author thanks the Lord Jesus Christ for blessing this project. He thanks Universidad de Córdoba in Colombia for supporting the Course Prophet Research Project (grant FI-01-22). He also thanks all students who collaborated by answering the survey conducted for collecting the dataset used in this study. Finally, the author thanks the anonymous reviewers for their comments that contributed to improve the quality of this article.

References

[1] I. Caicedo-Castro, M. Macea-Anaya, and S. Rivera-Castaño, "Early Forecasting of At-Risk Students of Failing or Dropping Out of a Bachelor's Course Given Their Academic History - The Case Study of Numerical Methods," in *PATTERNS 2023: The Fifteenth International Conference on Pervasive Patterns and Applications*, ser. International Conferences on Pervasive Patterns and Applications. IARIA: International Academy, Research, and Industry Association, 2023, pp. 40–51.

[2] I. Caicedo-Castro, "Course Prophet: A System for Predicting Course Failures with Machine Learning: A Numerical Methods Case Study," *Sustainability*, vol. 15, no. 18, 2023.

[3] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Computers and Education*, vol. 53, no. 3, pp. 950–965, 2009.

[4] J. Kabathova and M. Drlik, "Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques," *Applied Sciences*, vol. 11, p. 3130, 04 2021.

[5] D. E. M. da Silva, E. J. S. Pires, A. Reis, P. B. de Moura Oliveira, and J. Barroso, "Forecasting Students Dropout: A UTAD University Study," *Future Internet*, vol. 14, no. 3, pp. 1–14, February 2022.

[6] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100066, 2022.

[7] V. Čotić Poturić, A. Bašić-Šiško, and I. Lulić, "Artificial neural network model for forecasting student failure in math course," in *ICERI2022 Proceedings*, ser. 15th annual International Conference of Education, Research and Innovation. IATED, 2022, pp. 5872–5878.

[8] S. Zihan, S.-H. Sung, D.-M. Park, and B.-K. Park, "All-Year Dropout Prediction Modeling and Analysis for University Students," *Applied Sciences*, vol. 13, p. 1143, 01 2023.

[9] I. Caicedo-Castro, "Quantum Course Prophet: Quantum Machine Learning for Predicting Course Failures: A Case Study on Numerical Methods," in *26th International Conference on Human-Computer Interaction*. Springer, July 2024, to appear.

[10] V. Čotić Poturić, I. Dražić, and S. Čandrlić, "Identification of Predictive Factors for Student Failure in STEM Oriented Course," in *ICERI2022 Proceedings*, ser. 15th annual International Conference of Education, Research and Innovation. IATED, 2022, pp. 5831–5837.

[11] S. Merchán-Rubiano, A. Beltrán-Gómez, and J. Duarte-García, "Formulation of a predictive model for academic performance based on students' academic and demographic data," in *IEEE Frontiers in Educacion Conference*, Texas, USA, 2015, pp. 1–7.

[12] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *CoRR*, vol. abs/1810.11363, 2018.

[13] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[14] I. Caicedo-Castro, M. Macea-Anaya, and S. Castaño-Rivera, "Forecasting Failure Risk in Early Mathematics and Physical Science Courses in the Bachelor's Degree in Engineering," in *IARIA Congress 2023: International Conference on Technical Advances and Human Consequences*. International Academy, Research, and Industry Association, 2022, pp. 177–187.

[15] I. Pacheco-Arrieta *et al.*, "Agreement No. 004: Student's code at the University of Córdoba in Colombia," http://www.unicordoba.edu.co/wp-content/uploads/2018/12/reglamento-academico.pdf [retrieved: February, 2024], 2004.
[16] C. Williams and C. Rasmussen, "Gaussian Processes for Regression," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds., vol. 8. MIT Press, 1995, pp. 514–520.
[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
[18] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, 2nd ed. The MIT Press, 2018.
[19] C. Cortes and V. Vapnik, "Support Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
[20] C. C. Aggarwal, *Neural Networks and Deep Learning*. Springer, 2018.
[21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-propagating Errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[22] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," http://arxiv.org/abs/1412.6980 [retrieved: February, 2024], 2014.
[23] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.
[24] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *ICML*, vol. 96, 1996, pp. 148–156.
[25] L. Breiman, "Random forests," in *Machine Learning*, vol. 45, no. 1. Springer, 2001, pp. 5–32.
[26] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794.
[27] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, April 1997.
[28] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[29] "Google Colaboratory," https://colab.research.google.com/ [retrieved: February, 2024], 2004.
[30] R. Ajjawi, D. Boud, N. Zacharias, M. Dracup, and S. Bennett, "How Do Students Adapt in Response to Academic Failure?" *Student Success*, vol. 10, pp. 84–91, 12 2019.
[31] E. T. Khor and M. K, "A Systematic Review of the Role of Learning Analytics in Supporting Personalized Learning," *Education Sciences*, vol. 14, no. 1, 2024.
[32] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers*. MIT Press, 1999, pp. 61–74.

Local-Global Reaction Map for Classifying Listeners by Pupil Response to Sentences with Emotion Induction Words and Its Application to Auditory Information Design

Katsuko T. Nakahira*, Munenori Harada*, Shunsuke Moriya*, Muneo Kitajima*
* *Nagaoka University of Technology*
Nagaoka, Niigata, Japan
Email: {katsuko@vos, s193369@stn, s213345@stn, mkitajima@kjs}.nagaokaut.ac.jp

***Abstract*— When a person acquires a text as auditory information and derives the meaning of the text, he or she may simultaneously generate an emotion in response to the content of the text. Emotions are said to have a certain relationship with decision-making and memory. Therefore, it is expected that even sentences with the same meaning will be remembered differently depending on the emotion evoked. This study aims to clarify the relationship between the emotions that arise when listening to a text and the memory of the presented text. The classification of emotional states held by people is performed by a method based on subjective quantities by impression rating or by a method based on objective quantities by biometric information. In this study, we focus on pupil response, which is biological information that has been suggested to change with emotion. Based on this, this paper proposes the Local-Global Reaction Map (LGR-Map) as a classification method for pupil changes accompanying emotional changes, as a basic research for the construction of adaptive content design methods that utilize the degree of human emotional arousal. The LGR-Map is generated by capturing the emotional changes during listening to a text from the following two perspectives; Those generated by words in a specific region of a sentence (Local reaction); those generated by the context of the entire sentence (Global reaction). The total pupil diameter change within a certain time period is obtained as the characteristic quantity for each response. Error ellipses are defined for the distribution of listeners in the LR-GR for the presented text (LGR-Map), and classified into five types based on the rotation angle and flattening ratio of the error ellipses. The basic properties of the LGR-Map were investigated by using auditory stimuli presented in short sentences containing Affective Norm for English Words (ANEW). As an extension, we will attempt to create an extended LGR-Map for sentences with multiple ANEWs and consider whether it is possible to extract features of the pupillary response. In addition, we discuss the consistency of the results of a recall test in relation to the cognitive model.**

Keywords*— *Local-Global Reaction Map; Pupil Response; Affective Norm for English Words; Emotion Induction; Contents Design of Auditory Information.

I. Introduction

This paper is based on the previous work originally presented in COGNITIVE2023 [1]. An extended LGR-Map analysis was added in Section V.

With the penetration of mobile devices and the development of eXtended Reality (XR) technology, we are surrounded by an increasing number of services that disseminate content via electronic media. Many of these services are designed to enrich the experience of individuals, and their range of application is wide, from sensory experiences, such as sightseeing and movies to educational materials that make it easier for people to acquire knowledge. In recent years, there has been a movement to expand content provision services from an inclusivity perspective (e.g., Vallez et al. [2]).

Content design is essential to content provision in the sense of striving to convey what is to be conveyed as accurately as possible. Content design has the issue of the quality and quantity of the presenting stimulus as the material contained in the content. Visual and auditory information are the central presenting stimuli, and how to handle their quality and quantity is one of the key factors.

Regarding the amount of content, since perceptual information is basically a physical quantity, the amount of processing is determined by the structure of the human cognitive system itself, and individual differences are usually negligible. This is described by Hirabayashi et al. [3] as the relationship between the amount of information and the timing at which the information is given, and it is possible to maximize human memory by giving visual and auditory information, or explicit and implicit information in the appropriate order and intervals.

Furthermore, the quality of content is largely related to the viewer's cognitive process. The cognitive process depends on the richness of information nodes and the state of node connectivity of the information receiver, and thus varies from person to person. Murakami et al. [4][5] discussed the quality of content for short auditory information. They classified the emotions of short sentences into positive, negative, and other categories (in this case, we assign neutral), and calculated memory scores for each category, suggesting that short sentences belonging to a specific category improve memory scores. They also suggested the possibility of using pupil response to measure human emotion induction from short sentences. In addition, Moriya et al. [6] found that pupil responses to Affective Norm for English Words (ANEW) contained in short auditory information may be characterized based on ANEW categories.

Therefore, in order to design content that facilitates better emotional experiences and knowledge acquisition, it is desirable to be able to adaptively provide content according to the viewer's cognitive characteristics. For this purpose, it is necessary to monitor the viewer's emotional state in real time. Biometric information is a suitable indicator for this purpose. There are many types of biometric information on emotion (e.g., Jim et al. [7], Shu et al. [8]), but considering the time scale and ease of measurement, the pupillary response is the most promising. Based on the above, this paper focuses on

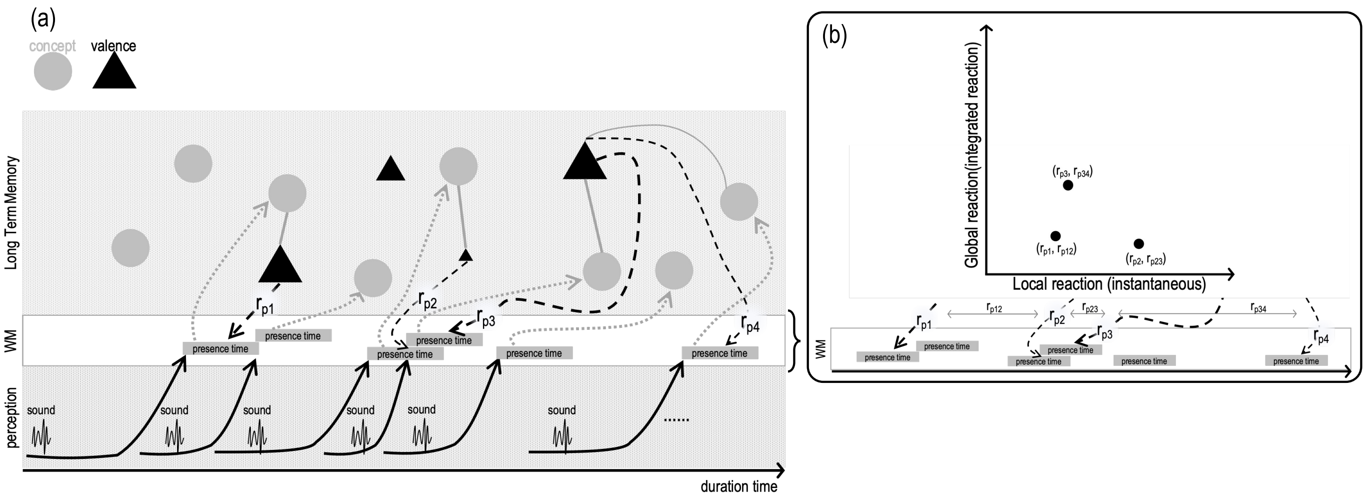

Figure 1. Cognitive model of this paper based on CI model. (a) Input - Cognitive process. (b) Working memory processing - output process.

pupillary response and proposes the Local-Global Reaction Map (LGR-Map) as a classification method for pupillary changes associated with emotional changes.

This paper is organized as follows. In Section II, we construct a base cognitive model and propose an LGR-Map based on it. In Section III, we show the usefulness of the LGR-Map by actually applying it to the HUCAPP 2023 data [6]. In Section IV, we discuss the usefulness of LGR-Map. In Section V, as an extension of the LGR-Map, we attempt to interpret the relationship between pupillary response, and the result of recall test when multiple ANEWs are present.

II. Designing LGR-Map based on Construction-Integration Model

A. Basic Design

In this paper, we construct a reaction model for human emotion based on the Construction-Integration Model (CI-model, e.g., [9][10][11]) proposed by Kintsch. The CI-model is a theory of discourse comprehension consisting of a construction step and an integration step.

The scenario in this paper is modeled based on the CI-model as shown in Figure 1. Figure 1 (a) shows the construction process that encodes information (packet of sound waves) input from the outside world, retrieves information stored in long-term memory using it as a clue, and constructs a network. Figure 1 (b) shows the integration process in which the retrieved information is pruned and integrated in the working memory by pruning information that does not fit the context, and the physical response is output.

First, when a single stimulus (a packet of sound waves in the auditory case) is perceived from a sensory organ, it is sent as encoded perceptual information from the sensory organ to the working memory. The sent information is matched with a large number of nodes (knowledge concepts) in the brain's long-term memory. The corresponding knowledge concept and its associated knowledge concept are then returned to the working memory. In this case, the information of the chunk of emotion (defined by valence and arousal) r_{p_i} associated with the knowledge concept is also returned, so that the working memory temporarily retains the emotion of the perceived packet of sound waves. Based on the returned r_{p_i}, the cognitive process via the working memory activates the motor process in each part of the body, and a response is generated. The pupillary response we focus on in this paper is produced by the activity of the pupillary sphincter and pupillary dilator muscles, which are considered to be one of their responses. The story so far can be expressed as follows.

Let $\boldsymbol{K}$ be a row vector of $^{\forall}$word concepts (knowledge concepts) in the long-term Memory (LTM) of $^{\exists}$person, and the word concept i input at time t_j is denoted by the element $K_i(t_j)$ in $\boldsymbol{K}$. where $K_i(t_j)$ are the values of valence V_i and arousal Ar_i that characterize the emotion [12]. The number of elements is $i = 1 \cdots n_K$ (n_K is the total number of word concepts), with only one i value of 1 for some time t_j. Here, V_i or Ar_i or both may have no value ($\sim$ 0) (in that case, $V_i = 0,\ Ar_i = 0$). The range of values for V_i and Ar_i is $1 \leq V_i \leq 9$, $1 \leq Ar_i \leq 9$.

Next, let $\boldsymbol{A}$ be a column vector of $^{\forall}$emotion concepts in the LTM of $^{\exists}$people, consisting of elements $A_k(V_k,\ Ar_k)$. The number of elements is $k = 1, \cdots, m_A$ (m_A is the total number of emotion concepts), and there always exist $V_k,\ Ar_k$ values.

The $\boldsymbol{K}$ and $\boldsymbol{A}$ are connected by a $n \times m$ matrix $\boldsymbol{W(t_j)}$ that shows their connectivity at time t_j. The element $w_{ik}(t_j)$ of $\boldsymbol{W(t_j)}$ indicates the degree of coupling between $K_i(t_j)$ and A_k. If $K_i(t)$ in $\boldsymbol{K}$ is input and co-occurs with A_k in $\boldsymbol{A}$ on t_j, the probability that $K_i(t_j)$ retrieved from LTM is $p(K_i(t_j))$ and that A_k retrieved from LTM is $p(A_k(t_j))$, the probability of A_k being retrieved from LTM is expressed by the following equation.

$$w_{ik}(t_j) = w_{ik}(p(K_i(t_j)),\ p(A_k(t_j)))$$

In this case, the temporary emotion $E(t_j)$ generated from the

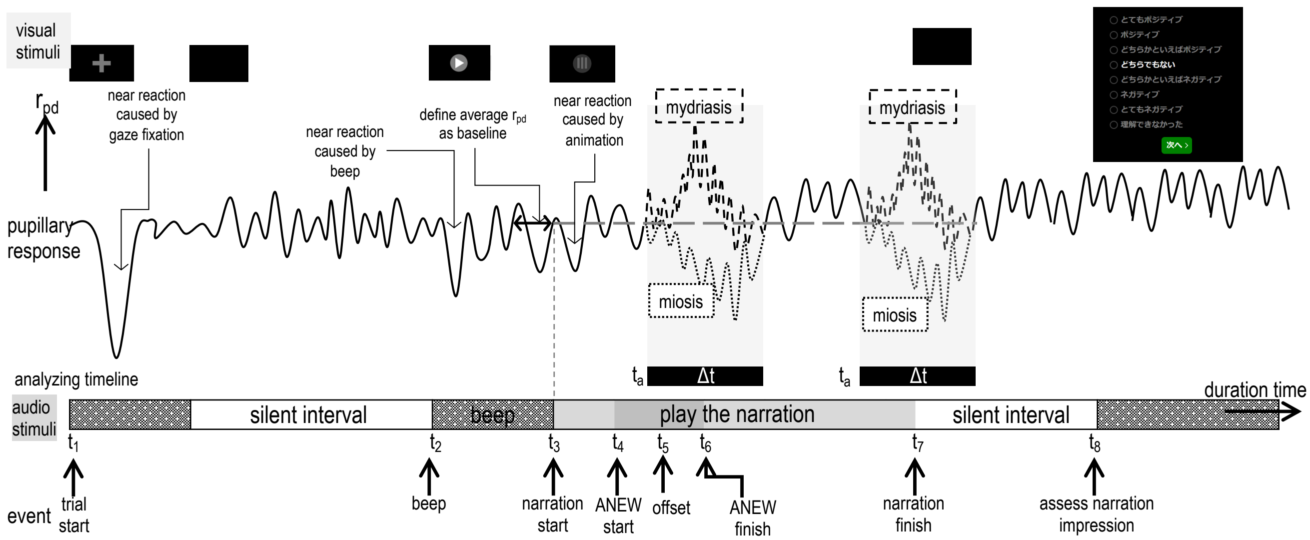

Figure 2. The full pupillary response when auditory information is given.

input $^{\exists}$packet of sound waves is (1).

$$E(t_j) = D(t_j) \sum_{i=1}^{n_K} \sum_{k=1}^{m_A} K_i(t_j) w_{ik}(t_j) A_k \tag{1}$$

Here, $D(t_j)$ is the damping factor. The above explanation represents the construction process.

The sentence, which consists of n_{wp} packets of sound waves, repeats the process of (1) as one cycle up to this point, and continues to return the emotion associated with the knowledge concept to the working memory. In the process, $E(t_j)$ may or may not be integrated between t_j depending on the presence or absence of active sources and contextual relations. The damping factor is introduced as a quantity that indicates the degree of such emotional integration. When n_{wp} packets of sound waves are listened to, the emotion arises in the form of integration of $E(t_j)$ that has been cultivated up to that point. It is usually at the end of a sentence where the packets of sound waves are interrupted. This is the integration process in this research situation.

Based on this, we consider the 2D plane shown in Figure 1(b). We thought that we could show the characteristics of the emotion that occurs in listeners when they listen to narration by plotting the information on human reactions in this plane. The r_{p_i} in the figure indicates the emotional reaction to a specific packet of sound waves. $r_{p_{i,i+1}}$ indicates the emotional reaction generated by the integration of the emotional reactions generated by multiple packets of sound waves. By treating it in this way, two measured reaction quantities can be plotted on a plane as $(r_{p_i},\ r_{p_{i,i+1}})$. In this paper, we call this plane as *Local-Global Response Map* (LGR-Map).

B. *Representation of Pupillary Response based on CI-model*

In order to apply the LGR-Map to pupillary responses, the measurement design of pupillary response should keep an adequate time interval by both inducing time interval both inducing a specific emotion induction word in narrating (instantaneous response) and context of narration (integrated response). Figure 2 represents a measurement design of pupillary response when listening to the narration stimuli based on Figure 1. Here, we adopt the Japanese version of ANEW [15], which induce emotion as the result of instantaneous response. For the integrated response, we assumed that the effect appears at the end of the sentence. We measured participants' pupillary responses to short sentences containing one ANEW word.

The auditory stimuli are adjusted for the event specified in Figure 2 as follows. The beep sound for the mental preparation to initiate auditory stimuli is uttered at t_2. Narration starts at t_3 and ANEW is uttered at t_4. After that, auditory stimuli are terminated at t_7. During this period, the auditory elements related to the evocation of emotion are the ANEW and the atmospheres at the end of the sentence. When analyzing the pupillary response, it is necessary to analyze the data in the vicinity of these elements.

Next, pupil diameter $r_{pd}(t)$ at elapsed time t is processed as follows, in the following order: determination of baseline, calculation of pupil diameter change, and total pupil diameter change.

First, when we set Δt_b as the interval necessary to calculate baseline, baseline $\tilde{r}_{pd}$ is calculated as follows.

$$\tilde{r}_{pd} = \frac{1}{\Delta t_b} \int_{t_{ns}-\Delta t_b}^{t_{ns}} r_{pd}(t)\ dt \tag{2}$$

Here, pupil diameter change value in t $\Delta r_{pd}(t)$ is calculated by the equation (3).

$$\Delta r_{pd}(t) = r_{pd}(t) - \tilde{r}_{pd} \tag{3}$$

The pupil diameter change $\delta r(t)$ between the duration time t and δt is calculated by the equation (4).

$$\delta r(t) = \Delta r_{pd}(t + \delta t) - \Delta r_{pd}(t) \quad (4)$$

Mydriasis (dilation) and miosis (constriction) are typical quantities that show pupillary response. Since the instantaneous changes in either of them are minute, we represent the total amount of change in only mydriasis or only miosis at $[t_a,\ t_a + \Delta t]$. These can be expressed as total amount of mydriasis r_{myd}. The total amount of miosis r_{mio} is calculated by the equation (5).

$$r_{myd} \text{ or } r_{mio} = \int_{t_a}^{t_a+\Delta t} \delta r(t)\ dt \quad (5)$$

In order to obtain a clearer picture of the change in pupillary response, it is better to capture the absolute change in $r_{myd},\ r_{mio}$. Here, we define r_{all} as the total change in pupillary response calculated by (6).

$$r_{all} = |r_{myd}| + |r_{mio}| \quad (6)$$

In LGR-Map, r_{all} is assumed to be a local (instantaneous) or global (integrated) reaction around calculated R. The pupillary response analysis start time t_a and analysis interval Δt can be arbitrarily determined. In the LGR-Map, we set t_a and Δt using the event time in Figure 2 as follows: For the local reaction, t_a is set to t_5, the offset time at which the pupil response is expected to start after the appearance of the ANEW that causes the reaction. For the global reaction, t_a is set where the integrated effect can be easily confirmed. In this paper, t_a is set a little before t_7, when the narration ends. Since the actual narration has n_s sets of calculated points or consists of n_s sentences, at most n_s points are plotted on the LGR-Map.

C. Typology based on LGR-Map

r_{all} distribution on LGR-Map is regarded as a description of induced emotion by narration stimuli for each participant. We design the method of categorization of typology for r_{all} distribution on LGR-Map. r_{all} distribution has x axis for local response and y axis for global response. r_{all} is the information including the individual differences. Now, each individual difference is assumed to obey a normal distribution. If the distribution obeys a two-dimensional Gaussian distribution, we can draw the error ellipsoid on LGR-Map.

The error ellipsoid is represented by the following equation using the transformed coordinates $u,\ v$. Hence $\sigma_u^2,\ \sigma_v^2$ are the variances of the transformed coordinates with respect to the respective axes.

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} = c^2$$

Here, $\sigma_u^2,\ \sigma_v^2$, and rotation angle of error ellipsoid α can be converted as the equations (7) – (9) using σ_x^2 as variance for local response, σ_y^2 as variance for global response, σ_{xy} as covariance of local-global response.

$$\sigma_u^2 = \frac{\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}}{2} \quad (7)$$

$$\sigma_v^2 = \frac{\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}}{2} \quad (8)$$

$$\tan\alpha = \frac{\sigma_{xy}}{\sigma_u^2 - \sigma_y^2} \quad (0 < \alpha < 180^\circ) \quad (9)$$

We consider the shape of error ellipsoid depending on the behavior of $\sigma_x^2,\ \sigma_y^2,\ \sigma_{xy}$. First, we relate σ_x^2 and σ_y^2 as (10).

$$\sigma_y^2 = \gamma\sigma_x^2 \quad (\gamma > 0) \quad (10)$$

The error ellipsoid can then be classified by the value of γ. First, we can set $\sigma_x^2 = \sigma_y^2 = \sigma_0^2$ when $\gamma = 1$. Therefore, $\alpha = 45^\circ$ as shown in the following calculation.

$$\begin{aligned}
\sigma_u^2 &= \frac{\sigma_0^2 + \sigma_0^2 + \sqrt{4\sigma_{xy}^2}}{2} = \sigma_0^2 + \sigma_{xy} \\
\sigma_v^2 &= \frac{\sigma_0^2 + \sigma_0^2 - \sqrt{4\sigma_{xy}^2}}{2} = \sigma_0^2 - \sigma_{xy} \\
\tan\alpha &= \frac{\sigma_{xy}}{\sigma_0^2 + \sigma_{xy} - \sigma_0^2} \\
&= 1
\end{aligned}$$

If $\sigma_{xy} \sim 0$, the distribution has a circle shape; if it has a large value, the distribution has an ellipsoid shape.

Next, we consider the case of $\gamma \neq 1$ in the equation (10), where we apply the observed data properties to the variables in the equations (7) – (9). Since $\sigma_x^2,\ \sigma_y^2$, and σ_{xy} are at most on the order of 10^{-2} given the experimental environment, the σ_{xy} term is on the order of 10^{-4}. Therefore, we can ignore the σ_{xy} term. The equations (7) – (9) can be approximated by the following equations.

$$\sigma_u^2 = \frac{\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2}}{2} \sim \sigma_x^2 \quad (11)$$

$$\sigma_v^2 = \frac{\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2}}{2} \sim \sigma_y^2 = \gamma\sigma_x^2 \quad (12)$$

$$\tan\alpha = \frac{\sigma_{xy}}{\sigma_u^2 - \sigma_y^2} = \frac{\sigma_{xy}}{(1-\gamma)\sigma_x^2} \quad (13)$$

In the situation, considering the range of γ and signum of σ_{xy}, we can predict the following categories. Hence, $L,\ G$ represent local or global reaction, and $+,\ -$ after the L or G represent strong or weak effect. $_{-}$, $^{-}$ represent the spreading to lower or upper side of data.

- case $\sigma_{xy} \sim 0$:
  - $\gamma \sim 1$: $L0G0$
    The error ellipsoid distribution has circle shape.
  - $0 < \gamma \ll 1$:*L*+*G*–
    The shape becomes parallel to the x axis, and $\alpha \sim 0^\circ$.

  - $\gamma \gg 1$: *L–G+*
    The shape becomes parallel to the y axis, and $\alpha \sim 90°$.
- case $\sigma_{xy} > 0$: $L_{-}G^{-}$
  The shape becomes parallel to the x (in case of $0 < \gamma < 1$) or y (in case of $\gamma > 1$) axis, and $0° \ll \alpha < 90°$.
- case $\sigma_{xy} < 0$: $L^{-}G_{-}$
  The shape becomes parallel to the x (in case of $0 < \gamma < 1$) or y (in case of $\gamma > 1$) axis, and $90° \ll \alpha < 180°$. In case of $\sigma_x^2 \sim \sigma_y^2$ $(\gamma \sim 1)$, $\alpha \sim 135°$.

III. Typology of Pupillary Response based on LGR-Map

To evaluate the validity of the LGR-Map designed in Section II, we analyzed the pupillary response. The LGR-Map analysis was conducted using the pupillary response data measured for the controlled narration in the form of Figure 2.

A. *Characteristics of Data for Generating LGR-Map*

The data used are those obtained by [6]. The data profile is as follows. The narration source used in the experiment is designed as shown in Figure 2.

1) The narration is played back in Japanese, and is a short sentence consisting of about 30 syllables.
2) One ANEW corresponding to either high-positive valence V_{++}, high-negative valence V_{--}, or neutral valence V_N was placed at t_{vs} in one sentence.
3) After the appearance of an ANEW, we assigned an expression that characterizes the mood of the whole sentence as positive (At_+) / neutral (At_N) / negative (At_-).
4) After t_4, the analysis interval from t_a as t_5 to Δt, where the pupillary response is expected to start, was set as analysis interval 1.
5) The response that occurs at $0.5\Delta t$ before and after the end of narration was defined as analysis interval 2.

Therefore, analysis interval 1 was defined as local reaction (instantaneous reaction) and analysis interval 2 as global reaction (integrated reaction). Twenty-one participants in their 20s were included, but data of two participants were excluded due to inaccuracy.

Table I shows the results of subjective evaluation of narration stimuli by participants. The narration stimuli are composed of V_{--}, V_{NN}, V_{++} and At_-, At_N, At_+. The participants listened to each stimulus and then evaluated their impressions on a 7-point scale from high negative to high positive, indicating whether their ratings were consistent with the valence or the atmosphere. However, cases in which the impression matched less than 10 participants were excluded.

B. *LGR-Map to Represent Individual Participants' Response Sensitivity*

For each participant, an LGR-Map was created for all narration stimuli for the pupillary responses obtained under the above conditions. In order to confirm that the distribution was independent of the size of the individual pupillary response, median-normalized values within analysis interval 1 and analysis interval 2 were used for the plots.

TABLE I. The results of the emotional arousal effect of sentences and the impression evaluation. V denotes valence, At denotes atmosphere, S_I denotes Score of Impression.

V	At	$V = S_I(\%)$	$At = S_I(\%)$	Number of Trial
V_{NN}	At_N	38	38	50
V_{--}	At_-	58	58	54
V_{++}	At_+	54	54	49
V_{NN}	At_-	16	39	24
V_{NN}	At_+	17	54	34
V_{--}	At_+	8	28	29
V_{++}	At_-	4	52	49

TABLE II. The α and flattening rate of the error ellipse in the LGR-Map for the characteristics of the narration stimulus. F_{ee} denotes the flatness.

V	At	α	F_{ee}	V	At	α	F_{ee}
V_{NN}	At_-	63.6°	0.506	V_{--}	At_-	28.4°	0.091
V_{NN}	At_N	43.1°	0.350	V_{--}	At_+	12.3°	0.434
V_{NN}	At_+	32.4°	0.246	V_{++}	At_-	52.2°	0.182
				V_{++}	At_+	−7.43°	0.194

From the equation (1), we expect that the distribution of individual participants' pupillary responses in the LGR-Map can be classified into five types. Figure 3 shows a representative example of an LGR-Map created using the pupillary responses of individual participants to narration stimuli. As shown in Section II, (a) in Figure 3 is $L+G-$, same as (b) is $L0G0$, (c) is $L^{-}G_{-}$, (d) is $L_{-}G^{-}$, and (e) is $L-G+$. When creating the LGR-Map for individual participants, we also examined whether there was a bias in the pupillary response to a particular valence or atmosphere, but no bias was found.

IV. Discussion: Implications of LGR-Map

A. *LGR-Map for Characterizing Individual Participant*

The classification of individual participants was not characterized by a distinctive response to the combination of (valence, atmosphere), which indicates emotion, suggesting that it was simply determined by the distribution of $w_{ik}(t_j)$, which is indicated by the equation (1). The intensity of $w_{ij}(t_j)$ is considered to change depending on the intensity of the individual's experience of emotion. If the overall experience of emotion is weak, or if the experience of emotion is weak for some reason and almost no emotion is generated, the response of $L_{-}G^{-}$ is expected to be shown. When the reaction is triggered by either valence or atmosphere, it is considered to have a reaction of *L+G–* or *L–G+*. If the reaction is equally distributed between valence and atmospheres, the reaction is considered to be $L^{-}G_{-}$. If the reaction is completely random, it is considered to be $L0G0$.

B. *LGR-Map for Categorizing Narrations*

Next, we consider human responses to ANEWs used as narration stimuli. Since ANEWs are basically emotion references elicited when people hear the word, we believe that it is possible to evaluate the validity of narration stimuli that show the same atmospheres as ANEWs by using the LGR-Map type classification.

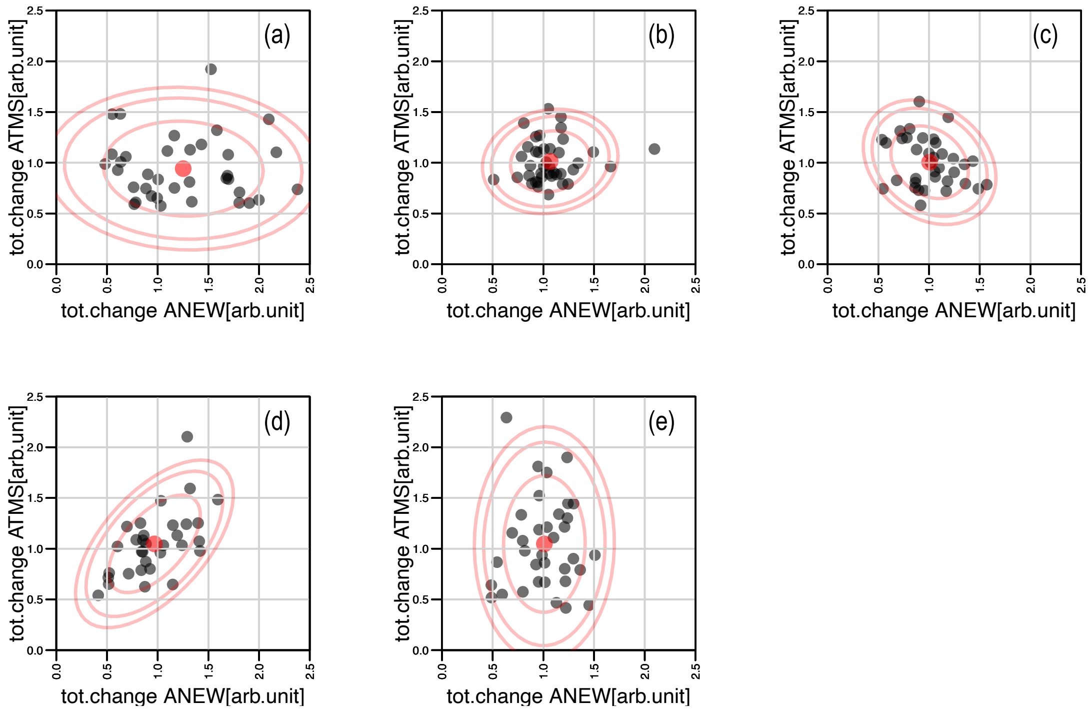

Figure 3. Examples of LGR-Maps for individual participants. The LGR-Map an individual participant is normalized by a median of r_{all} near ANEW and near the end of the sentence, respectively. The oval lines indicate 66%, 90%, and 95% confidence levels from the inside. The categories in LGR-Map are as follows: (a) $L+G-$, (b) $L0G0$, (c) L^-G_-, (d) L_-G^-, (e) $L-G+$.

Table II shows the values of α and flattening F_{ee}, which are the features of LGR-Map. The features in the LGR-Map are created by combining the valence and atmospheres into 9 patterns. There were three responses to each stimulus pair. Trials with fewer than 10 trials showing the level of response to a stimulus pair were excluded from the analysis, considering them to be less significant even if an error ellipse was written.

The table shows the following characteristics. For $V_{--}At_-$, F_{ee} is almost zero, indicating that it is a circular distribution. Therefore, (a) is classified as $L0G0$. For $V_{--}At_+$, α is 13.2°, almost parallel to the x axis, so it is classified as $L+G-$. For $V_{NN}At_N$, it is classified as L_-G^-, because $\alpha \sim 45°$. $V_{NN}At_-$, $V_{NN}At_+$ are not certain because the value of α is ambiguous, but we can classify them as L_-G^- for the reason described later. Since $V_{++}At_-$, $V_{++}At_+$ have ambiguous α values and F_{ee} values are not circular, we cannot indicate which type they can be classified into at this time.

From the above, the following possibilities are considered for $V_{--}At_-$, $V_{--}At_+$, $V_{NN}At_N$. For $V_{--}At_-$, values of valence and atmosphere have negative each. In this case, valence and atmosphere are the same characteristics, so that we anticipate that participants' pupillary responses are almost uniform as indicated by Murakami et al. [4][5]. Taken together, these results suggest that the distribution of the LGR-Map is random, centered on a representative LGR-Map value.

For $V_{--}At_+$, the response is negative valence, with positive atmosphere. This, together with $V_{--}At_-$, can be interpreted as follows. The response to the negative valence was scattered, but the response to the pupillary response in the atmospheres was reasonably consistent, resulting in the values in the y axis being almost consistent and the distribution in the x axis being broadened. This is thought to be due to the broadening of the x-axis distribution.

Next, we consider the case of $V_{NN}At_N$. Both valence and atmosphere were neutral, that is no emotion is induced. It indicates that no matter where in the instantaneous or integrated area the pupillary response is measured, no change in emotion occurs for the same person or narration. Therefore, both pupillary responses show almost similar values, which is a good sign that the distribution is close to a straight line with $y = x$.

V. Generated Emotion from Sentences with Multiple ANEWs

The LGR-Map introduced in Section II is based on the relationship between a single ANEW and sentence-final expressions (positive/negative). In this case, since the factors that promote emotional change are either emotion-inducing words or sentence ends, the LGR-Map can be created by examining the degree of total change in respective pupillary responses.

However, in the real world, it is not uncommon to find multiple ANEWs in a single sentence, or to induce emotions from information given by a sequence of multiple sentences. In order to adapt LGR-Map to such a situation, several modifications may be necessary for the definition of the LGR-Map. In this section, as a first step, we discuss the total change of pupillary response expected when auditory stimuli consisting of one or two sentences with two ANEWs are presented to a participant. In addition, the degree of memorization of the given stimuli is also discussed with reference to Figure 1.

A. Basic Design

The emotion induction when participants hear sentences with multiple ANEWs is affected by various among ANEWs, compared to the emotion induction with a single ANEW. The interference is caused by the valences of ANEWs or the presentation interval between them. In this subsection, we discuss the effects of each type of interference.

Table III shows the classification of emotion induction by the valences of two ANEWs, *ANEW*$_1$ and *ANEW*$_2$, vis-a-vis the interval between them, and predicted degree of memorization for the content of stilumi. The valences V and V' are defined as follows:

$$\begin{aligned} V,\ V' &\in \{V_{++},\ V_{--}\}, \\ \text{where } (V,\ V') &= (V_{++},\ V_{--}) \text{ or } (V_{--},\ V_{++}). \end{aligned}$$

We set the interval between *ANEW*$_1$ and *ANEW*$_2$, i.e. T_{int} as short or long.

Depending on the combination of possible values of valence for ANEWs, the following four cases are possible for emotion induction and the resulting degree of memorization of the concept carried by the presented stimuli.

(a) ANEWs valence: $(V,\ V)$, T_{int}: short
After the short T_{int} from the presentation of *ANEW*$_1$, *ANEW*$_2$ appears, which means emotionally similar concepts to *ANEW*$_1$ are triggered in close proximity to the presentation of *ANEW*$_2$. The concepts carried by the stimuli's contents might be highly activated as a result of triggering the adjacent regions in long-term memory, which have overlapping areas of the presented two ANEWs, at short time intervals. This would result in the strengthening of the association between the concepts and the ANEWs. As a result, the concept, i.e., the stimuli's contents, will be recalled easily, and can be answered easily on the recall test.

(b) ANEWs valence: $(V',\ V')$, T_{int}: long
After the long T_{int} from the presentation of *ANEW*$_1$, *ANEW*$_2$ appears, which means emotionally similar concepts are triggered at a long interval. The concepts carried by the presented stimuli will be activated in isolation due to the quasi-independent triggering of long-term memory. Even if the triggered areas have overlapping areas, it is not likely that they reinforce each other because the timing of activation is not synchronized. The performance for the recall test will be limited because the mutual reinforcement of activities of the concepts of input stimuli would not be as effective as in the case of (a), in which the presentation of succeeding ANEWs in a short-interval assures the possibility of a synchronized activation of the concepts.

∗ The following two cases are incongruent versions of (a) and (b). The two ANEWs trigger isolated regions of long-term memory when presented. It is not likely that they jointly reinforce the activation of long-term memory. However, the short or long interval between the two ANEWs would have distinctive effects on memory test as described below.

(c) ANEWs valence: $(V,\ V')$, T_{int}: short
After the short T_{int} from the presentation of *ANEW*$_1$, *ANEW*$_2$ appears, which means emotionally different concepts from *ANEW*$_1$ are triggered in close proximity to the presentation of *ANEW*$_2$. The concepts carried by the presented stimuli will be activated as a result of triggering long-term memory, which might not have overlapping areas, at short time intervals. This would result in a weak association between the concepts and the ANEWs. As a result, the concept, i.e., the stimuli's contents, will be hard to recall, and can not be answered easily on a recall test.

(d) ANEWs valence: $(V',\ V)$, T_{int}: long
After the long T_{int} from the presentation of *ANEW*$_1$, *ANEW*$_2$ appears, which means emotionally different concepts from *ANEW*$_1$ are triggered at a long interval. The concepts carried by the presented stimuli will be activated independently as a result of triggering of long-term memory with possibly the least overlapping areas of the two ANEWS. As a result, the performance of the recall test does not improve, and the memory recall depends on the degree of past experience.

Among these states, we predict that interference for association strength between ANEWs is expected in the behavior of r_{all}, especially for cases (a) and (c), i.e., when ANEWs have a short appearance interval. We describe r_{all} for *ANEW*$_1$, *ANEW*$_2$, and *ES* by $r_X(X = \{ANEW1/ANEW2/ES\})$; we regard r_X as a quantity that indicates the degree of response to *ANEW*$_1$/*ANEW*$_2$/*ES*. The relationship between r_X's in terms of the degree of their amount indicates which partial ANEWs or whole sentences should have caused a stronger response. Based on this assumption, we extend the LGR-Map introduced in Section II. Toward constructing the extended LGR-Map, we attempt to find the influence of emotion induction as follows: to see the local response, we calculate the difference of r_{ANEW1} and r_{ANEW2}; for global response, we calculate the difference of r_{ANEW1} or r_{ANEW2} and r_{ES}.

Figure 4 shows the relationship between the profile of r_{PD}, i.e., numerical characteristics of r_X, and the characteristics on the extended LGR-Map. Figure 4 (A) represents the possible observed pupillary response curve, which is derived from emotion induction response. The red/blue/green rectangle areas in Figure 4 (A) represent the area of the total change in pupillary response to the first ANEW (*ANEW*$_1$) / second ANEW (*ANEW*$_2$) / the end of sentence (*ES*), $r_{ANEW1/ANEW2/ES}$, respectively. $R1/R2/RS$ in Figure 4 represents the responses to *ANEW*$_1$/*ANEW*$_2$/*ES*. In this case, there are six different combinations of the two quantities between r_Xs, as shown in Table IV. Figures 4 (B) and (C) represent the behavior

TABLE III. Classification of emotion induction by the valences, $ANEW_1$ and $ANEW_2$, and the interval between ANEWs and prediction of the degree of memorization

Valence of $ANEW_1$	Interval T_{int}	Valence of $ANEW_2$			
		V		V'	
		degree of related activation of concepts	degree of memorization	degree of related activation of concepts	degree of memorization
V	short	high	high	dependent on concepts	dependent on concepts
V'	long	low (hard to associate)	memorization to the extent of past experience	low (hard to associate)	low

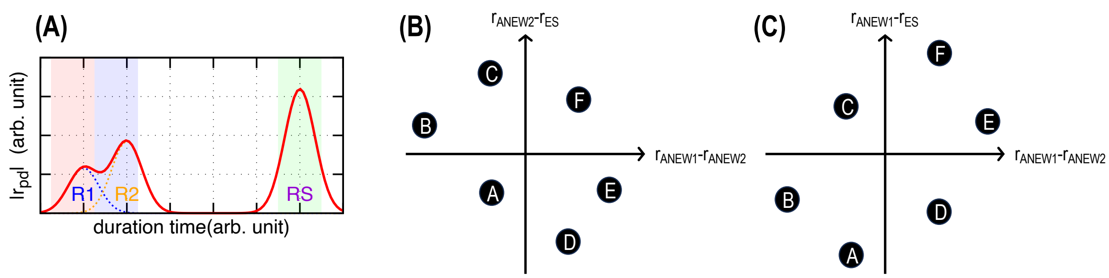

Figure 4. Observed images of the magnitude of the absolute total pupil response for $ANEW_1$, $ANEW_2$, and ES. (A) shows the composite of $|r_{pd}|$ for $ANEW_1$, $ANEW_2$, ES. (B) and (C) projected onto the extended version of the LGR-Map plane, which shows the behavior between r_X corresponding to Table IV.

for different r_X's in the extended LGR-Map, which are categorized in Table IV. Figures 4 (A) through (F) indicate the types based on the relationship between r_Xs, as shown in Table IV.

We considered that the degree of valence induced in conjunction with concepts, as predicted in Table IV, leads to emotion induction, which is directly related to the magnitude of r_X. Compared to the case where a single ANEW appears in isolation, the consideration shown in Table III for two ANEWs suggests that the following is expected: The presentation of two congruent ANEWs will affect each other in a direction that is more likely to induce emotion. A short interval presentation will, in particular, induce a particularly strong emotion induction. The concepts are then recalled under the influence of stronger association strength compared to weak emotion induction. As a result, the results of the recall test will improve. This situation can be rephrased as follows. Once an emotion is induced by $ANEW_1$, the emotion induction is continued by the same type of $ANEW_2$ shortly thereafter. Since $ANEW_2$ continues the emotion induction while the emotion was once induced by $ANEW_1$, at least the effect of local ANEWs is expected to be $r_{ANEW1} < r_{ANEW2}$. From this, we can specify that one of the types A∼C in Table IV is advantageous. The relationship between $ANEW_1/ANEW_2$ and ES depends on the relationship with r_{ANEW1}/r_{AEW2}, but it is reasonable to assume that r_{ES} will have a smaller value than r_{ANEW1}/r_{AEW2} under the current assumptions after stronger emotions are generated. Based on the above, the behavior of r_X in Table III (a) is expected to be consistent with the behavior shown in Table IV (B) and (C).

B. Preliminary Experiment

To confirm the prediction considered in Section V-A, we performed a preliminary experiment. The narration structure of the auditory stimuli designed in Figure 2 was modified as shown in Figure 5, and 32 auditory stimuli containing two ANEWs were prepared according to the following policy.

- We adopt ANEWs which arousal values are moderate and valence values are V_{++} with the range of $(7.70 - 9.00)$ or V_{--} with the range of $(1.00 - 1.99)$.
- Each sentence, i.e., an auditory stimulus, consists around 65 syllables.
- The interval between two ANEWs is as follows:
  - Each ANEWs are almost adjacent to each other, and has an inference on the pupillary response to ANEW (we adopt the 1 second interval for this experiment).
  - The two ANEWs are apart long enough to have little interference with each other on the pupillary response to ANEWs (we adopt the 7 second interval for this experiment).

In addition, a recall test was conducted on the content of every eight sentences to check the degree of memorization.

TABLE IV. Behavior between r_X when LGR-Map is extended.

	X-axis	Y-axis	
	$r_{ANEW1} - r_{ANEW2}$	$r_{ANEW2} - r_{ES}$ Figure 4(B)	$r_{ANEW1} - r_{ES}$ Figure 4(C)
A	$-$	$-$	$--$
B	$--$	$+$	$+$
C	$-$	$++$	$+$
D	$+$	$--$	$-$
E	$++$	$-$	$+$
F	$+$	$+$	$++$

C. Result of Preliminary Experiment

The analysis of the preliminary experiment was conducted as follows. Because this was a preliminary experiment, there were two participants. Since there are two ANEWs in each auditory stimulus, we need to extend the LGR-Map introduced

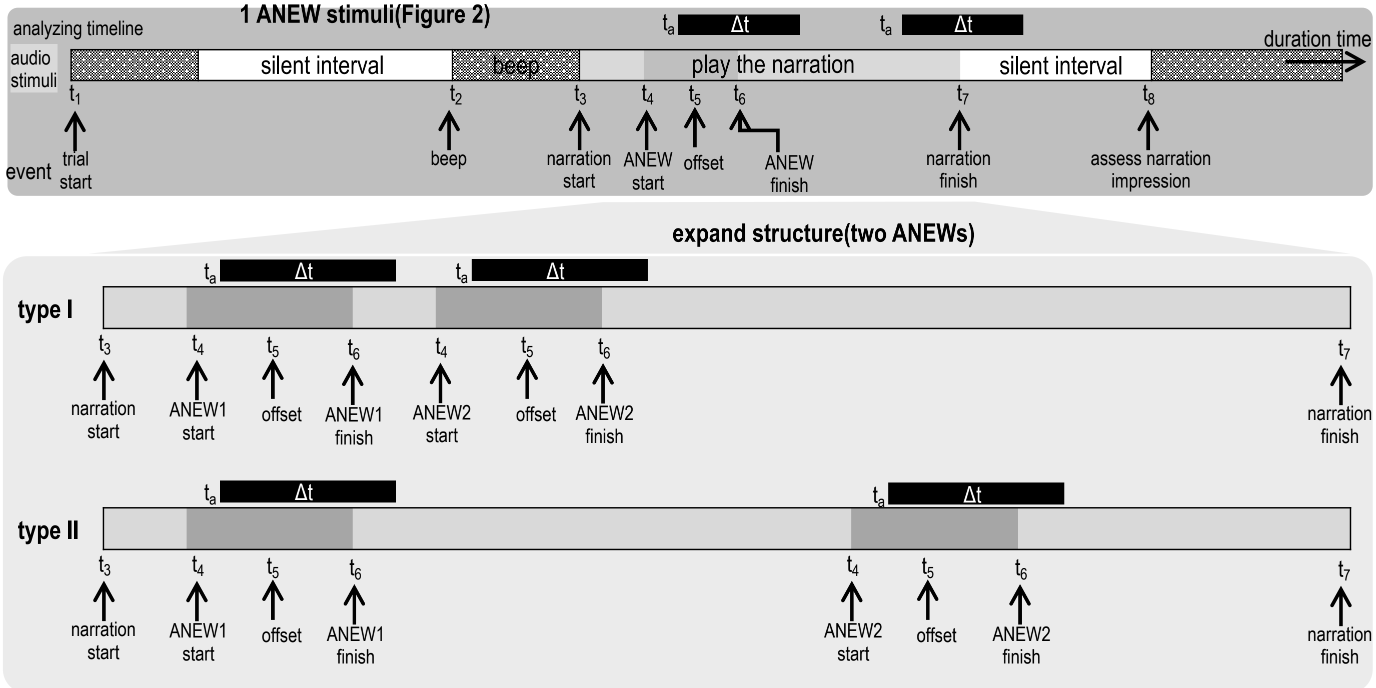

Figure 5. Timeline of presented stimuli including multiple ANEWs.

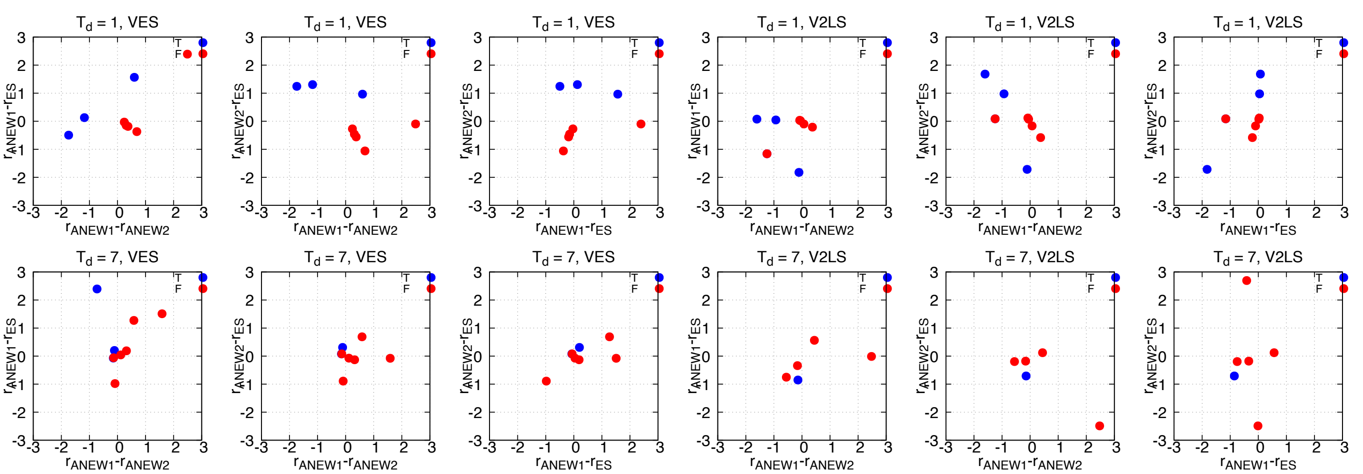

Figure 6. Measurement results between r_X. The horizontal and vertical axes are the differences between the observations in the combinations shown in the figure. Blue marks indicate that the corresponding presented stimulus was memorized, and red marks indicate that it was not memorized.

in Section II to capture the trend characterized by the local and global reactions. We devised the following two methods as candidates for a revised LGR-Map to examine its suitability for capturing trends. The purpose of the analysis is to estimate the appropriateness of the experimental design for defining the extended LGR-Map.

- Method 1: The graph was modified by taking $r_{ANEW1} - r_{ANEW2}$ on the horizontal axis to see the degree of the local impact for a pupillary response, and $r_{ANEW1} - r_{ES}$ and $r_{ANEW2} - r_{ES}$ on the vertical axis to see the impact of ANEW on the global pupillary response.
- Method 2: To directly assess the local-global influence, use $r_{ANEW1} - r_{ES}$ as the horizontal axis and $r_{ANEW2} - r_{ES}$ as the vertical axis.

In order to see the influence of ANEWs and induced emotion in each auditory stimulus, the data were classified based on the difference between the judged valence score S_p by participants for each stimulus and the valence $(V_{1/2})$ for each ANEWs in the stimuli. Some of the results are shown in Figure 6.

In describing the graphs, the relationship between the valence of ANEWs and S_p in auditory stimuli was checked as follows, and graphs were described for each category as listed below.

TABLE V. SUMMARY OF FEATURE FOR FIG.4, AND THE RESULTS OF THE RECALL TEST FOR AUDITORY STIMULI. THE RESULTS OF RECALL TEST ARE SUMMARIZED BY THE NUMBER AND PERCENTAGE OF THE AUDITORY STIMULI, WHICH ARE CATEGORIZED BASED ON FIGURE 4 TYPES, WHICH ARE MEMORIZED (MEM. TRUE) OR ARE NOT MEMORIZED (MEM. FALSE).

Type	$r_{ANEW1} - r_{ANEW2}$	$r_{ANEW2} - r_{ES}$	$r_{ANEW1} - r_{ES}$	Feature	Mem. True	Mem. False
A	$-$	$-$	$--$	$r_{ANEW1} < r_{ANEW2} < r_{ES}$	0.10(6)	0.08(5)
B	$--$	$+$	$+$	$r_{ANEW1} < r_{ES} < r_{ANEW2}$	0.05(3)	0.07(4)
C	$-$	$++$	$+$	$r_{ES} < r_{ANEW1} < r_{ANEW2}$ tend to mem.	***0.10***(6)	0.07(4)
D	$+$	$--$	$-$	$r_{ANEW2} < r_{ANEW1} < r_{ES}$ hard to mem.	0.07(4)	***0.19***(11)
E	$++$	$-$	$+$	$r_{ANEW2} < r_{ES} < r_{ANEW1}$ hard to mem.	0.02(1)	*0.08*(5)
F	$+$	$+$	$++$	$r_{ES} < r_{ANEW2} < r_{ANEW1}$ hard to mem.	0.05(3)	***0.12***(7)

$$\begin{aligned}
\text{VES}: &\quad |V_1 - S_p| < 1,\ |V_2 - S_p| < 1\\
\text{V1ES}: &\quad |V_1 - S_p| < 1,\ |V_2 - S_p| \geq 2\\
\text{V1LS}: &\quad 1 \leq |V_1 - S_p| < 2,\ |V_2 - S_p| \geq 2\\
\text{V2ES}: &\quad |V_1 - S_p| \geq 2,\ |V_2 - S_p| < 1\\
\text{V2LS}: &\quad 1 \leq |V_2 - S_p| < 2,\ |V_1 - S_p| \geq 2\\
\text{N/A}: &\quad |V_1 - S_p| \geq 2,\ |V_2 - S_p| \geq 2
\end{aligned}$$

Although there are 64 data points in total, only the categories that contain enough data for evaluation are shown, excluding those with a small number of data per category. Since the recall test was conducted at the same time, the r_X data for the memorized stimuli are shown in blue dots, and the r_X data for the not memorized stimuli are shown in red dots. The results suggest that the interval of 1 second between $ANEW_1$ and $ANEW_2$ and the VES category are the easiest to separate the memorized from not memorized stimuli.

Analysis of these results in comparison with Figure 4 leads to the following conclusions:

- $r_{ANEW1} - r_{ANEW2} < 0$, which means $r_{ANEW1} < r_{ANEW2}$,
- $r_{ANEW1} - r_{ES} \leq 0$, which means $r_{ANEW1} < r_{ES}$,
- $r_{ANEW2} - r_{ES} > 0$, which means $r_{ES} < r_{ANEW2}$.

Integrating all of them, we find that the situation might correspond to the situation defined in Table IV B or C. That is, the emotion initially generated by $ANEW_1$ is reinforced by $ANEW_2$ having the same type of valence, and the sentence presented in audio comes to an end before the emotional effect is fully extinguished.

Table V summarizes the results for all data. Column 1 in the table shows the graph type of Table IV, columns 2 to 4 show the positive and negative difference of r_X, and column 5 shows the relationship between r_X. The number of provided stimuli that were memorized was counted for each category of Table IV, and those that were memorized were described in column 6, Mem. True, and those that were not memorized were described in column 7, Mem. False. As a result, the state of Table IV C has a high degree of memorization, and the state of D or F has a low degree of memorization. The result indicates as follows:

$ANEW_1$ and $ANEW_2$, i.e., the distance between ANEWs in a provided auditory stimulus and the end of a sentence, should not be close together but should have a certain distance between them for easier memorization. ANEWs should include those with the same kind of valence as much as possible. One possibility is that the proximity of ANEWs generates a controlled valence, which is maintained for a long time if it exceeds a certain level of intensity, and that this valence, together with the concepts entered with the ANEWs, is stored and reinforced in the long-term memory. This is regarded to correspond to short-term learning. Therefore, in the recall test, the most recently learned items are presented in the order in which they were most recently learned, leading to improved memorization output.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we focused on the pupillary response and proposed the LGR-Map as a classification method for pupillary changes associated with emotional changes. The LGR-Map indicates whether an individual's pupillary response to a stimulus is more likely to respond to local informaion or contextual information.

In order to propose the LGR-Map, we needed a cognitive model that describes how people's emotions are induced in response to stimuli from the outside world. Thus, we constructed a model of human emotional responses based on the CI-model. We assumed that the input stimuli have the feature of auditory information, that is, transient information. At each time point, the auditory information was captured as a single packet of sound waves, and each packet was matched with the long-term memory only once. ANEW was assigned to one of the packets and assumed to appear somewhere in the sentence. We assumed a situation where the valence of the whole sentence is determined by the sentence-final expression. For each of them, we considered an emotional response appeared based on the CI-model. When considering the above situations, we thought that there was some kind of pupillary response for each emotional reaction. We described the pupillary response to ANEW as "local" response, and the pupillary response to the end of a sentence as "global" response. In this case, the local and global pupillary responses can be represented in a two-dimensional plane. Based on this idea, we proposed the LGR-Map. The shape classification of the LGR-Map was based on the variance of the error ellipsoid. The results indicated that the LGR-Map could be classified into five types according to the covariance of the local and global pupillary responses and the trend of the dispersion of the local pupillary response.

Based on a series of ideas, 36 auditory stimuli with various characteristics embedded in ANEWs and sentence-final expressions were actually given to 19 participants, and LGR-Maps were created. As a result, we confirmed that five types of shapes were recognized for individual pupillary responses. The effectiveness of the LGR-Map was confirmed by the fact that five different shapes were found for individual pupillary responses in 19 participants who were given 36 auditory stimuli with various characteristics embedded in ANEWs and sentence-final expressions.

As an extension of the LGR-Map, an application to the pupillary response in the case of multiple ANEWs was also considered and preliminary experiments were conducted. The results of preliminary experiment indicated the consistency with the Figure 1. Further analysis and experiment of auditory stimuli with multiple ANEWs will indicate further insights.

Image language plays a greater role than symbolic language in real-time communication. However, memes, which are words, play a major role in transmitting and accumulating knowledge over a long period [16]. Knowledge represented by a network, a meme, or a symbolic node, develops links to image nodes associated with it. Image nodes are formed in response to an individual's actual perceptual, cognitive, and motor experiences, represent something unique to each individual. This is very different from symbolic nodes, which are shared within a single culture. Communication through memes (words) is a form of communication through language nodes that aggregate a large amount of information, allowing for the exchange of a large amount of information with a small amount of information (words). This is achieved through the activation of image nodes that are associated with the language node. However, it is not guaranteed that the associated image nodes centered on the language node are consistent on both sides of the interlocutor. Therefore, errors in the transmission of information due to this discrepancy are inevitable [17].

For example, the proliferation of the Social Networking System (SNS) allows transmission errors to be amplified in an extremely short period of time. While considering these characteristics of SNS, it is necessary to establish a method to realize verbal communication that does not cause transmission errors and to build a community where people can communicate healthily. Toward this end, an approach that focuses on the activation of knowledge centered on language nodes, as shown in this study, is promising.

The application of the LGR-Map will make it possible to provide adaptive content for individuals. The method of implementation will be an issue in the future.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334 , and National University Management Reform Promotion Project. The authors would like to thank Editage (www.editage.com) for English language editing. MH also wants to thank to Nagai N · Promotion Foundation For Science of Perception for their finantial support.

References

[1] K. Nakahira, M. Harada, and M. Kitajima, "Local-global reaction map: Classification of listeners by pupil response characteristics when listening to sentences including emotion induction words -toward adaptive design of auditory information," in COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, 06 2023, pp. 79–85.

[2] N. Vallez et al., "Automatic museum audio guide," Sensors, vol. 20, no. 3, 2020. [Online]. Available: https://doi.org/10.3390/s20030779

[3] R. Hirabayashi, M. Shino, K. T. Nakahira, and M. Kitajima, "How auditory information presentation timings affect memory when watching omnidirectional movie with audio guide," in Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020), vol. 2, 2020, pp. 162–169.

[4] M. Murakami, M. Shino, K. T. Nakahira, and M. Kitajima, "Effects of emotion-induction words on memory of viewing visual stimuli with audio guide," in Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021), vol. 2, 2021, pp. 89–100.

[5] M. Murakami, M. Shino, M. Harada, K. T. Nakahira, and M. Kitajima, "Effects of emotion-induction words on memory and pupillary reactions while viewing visual stimuli with audio guide," in Computer Vision, Imaging and Computer Graphics Theory and Applications, A. A. de Sousa et al., Eds. Cham: Springer International Publishing, 2023, pp. 69–89.

[6] S. Moriya, K. T. Nakahira, M. Harada, M. Shino, and M. Kitajima, "Can pupillary responses while listening to short sentences containing emotion induction words explain the effects on sentence memory?" in Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2023, Volume V: HUCAPP, Online Streaming, February 19-21, 2023. SCITEPRESS, 2023, pp. 213–220.

[7] J. Z. Lim, J. Mountstephens, and J. Teo, "Emotion recognition using eye-tracking: Taxonomy, review and current challenges," Sensors, vol. 20, no. 8, 2020. [Online]. Available: https://doi.org/10.3390/s20082384

[8] L. Shu et al., "A review of emotion recognition using physiological signals," Sensors, vol. 18, no. 7, 2018. [Online]. Available: https://doi.org/10.3390/s18072074

[9] W. Kintsch, "The use of knowledge in discourse processing: A construction-integration model," Psychological Review, vol. 95, 1988, pp. 163–182.

[10] W. Kintsch, Comprehension: A paradigm for cognition. Cambridge, UK: Cambridge University Press, 1998.

[11] C. Wharton and W. Kintsch, "An overview of construction-integration model: A theory of comprehension as a foundation for a new cognitive architecture," SIGART Bull., vol. 2, no. 4, jul 1991, p. 169–173. [Online]. Available: https://doi.org/10.1145/122344.122379

[12] J. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, 12 1980, pp. 1161–1178.

[13] K. T. Nakahira, M. Harada, and M. Kitajima, "Analysis of the relationship between subjective difficulty of a task and the efforts put into it using biometric information," in Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 2: HUCAPP, Online Streaming, February 6-8, 2022, A. Paljic, M. Ziat, and K. Bouatouch, Eds. SCITEPRESS, 2022, pp. 241–248. [Online]. Available: https://doi.org/10.5220/0010906800003124

[14] P. Jerčić, C. Sennersten, and C. Lindley, "Modeling cognitive load and physiological arousal through pupil diameter and heart rate," vol. 79, no. 5, pp. 3145–3159. [Online]. Available: https://doi.org/10.1007/s11042-018-6518-z

[15] M. M. Bradley, L. S. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation." Psychophysiology, vol. 45, no. 4, 2008, pp. 602–607. [Online]. Available: https://doi.org/10.1111/j.1469-8986.2008.00654.x

[16] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. W W Norton & Co Inc, 2 2018.

[17] M. Kitajima, M. Toyota, and J. Dinet, "How Resonance Works for Development and Propagation of Memes," International Journal on Advances in Systems and Measurements, vol. 14, 2021, pp. 148–161.

Cyber Fusion in Financial Services

How Cyber Fusion Centres are being utilised in Australian Financial services organisations

Anne Coull
College of Science and Engineering
Flinders University
Sydney, Australia
anne.coull@proton.me

Elena Sitnikova
College of Science and Engineering
Flinders University
Adelaide, Australia
elena.sitnikova@flinders.edu.au

***Abstract*—Critical infrastructure organisations with operating environments spanning multiple domains and/or with multi-dimensional threats are looking for ways to accelerate response to cyber incidents. Cyber Fusion Centres are emerging as potential models for managing inter-related multi-domain and multi-dimensional threats. Through history, military fusion centres have facilitated integrated strategic operational multi-domain situational awareness in times of conflict. The Cyber Fusion Centre has evolved from a military and antiterrorist intelligence gathering centre to become an intelligence foci for collating information and facilitating cyber incident management in organisations. Some benefit is being realised in Australia's larger banks as they manage the challenge of coordinating cyber response across disparate and siloed teams. These simple Cyber Fusion Centres provide basic, manual, reactive coordination of cyber incidents by generating open communication between response teams. But the true visionary potential of the Cyber Fusion Centre models described in the literature are not yet being achieved. These theoretical centres of response-excellence incorporating strategic threat intelligence, orchestration, crisis simulations and real-time response capability are well beyond the current reality. Analysing the original military fusion centres to fully understand how these models function, and applying this more wholistic approach to implementing fusion for cyber closes the gap between theory and practice to deliver the anticipated benefits from Cyber Fusion Centres.**

Keywords- Cyber Fusion Centre; Intelligence; Counterinsurgence Operations; Counterterrorism; Crisis Management; Continuous Performance Improvement.

I. INTRODUCTION

As the coordination centre for cyber intelligence and response within an organisation, the Cyber Fusion Centre (CFC) appears to be the logical place from whence to drive accelerated response to cyber incidents. The literature describes the CFC as a collaboration between threat intelligence, incident response, threat hunting, and vulnerability management, with the purpose of accelerating identification and response to security threats [7][10][11]. A fusion centre of this nature enables an organisation to accelerate response, removing delays by orchestrating cyber response activities that span multiple departments and teams. It allows the organisation to be more proactive in their cyber response, identifying potentially large-scale threats by collating intelligence and observations from multiple teams and systems. A centre of this nature is in a unique position to see horizontally across and vertically within each aspect of the response process, enabling end-to-end response optimisation. Ultimately, the mature CFC facilitates more proactive threat response by mitigating threats as they are identified rather than just responding after the alerts have been generated, and the incidents have occurred.

The CFC emerging in Australian banks, and documented in a whitepaper by the Financial Services Information Sharing and Analysis Center (FS-ISAC), is a simple model of collaboration between security, service management, and customer service [15]. This model is aligned with equivalent CFC capabilities operating in banks and organisations in the United States, Canada, Singapore, and Australia. Utilising fusion in this way reduces potential threat impact by decreasing time to identify complex and critical incidents and time to respond. However, the implemented CFCs have not demonstrated the degree of uplift nor the benefits anticipated in the literature.

Section II outlines the evolution of fusion centres from military coordination centres to intelligent CFCs. Section III looks at the types of CFC. Section IV highlights the uplift in capability resulting from appropriate data and technology. Section V looks at the motivating factors influencing CFC creation in Australian Critical Infrastructure (CI). Section VI assesses how CFC have been implemented in Australia. Section VII outlines the factors limiting CFC capability and Section VIII provides insight into how the gap between the theory and reality can be closed to achieve the envisioned response capability uplift. Section IX surmises the current gap between theory and practice, and the future research to monitor the evolution of CFCs.

II. CYBER FUSION EVOLUTION

Fusion centres have operated as operations response coordination centres since mankind participated in multi-domain warfare. Over time, the Fusion Centre model has evolved into a centre for intelligence, co-ordination, and information sharing, in response to terrorist incidents and the rise of cyber-crime.

A. Military Fusion Centres

During the second world war, Winston Churchill directed the British and allied forces from the underground war room headquarters beneath London [17].

As the domains of war extended to include cyber space in the 1980s, Fusion centres have operated in the military as Joint Operations Centres, to co-ordinate operations across the multiple domains of war: land, sea, air, and space, and more recently cyberspace [2][7][13][19][23][24] (see Figure 1). Military fusion centres provided a strategic perspective of battles, facilitating coordinated information flow and driving greater efficacy in offensive and defensive operations spanning multiple regiments operating across the different locations and domains.

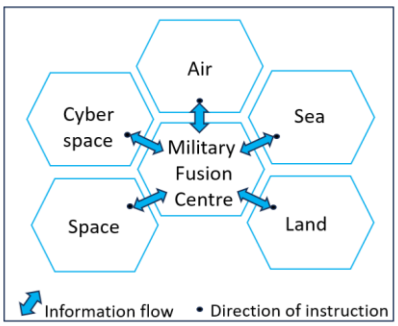

Figure 1. Military Fusion Centre model [1].

B. Counterinsurgency Operations' Intelligence Fusion and Flow

Insurgencies involve combinations of conflict and tactics across multiple domains, topographies and offensives, and counterinsurgency operations (COINOPS) need tangible real-time intelligence to stay abreast of enemy movements. This sensitivity is driven by COINOPS role working closely with civilian populations rather than conventional military forces. Counterinsurgency field commanders rely on local civilians to understand the complete geopolitical situation in which they are operating, including the insurgent actors and the motives for their behaviours. During counterinsurgency operations, this information needs to be disseminated from/to headquarters (HQ) and the field commanders in real-time. Rather than having all the intelligence capabilities centralised in military HQ, the key to the COINOPS model is to have technology and specialist personnel such as language translators and intelligence analysts, lower in the chain, implanted through all the layers from front line platoons and commanders to HQ (See Figure 2). This facilitates the flow of intelligence information, and generates greater situational transparency for the commanders at all levels. This model was demonstrated to be extremely effective in Iraq through 2006 and 2007 [19][28].

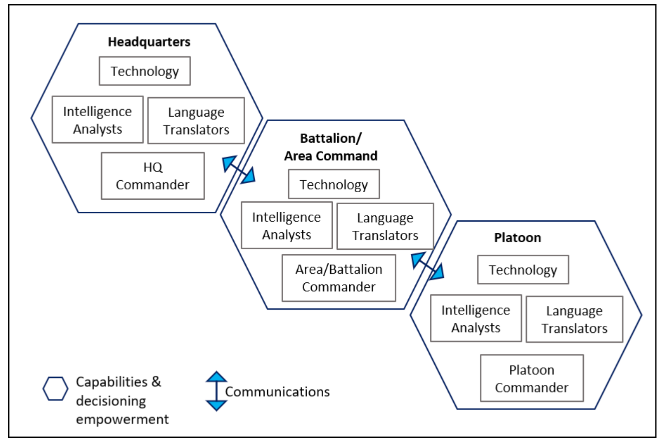

Figure 2. COINOPS model [19].

C. *Counterterrorism Intelligence Fusion Centres*

Following the New York twin tower attacks on September 11, 2001, fusion centres evolved from wartime and operational co-ordination centres into centres for collating and correlating terrorist intelligence. In the U.S.A., the Department of Homeland Security (DHS) was created at the national level, to bring together intelligence and law enforcement (See Figure 3). Correspondingly, law enforcement, public security, and emergency response were also centralised at the state level. The concept of a Fusion Process emerged, with the goal of implementing "risk-based, information-driven prevention, response, and consequence management programs" to "address immediate or emerging threat related circumstances and events" [12]. Fusion centres were created to connect the local and state intelligence centres with federal intelligence organisations and services. Aiming to prevent another successful attack on the U.S.A. through the open exchange and dissemination of analysed counterterrorism (CT) information from multiple intelligence sources [12][21][26].

This integrated model enabled more streamlined communications, collaboration, and coordination across intelligence, law enforcement and emergency response at the state and national levels, and provided actionable intelligence as the basis for security, public safety, and emergency response [26]. It was later recognised that these fusion centers also offered valuable foci "for coordinating the response to, and investigation of cyber-crimes and cyber threats against state assets and critical infrastructure. Thus emerged the Intelligent CFC [21].

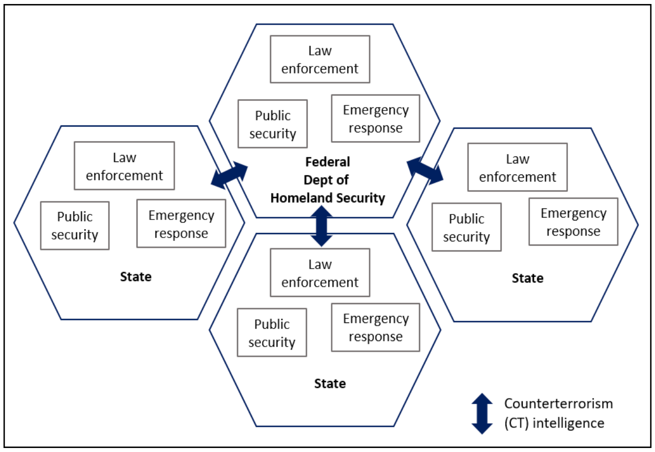

Figure 3. DHS Fusion Centre model [26].

D. Intelligent Cyber Fusion Centres

As security leaders moved from roles in military defence into business enterprises they saw the need, in their new organisations, for more efficient and effective intelligence-enabled cyber response and incident management. As a result, CFCs have been established in a number of larger organisations across the U.S.A. and in some of the larger Australian Banks, to more closely integrate cyber intelligence and operations

A CFC is described in the literature as a physical or virtual entity created through collaboration between threat intelligence, incident response, threat hunting and vulnerability management with the purpose of identifying, managing, and rapidly responding to security threats [25]. This may be a separate team, a virtual team with representation from the local response teams, or a blend, with a small group of individuals facilitating and coordinating aggregation, collation, and distribution of information across the participating teams, and analysing this integrated information to identify themes and correlations [2][8][9][26][27].

III. Cyber Fusion Centre Theory

Cyber Fuson Centres (CFC) are most relevant to organisations managing multi-domain and multi-dimensional cyber threats:

a) Multi-domain

Critical infrastructure organisations, such as energy, communications and transport manage cyber threats across multiple Information Technology (IT) and operational technology (OT) domains such as: power plants, satellites, fibre networks, railway tracks and signals, etc; along with the associated networking, connectivity and access management. The operating environments in the older banks established through last century, incorporate physical and virtual technology spanning mainframes, midrange, desktops, public and private cloud, and all the associated networking, connectivity and access management.

b) Multi-dimensional

Banking and finance enterprises manage cyber threats under many guises and with many dimensions. Along with the classic external cyber threats, these organisations are also managing insider threats, fraud, money muleing, money laundering, bribery and corruption, and regulatory requirements such as sanctions and politically exposed persons (PEPs). Combatting the added layer of criminal activity in these money-motivated instances requires close co-ordination with law enforcement agencies as, in many instances, the funds are linked to crimes including extortion, drug and human trafficking, child exploitation and investment scams.

Establishing a CFC in these complex organisations, that incorporates the different cyber teams, fraud, financial crime, service management and customer experience monitoring has the potential to deliver tangible benefits:

1. Accelerated detection of multi-domain incidents such as cyber-fraud scams and website spoofing;
2. Increased accuracy, speed, and reduced cost of incident response for simple incidents such as credential compromise and malicious phishing emails;
3. Increased accuracy, speed and reduced cost of incident response for multi-domain and multi-dimensional incidents such as mule accounts, extortion, blackmail, and money laundering;
4. Early intervention on incidents before they escalate to become major;
5. More accurate reporting showing the extent of multi-domain and multi-dimensional incidents;
6. Improved ability to track and analyse trends;
7. Improved opportunity for streamlining processes and building team synergies;
8. Reduced customer impact and reputational damage with corresponding improvement in customer confidence.

In addition, when industry peer organisations work together to create cross-organisation fusion, this provides broader insights into threats, threat actor behaviours and a deeper understanding of the mitigating responses being applied, which further enhances the opportunity for response acceleration [2][7]-[11][18].

A. The Cyber-centric Fusion Centre

The literature focuses primarily around Cyber-centric Fusion Centres and describes a capability that brings together:

1. Technical Threat Intelligence such as attack vectors, suspicious domains, malware hashes, and exploited vulnerabilities to assess the cyber threats facing the organisation;
2. Strategic Threat Intelligence to map attack trends, motivations and characteristics;

3. Analysis of this intelligence to generate insights about threats and adversary behaviours, tactics, techniques and procedures (TTPs), and indicators of compromise (IOC) [2][7][8].
4. Cyber incident management [11] to co-ordinate incident response activities that span multiple teams and organisational divisions.

As it matures, the CFC described extends to incorporate:

1. Security orchestration, automation and response (SOAR), with automated operational workflows to facilitate incident triage, threat pattern analysis, and automated threat response capability;
2. Response plan testing, and crisis simulations to prepare for major incidents; and
3. Short and long-term recovery planning [7][8][26] (See Figure 4).

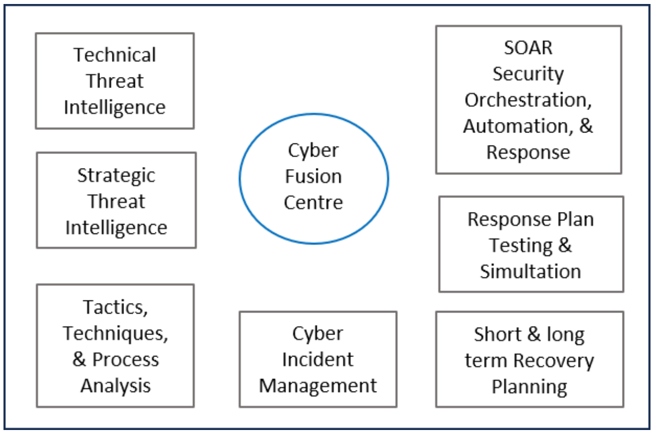

Figure 4. Intelligent Cyber Fusion Centre model [8].

B. Cyber Fusion in Financial Services

A model for Financial-services-centric CFCs has been developed by the Financial Services Information Sharing and Analysis Centre (FS-ISAC) [15]. FS-ISAC is a collaborative not for profit venture whose mission is to "advance cybersecurity and resilience in the global financial system, protecting financial institutions and the people they serve" [14]. The 2023 whitepaper released by FS-ISAC and authored by a subcommittee of its members, provided recommendations for establishing and implementing a CFC in a bank. According to the FS-ISAC whitepaper, the CFC's primary benefit is derived from sharing information during an incident, by "synchronising response activities across different regions, business units, and other fusion centers." In addition, the whitepaper highlighted that the CFC establishes a common language, streamlining communications between responders and leadership prior to and during security events, and improving c-suite risk reporting [15].

The expected benefits revolve around the resultant uplift in response capability based on:

- "Standardised, repeatable, incident response and management processes;
- Enhanced transparency into tactical reactions to events;
- Dedicated, trained, and experienced incident commanders;
- Improved adherence to regulatory disclosure requirements;
- Demonstrated overall security posture to regulators/clients/and executives" [15].

a) Fusion Centre Participants

The FS-ISAC CFCs whitepaper (2023) described a centralised, co-located or distributed, virtual model focused on response and incident management, where multiple areas in the business are impacted [15] (See Figure 5). They recommended the core participants in the fusion centre include representatives from:

- Security Operations Centre (incl. Cyber & Technology)
- Incident & Crisis Management
- Fraud Management
- Physical Security
- Intelligence
- Third Party Management
- Communications
- Legal, and Compliance.

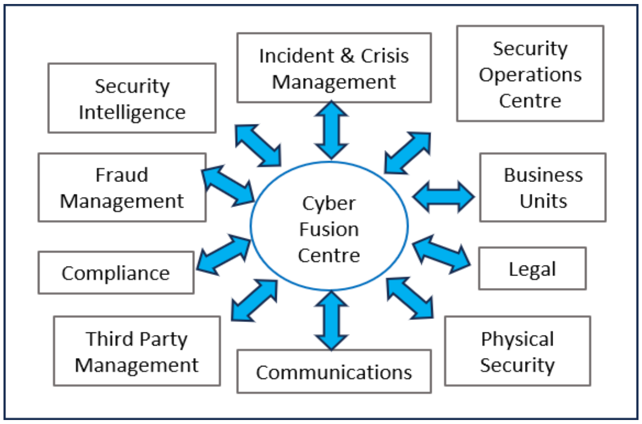

Figure 5. FS-ISAC Fusion Centre Model, based on [15].

A secondary group of participants were recommended to participate when an incident is relevant to their areas of responsibility. These secondary members include:

- Accounting
- Anti-Money Laundering (AML)
- Business Continuity
- Digital Protection & Forensics
- Data Privacy / Breach Incident Response
- Human Relations
- Group Insurance
- Internal Investigations (Insider Threat)
- Risk
- Public Relations
- Security Architecture
- Security Awareness
- Service Management (e.g., Payments, Customer Service, Internet Banking), and
- Vulnerability Management [15].

b) Implementation Model

The FS-ISAC paper outlined the method for implementing a CFC starting with a daily standup, where participants share observations and insights from the previous 24 hours. The purpose of the daily standup is to facilitate collaboration between participating teams, capture the updates they provide. Participants raise items of interest, question one another, and look for common areas of interest. The coordinating CFC team documents and tracks items raised and actions involving multiple participating teams. As the CFC matures, trends and patterns may be identified and tracked [15].

IV. FUSION ENABLING DATA & TECHNOLOGY

The strength and benefits of fusion centres comes from their ability to bring information together, in such a way that the whole is greater than the sum of the parts. Actionable intelligence comes from collating and analysing intelligence from multiple sources [25]. Fusion generates the complete threat-picture by bringing together the threat elements, or puzzle pieces [20] of techniques, tactics, and processes that only become clear when the threat is viewed from multiple perspectives. This can be achieved manually, through the standups, when participating representatives from different areas of security, fraud, service management and customer service share their observations and insights. But for accelerated fusion targeting real-time detection, supporting technology is needed. The technology platform's role is to collate the threat elements, apply pre-defined algorithms to analyse them, identify patterns and correlations, and when a defined threshold is exceeded, actionable intelligence in the form of alerts enriched with supporting information are generate, so these alerts can be actioned. Unless the alerts are analysed, filtered and prioritised, the response teams are at risk of viewing every threat as a priority, or becoming overwhelmed by the mass of alerts and missing the important ones [22][25].

Short term benefit can be realised by publishing alerts from specialist detection systems more broadly across the CFC participating teams, for example, sharing fraud and cyber-crime alerts between these respective teams.

Figure 6 illustrates the elements and interactions within the fusion enabling technology described herein. Fusion-enabling technology encompasses:

a) Data Storage

A data store, such as data lake is used to store large volumes of data from a range of sources. These may be in any format, from excel and comma delineated flat files, through to structured relational database models. Examples include intelligence feeds, vulnerability scan results, customer and employee data.

b) Utilising existing data sources

The data lake is used to bring together the data from its various sources. Existing systems will have their own data stores. Ideally, this data is ingested directly from the source system into the Fusion Data Store [DHS].

c) Intelligent Analytics Platform

In order to interrogate and analyse information from multiple sources, the Fusion Centre requires extensive analytical, discovery, and entity mapping capabilities. The scope of the analytics platform capabilities required for an organisation's fusion centre will depend on what is already available through their existing tools and platforms. As a minimum, the CFC's analytics platform needs to be capable of mapping relationships between entities sourced from multiple data sources, provide risk ratings and insights on these entities, and generate alerts where the risk rating exceeds a pre-determined threshold.

Enabling entities of interest to be discoverable will make it easier for the CFC to work with law enforcement agencies (LEA) such as the Australian Federal Police and the state level Crime Squads. Integrability, flexible data ingestion, and configurable modelling capabilities are fundamental. Additional capabilities include case management workflow, artificial intelligence and machine learning.

d) Integration

Bi-directional integration between the Analytic Platform and the data store is essential, as is integration between the data store and the source systems. Integration of the data store and/or the analytics platform with existing decisioning engines, AI tools, reporting, case management, and access management utilities will all depend on the architecture of the systems within the organisation. The platform and its capabilities will need to be in alignment with the organisation's policies and standards.

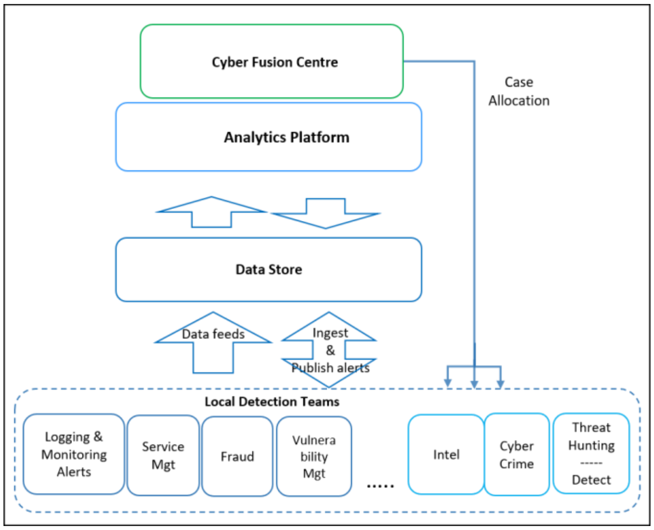

Figure 6. Fusion Enabling Technology

e) Utilising existing decision engines

Rather than recreating everything from scratch, the fusion analytics platform and data store can integrate with existing analytics capabilities. These may be machine learning tools, or AML and sanctions decisioning engines. By ingesting and publishing alerts, these analytic engines can operate collaboratively, identifying risks that only become visible by bringing information together.

f) Real-time vs Batch

Response times in cyber defence are a critical measure. Waiting overnight for a batch job will provide plenty of scope for a cyber threat actor or to move laterally through the systems domain before being found, or for a financial criminal to gain approval for a loan or credit card prior to being identified. Deeper analytics may take longer, but real-time fusion processing and risk rating is a must.

V. Motivating Factors for Establishing Cyber Fusion Centres in Australian Critical Infrastructure

The motives for implementing CFC in Australia appear to relate primarily to the scale and complexity of the organisation, perceived ease of internal communications, and the degree of scrutiny from regulators.

A. *Size and Complexity Matters*

Industry research indicates that only the large scale banks in Australia have implemented, or considered implementing CFCs at this time. In these large-scale organisations, the complexities of communicating between multiple teams who participate in cyber, fraud, and service management incident detection and response, with their different perspectives and priorities, has hampered fluid information flow. The large security departments that have evolved in these banks naturally segregated into silos, with each team focusing on their local accountabilities [9].

B. *Smaller Scale*

Smaller organisations have relied on open communications and close interpersonal relationships when coordinating their response efforts, but this is not scalable. The smaller scale organisations that were assessed in the energy and financial sectors did not see a need for a CFC, as communications and coordination during priority incidents was straightforward. Analysis found that the communications within smaller, less complex organisations, such as those within the insurance and energy sectors, is naturally more open and less arduous. With only a handful of individuals involved in incident management and cyber response, it has been easy for each participant to have a deep understanding of their own area of accountability, as well as visibility across the cyber and business landscape. In these smaller organisations, there is less opportunity for information to fall through the gaps.

C. *Regulator Attention*

Australian banks and financial services organisations have received close scrutiny from the regulators such as AUSTRAC and APRA for more than 2 decades. This has provided the impetus that has driven these organisations to invest in uplifting their cyber security capabilities.

The introduction of the Security of Critical Infrastructure act (SOCI), and the increased emphasis on cyber security posture across all critical infrastructure (CI) organisations reflects the shift in focus of the regulators from banks to all CI [6]. As the regulators increase the pressure across the communications, energy, transport sectors etc, there is increased motivation for these organisations to lift their cyber security capability. CFCs may emerge in these sectors as a result.

VI. CYBER FUSION CENTRE IMPLEMENTATIONS IN AUSTRALIAN CRITICAL INFRASTRUCTURE

The few fusion centres in Australia are concentrated in the larger banks. These organisations are highly complex, heavily regulated, and potentially lucrative targets for threat actors and criminals [3]-[6][9]. Two large Australian Banks that have implemented CFCs were analysed. In these organisations, the CFC has played a role in bridging the gaps across disparate teams, facilitated open communications, and created an integrated perspective for cross-domain and cross-dimension response activities.

The first of the big four banks to implement a CFC established a virtual capability where people from different teams across security came together to facilitate incident response. This virtual model, while successful in achieving greater collaboration and focus in incident response was disbanded and then reformed when the Chief Information Security Officer (CISO) role changed hands.

The CFC has become more operative as it progressed through different iterations, starting with representatives from the cyber teams and evolving into a small, separate team. Along the way, it has experienced many challenges. Initially, the attendees all had day jobs so there was no-one responsible for prioritising the incidents that were identified, nor anyone assigned to capture or document them. As such, no reports were generated and the insights were only available to those present at the daily catchups. This limited the ability for others to contribute or benefit.

In addition, there were problems with managing sensitive and classified information, as the core cyber team were reticent to offer insights on situations they considered sensitive, when non-cyber participants were present. As a result, discussion became stifled, shallow, and limited, and few insights or actions emerged.

In the second Australian banks, the CFC was established with an initial focus on facilitating information flow. A new team was added alongside the detect and response teams. The fusion team coordinated daily communications forums each morning, with representatives from the different teams across cyber and physical security, fraud, IT service management priority incident response, crisis management, supplier management, and customer service (See Figure 7). Attendance included analysts through to general managers. Prior to the creation of the CFC, these specialised teams had been functioning independently, with information flow only within the core cyber teams. Coming together daily to share updates and insights with the broader group, on what they had seen in the previous 24 hours, facilitated greater transparency and visible cooperation between the teams. The CFC team was active in encouraging this cooperation, involving themselves when an incident spanned multiple domains and/or dimensions.

Beyond initial benefits elicited from the sharing of insights and improved cooperation, the value derived from the CFC has been limited. While the non-cyber teams shared their experiences openly, the core-cyber teams were slow to open up and continued to show resistance to imparting any real information. The Vulnerability Management team shared some insight into CVEs and the corresponding vulnerabilities and response actions, but the updates provided by the remaining cyber participants did not include detailed technical threat intelligence regarding the threats facing the bank, existing or missing controls, alert details, nor strategic threat intel showing trends, motivations and characteristics, and adversary behaviours. Similar to the first instance, this reticence effected the depth of discussion and the level of situational awareness across the participants, which continued to be limited and localised. Further work was needed to develop trust and a sense of shared purpose for the cyber teams.

The expected benefits from the CFC, such as accelerating threat response, were not yet being seen. The CFC had not played a role in developing SOAR capabilities, nor had they made plans to facilitate practice sessions in preparation for major or significant incidents, nor were they involved in short- and long-term recovery planning. While the CFC team supported incident management spanning multiple domains, the majority of cyber incident management continued to be accomplished locally within the specialist teams.

Despite a higher level of investment, resourcing, and visible management support, the capability and performance of this less mature CFC has been hindered by the inexperience of the CFC leader and their lack of knowledge and understanding of cybersecurity, fraud, and/or financial crime. In addition, the absence of a rousing vision for the CFC, coupled with a lack of direction and an inability to lead diverse teams and drive organisational change through inspirational leadership has stymied the CFCs ability to advance.

Without a clear vision and roadmap to propel the team forward, in this instance, the CFC operated reactively and at the task level. Continued aversion to implementing performance measures or to be involved in end-to-end process optimisation has restricted their ability to give outcome-focus their actions. This lack of strategic direction and value-add will need to be addressed if they are to demonstrate return on investment (ROI).

Neither of the bank's CFCs were enhanced with comprehensive, integrated data analytics capabilities or supported with intelligent-driven technology. The processes were mainly manual in nature, relying on regular human intervention.

Other forms of CFC were observed in different type of financial institutions. In a credit card organisation, a 2-person CFC team met with their participants monthly and focused on reviewing the events of the preceding period. The CFC summarised the events and provided a report for senior management to digest. This CFC appeared to lack the visible senior management backing they needed to drive greater levels of participation and more active interventions, and so they had become a reporting function.

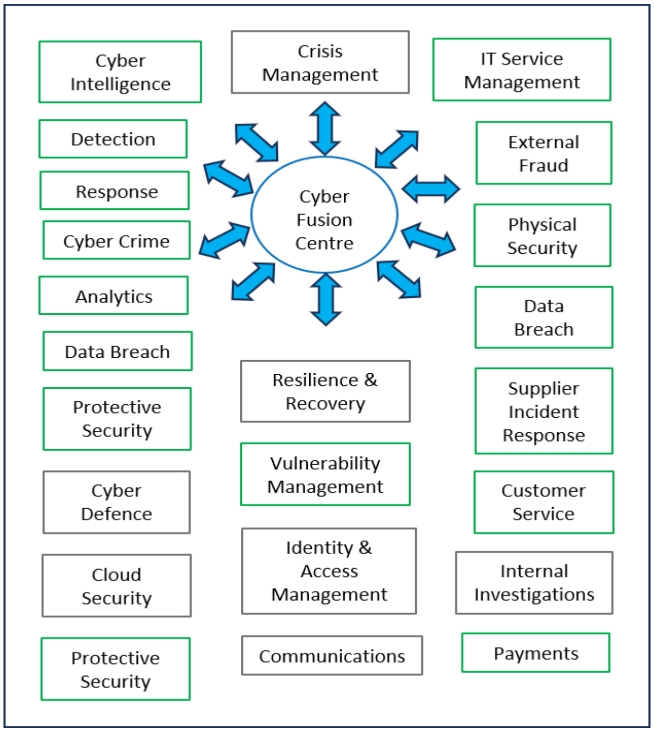

Figure 7. Cyber Fusion Centre in Australian Bank.

A. *Crisis Management*

In the Australian bank, where CFC facilitated a daily standup with representatives from across the different from the areas illustrated in Figure 7, observations were discussed, insights shared, and areas of overlap and interdependence highlighted. Where interdependencies were more complex and broader-reaching incidents were revealed, the CFC team stepped up and facilitated a more integrated response approach.

High Priority cyber incidents emerging from these collaborative sessions, whose scale of impact or potential impact exceeded an agreed threshold, were handed over to a Crisis Management Team (CMT). Crisis Management coordinated activities across IT Major Incident Management, cyber security, the affected business areas, and the internal and external communication teams. This ensured the crisis situation was prioritised, and appropriate resources were applied to accelerate recovery and minimise customer impact. It also ensured senior leadership, customer facing stakeholders and customers were aware of the outage and kept abreast of progress. The CFC team provided day-to-day support to the Crisis Management Team during a crisis situation (See Figure 8).

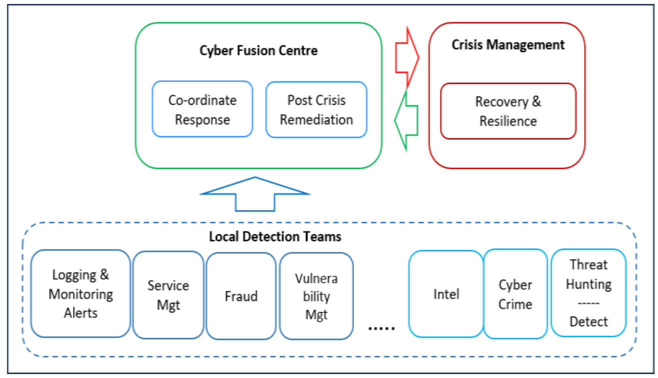

Figure 8. Fusion and Crisis Management in an Australian Bank

B. *Vulnerability Remediation*

Vulnerabilities and remediation requirements identified through this bank's Crisis Management process were captured and documented. These vulnerabilities were prioritised, funded, and remediated to ensure similar situations were not repeated. Many of these vulnerabilities had been previously reported by the accountable operations teams prior to the incident, but they had not been prioritised for funding. These larger scale incidents and the resultant crises, provided appropriate visibility and senior management focus to the potential risks, and the funding followed. The CFC was observed providing oversight and progress reporting on these remediation projects.

VII. Factors Limiting Cyber Fusion Centre Capability

The CFCs ability to influence incident response and containment times was constrained by several factors:

a) Regularity and Timeliness of Standups

The CFCs who met daily were timelier in their ability identify cross-team interdependencies, take action to intervene and thus affect response times. Those that ran weekly or monthly sessions operated primarily as a reporting and review function where events in the prior period were summarised, briefly discussed, and reported.

b) Visible Senior Sponsorship and Attendance

The priority given to the CFC, and the corresponding participation level directly correlated with the seniority and visibility of the sponsor support. Attendees joined more consistently and discussion was more open when senior leaders, such as General Managers, attended regularly, and asked questions in the CFC standup sessions.

c) Transparency, Openness and Trust

Reticence of the cyber intel teams to share details wask consistent across organisaitons. To facilitate and encourage open sharing of information and insights, the CFC need to develop a sanctum of trust amongst participants. To enable transparent and productive discussions, they will need up-front agreement from cyber leaders that it is ok for their representatives to share specific intel information regarding threats, threat actor behaviours and impacts or potential impacts to the organisation.

d) Resourcing and Accountabilities

CFCs operated better when specific individuals were accountable for leading the CFC. These people must have adequate capacity to prioritise and follow-up on incidents, document and distribute summary reports, and perform trend analysis. This may be achieved by having separate resources allocated specifically to the CFC team, or by ensuring adequate capacity and accountability for representative cyber team members participating in the CFC.

e) Cyber Experience

The level of cyber capability of the CFC leaders and team members facilitating interactions between the participating teams directly impacted the CFCs ability to influence. CFCs led by representatives from the cyber teams themselves generated greatest participation and information sharing. Those facilitated by people with little or no cyber knowledge failed to dig into issues and performed minimal, if any, root cause analysis.

f) Follow-up Report and Actions

Those CFCs that provided a brief summary of the discussions within the catch-up session, including updates from any follow-up actions, enabled greater participation from CFC members otherwise occupied during that session. This provided greater transparency and ensured all stakeholders could stay abreast of updates and insights.

g) Data Integration and Intelligent Analytics

Technology is the ultimate enabler for bringing together from multiple sources and perspectives and overlaying real-time data modelling and analytics Without access to real-time data and enabling technologies the CFCs remained manual and under-developed.

h) *Accumulative Effect*

These factors were accumulative. The ideal results observed from daily CFC sessions being facilitated by very experienced cyber personnel, with good communications skills, participation from General Managers, visible senior sponsorship, summary reports, adequate resourcing, and follow-up actions, enabled with real-time intelligent analytics.

VIII. Realising Cyber Fusion Centres Potential: Addressing the Gap

The lack of maturity of the existing CFCs in Australia is reflected in the limited benefits they deliver. These fall far short of the goal, but the capability uplift described in the literature is attainable. The keys to addressing the gap between CFC theory and practice can be found in the fusion models that have been most successful: The COINOPS intelligence fusion and flow model, and the DHS fusion centre guidelines are readily applicable to organisations managing the risks associated with cyber and criminal activity. These models highlight the need for:

a) A Shared Vision

The COINOPS commander in the field is clear on their direction, with a strong vision of the mission objectives. The vision of a cohesive mature CFC function, that brings together every aspect of cyber: intelligence; vulnerability management; detection; response; and recovery, with technology, and customer support, for complete situational awareness, is exhilarating. The CFS vision needs to be clearly and inspiringly communicated from the top echelons of leadership through the CFC leader, to the analysts and response teams working day-to-day with the CFC [12].

b) The Right Skills and Leadership Capability

CFC effectiveness relies on the right mix of skills and capabilities, in the same way the COINOPS effectiveness relies on the right mix of skills for intelligence fusion and flow. The effective COINOPS platoon in the field incorporates both military specialists and professionals who understand the environment, with language and technology specialists, and intelligence analysts who generate situation awareness [11]. The platoon commander's understanding of the civilian and military context, in that moment, in the field, is crucial. Their depth of experience and capability is reflected in their ability to lead a diverse team of specialists, through challenging situations; distilling intelligence, providing direction; and retaining grasp of the goal while flexing to fit with the constantly changing circumstances [12][19]. The fusion centre leader requires an equivalent level of contextual appreciation, depth of leadership capability and experience, focus on outcomes, and the ability to distil information and lead diverse teams of specialists through potentially challenging situations.

c) Clear Information Flow and Accountabilities

The DHS Counterterrorism fusion model illustrates how different accountable teams can be brought together into Fusion Centres to work more collaboratively and to facilitate information flow from state to the national level [26]. The COINOPS model has taken this to the next level, accelerating the flow of information and intelligence bi-directionally through the layers of command to enable the field commanders to make informed decisions real-time [19]. Similarly, the efficacy of Cyber Fusion and Incident Response in organisations relies on clear accountabilities along with fluid and transparent flow of intelligence information, vertically and horizontally through the organisational [12].

d) Robust Strategy and Roadmap

Turning the CFC vision into reality relies on having a roadmap that outlines the steps to get from the current, manual, reactive reality, to the proactive, informed real-time intelligent fusion analytics and integrated response capability. This roadmap needs to include all the relevant changes for policies, processes, technology and people.

e) Fusion Enabling Technology

The ultimate objectives of the CFC will only be achieved when intelligent analytics and modelling is applied across merged transaction data, alerts, and intelligence feeds from the multiple source domains. Generated insights will then highlight risks and issues not previously visible to the constituent teams.

Significant performance uplift can be attained by strategically utilising existing technology and intelligence already available within the organisation, to facilitate situational transparency and awareness across the response teams. To perform at its best as it matures, CFC will be required to leverage technology for timely information flow, integrated intelligence analytics, and orchestrated response capability [12]. Trend analysis and problem management can then be used to identify endemic issues, analyse and address the root cause(s).

f) Practice

Simulations allow processes to be refined and skills uplifted. The CFC is in an ideal position to plan and implement controlled simulations to address areas of weakness, test new process, and enhance capability and confidence in managing scaled cross-domain and cross-dimensional incidents. This capability can sit side-by-side with Crisis Management, enabling the CFC to manage incidents before they escalate to crisis level and reduce impact to customers.

g) New Ways of Working

Influencing stakeholders to overcome resistance to the new ways of working is the most challenging aspect of building a fusion centre. The CFC is a shared responsibility with potential benefits that span the business. Effective organisational change management, with visible senior-leader sponsorship, and hands-on and capable leadership from the CFC, will inspire and encourage teams to participate, learn from one-another, build mutual trust, and share in the collective gain of fusion [8][9].

h) Tuning the Model to Fit

The literature describes a CFC as a command-centre coordinating response activities, but in large complex organisations with established and experienced cyber response teams, the CFC team are facilitators, influencers, and ultimately drivers of improved response through their unique cross-silo perspective.

i) Performance Measures

Performance measures help team members to focus on the elements that make a difference. To demonstrate how the CFC can affect cyber threat response times, performance metrics such as: the Mean Time To Detect (MTTD), Mean Time To Respond (MTTR), and Mean Time To Contain (MTTC) need to be baselined and tracked [11]. Improvements in these measures will highlight the CFC's value, as well as point to areas requiring their attention.

j) Collaborative Space

Collaboration is the back bone of cyber fusion, both within the organisation, and between the CFC and other centres and cyber partners. Creating a collaborative, safe environment, where the CFC have a deep understanding of the environment and participants can openly contribute through sharing information and intelligence insights, will ensure a constant flow of information, peer to peer learning, and a continuous uplift of capabilities [12][16][25].

k) Continuous Learning and Improvement

Cyber Security, as a field, is constantly evolving: Technologies change, threat patterns change, threat actor techniques, tactics and processes constantly evolve. New vulnerabilities emerge every day with threats closely behind. The CFC is in the unique position of seeing horizontally across and vertically within the participating response teams. By working with these teams actively seeking to uplift and streamline processes, they can sustainably improve response outcomes.

l) Enduring Value and Funding

Secure adequate funding to establish, maintain, and sustain the CFC is essential for the CFCs continued existence. This will include people in the CFC, licensing costs for technology platforms and tools, technology delivery and maintenance intelligence feeds, if applicable, and more.

IX. CONCLUSION AND FUTURE RESEARCH

The CFC holds great promise for organisations faced with coordinating multiple divisions, departments, and domains when responding to cyber incidents. The literature paints a picture of CFCs as hubs of intelligence, knowledge, and response coordination excellence; where expertise comes together to problem solve and drive actionable outcomes. The reality is much simpler and more basic. The CFCs described in the FS-ISAC whitepaper and being implemented in Australian banks, focus on basic manual and reactive response coordination through daily standups where representatives share their observations and insights with one another. While this has provided some benefit through great cross-team transparency, it is not delivering the anticipated improvements.

Building a mature intelligence-enabled cyber fusion capability and realising the associated benefits, requires visionary and strategic leadership, a broad appreciation of cyber security in all its aspects, an ability to engage and inspire cyber professionals to join-in, and a deep understand of the problems the fusion centre is addressing, along with the skills, technology, data, and investment to make it happen.

Future research will monitor the evolution of these CFCs, as well as other approaches and factors contributing to accelerated cyber response in critical infrastructure organisations.

References

[1] A. Coull, “Fusion or Fantasy: Is Cyber Fusion Living up to the Dream?,” Proc. CYBER 2023, pp. 38-45. Available from: https://www.thinkmind.org/library/CYBER/CYBER_2023/cyber_2023_1_70_80040.html, accessed June 2024.

[2] Anomali, “What is a Cyber Fusion Centre,” Available from: https://www.anomali.com/blog/what-is-a-cyber-fusion-center, accessed July 2023.

[3] APRA, “APRA 230 Operational risk management,” Available from: https://www.apra.gov.au/operational-risk-management, accessed February 2024.

[4] APRA, “APRA 234 Information Security,” Available from: https://www.apra.gov.au/information-security, accessed February 2024.

[5] AUSTRAC, “New financial crime guides,” Available from: https://www.austrac.gov.au/new-financial-crime-guides, accessed February 2024.

[6] CISC, “Security of Critical Infrastructure Act 2018 (SOCI),” Available from: https://www.cisc.gov.au/legislation-regulation-and-compliance/soci-act-2018, accessed February 2024.

[7] Cyware, “What is a Cyber Fusion Center Center and how is it different from Security Operations Center (SOC)?,” August 2018, Available from: https://cyware.com/security-guides/cyber-fusion-and-threat-response/what-is-cyber-fusion-center-and-how-is-it-different-from-security-operations-center-soc-b13a, accessed May 2023.

[8] Cyware, “Building a Cyber Fusion Center, November 2020,” Available from: https://cyware.com/educational-guides/cyber-fusion-and-threat-response/building-a-cyber-fusion-center-ae08/, accessed May 2023.

[9] Cyware, “Why are Financial Institutions Adopting Cyber Fusion Strategies,” May 2022, Available from: https://cyware.com/security-guides/cyber-fusion-and-threat-response/why-are-financial-institutions-adopting-cyber-fusion-strategies-57b5, accessed May 2023.

[10] Cyware, “How Can You Improve Your Security Posture with Cyber Fusion?,” June 2022, Available from: https://cyware.com/security-guides/cyber-fusion-and-threat-response/how-can-you-improve-your-security-posture-with-cyber-fusion-3afb, accessed May 2023.

[11] Cyware, “How Cyber Fusion Provides 360-degree Threat Visibility?,” July 2020, Available from: https://cyware.com/security-guides/cyber-fusion-and-threat-response/how-cyber-fusion-provides-360-degree-threat-visibility-8fda, accessed May 2023.

[12] DHS, “Fusion Center Guidelines Developing and Sharing Information and Intelligence in a New Era,” U.S. Department of Homeland Security, Issued August 2006, Available from: https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/fusion_center_guidelines_law_enforcement.pdf, accessed August 2023.

[13] D. P. Fidler, “Inter arma silent leges Redu?: The law of armed conflict and cyber conflict,” Cyberspace and national security threats, opportunities, and power in a virtual world, pp. 71-88. Georgetown University Press / Washington, DC 2012, Available from: https://www.jstor.org/stable/j.ctt2tt6rz, accessed June 2024.

[14] FSISAC, “Financial Services Information Sharing and Analysis Center,” Available from https://www.fsisac.com/, accessed June 2023.

[15] FSISAC Fusion Council, “Considerations for Implementing a Fusion Operating Model in Financial Services– Whitepaper,” Financial Services Information Sharing and Analysis Centre (FS-ISAC) May 2023. Available from https://www.fsisac.com/, accessed June 2023.

[16] Global Advisory Committee, “Cyber Integration for Fusion Centers, Global Advisory Committee (2015),” Available from: http://www.cisecurity.org/documents/CyberIntegrationforFusionCenters_000.pdf, accessed August 2023.

[17] Imperial war rooms, “Visit Churchill War Rooms,” Available from: iwm.org.uk, accessed February 2024.

[18] K. L. Mclaughlin, “Cybersecurity and fusion centers,” The EDP Audit, Control, And Security Newsletter 2023 Vol. 67, No. 4 2023, Available from: https://www.tandfonline.com/doi/pdf/10.1080/07366981.2023.2205689, accessed June 2023.

[19] A. Mishra, “Synchronising Counterinsurgency Ops with Effective Intelligence,” Available from: https://theforge.defence.gov.au/publications/synchronising-counterinsurgency-ops-effective-intelligence, accessed July 2023.

[20] NEC, “NEC cyber fusion – actionable intelligence,” NEC Corporation 2015, Available from: https://uk.nec.com/en_GB/en/global/solutions/safety/pdf/NEC_Cyber_Fusion.pdf, accessed August 2023.

[21] NGA, “Enhancing the Role of Fusion Centers in Cybersecurity,” National Governors Association, 2019, Available from: https://www.nga.org/wp-content/uploads/2019/04/1507EnhancingTheRoleOfFusionCenters.pdf, accessed August 2023.

[22] A. Reeves and D. Ashenden, “Understanding decision making in security operations centres: building the case for cyber deception technology,” 2023. Available from: https://www.researchgate.net/search.Search.html?query=security+operations+centre&type=publication, accessed June 2023.

[23] S. Reveron, “An introduction to National Security and Cyberspace,” Cyberspace and national security threats, opportunities, and power in a virtual world, pp. 3-20. Georgetown University Press / Washington, DC 2012, Available from: https://www.jstor.org/stable/j.ctt2tt6rz, accessed June 2024.

[24] J. B. Sheldon, “Toward a theory of cyber power: Strategic purpose in peace and war,” Cyberspace and national security threats, opportunities, and power in a virtual world, pp. 207-224. Georgetown University Press / Washington, DC 2012, Available from: https://www.jstor.org/stable/j.ctt2tt6rz, accessed June 2024.

[25] K. Shouse, "Actionability of cyber threat intelligence," Utica College December 2015, Available from: https://www.proquest.com/docview/1739017603/fulltextPDF/AF5BEC7091EE448CPQ/1?accountid=10910, accessed August 2023.

[26] J. E. Steiner, "Needed: State-level, Integrated Intelligence Enterprises," Studies in Intelligence Vol. 53, No. 3 (Extracts, September 2009), Available from: https://www.cia.gov/static/Needed-State-Level-Integrated.pdf, accessed May 2023.

[27] J. Steinke, B. Bolunmez, L. Fletcher, V. Wang, A. J. Tomassetti, K. M. Repchick, S. J. Zaccaro, R. S. Dalal, and L. E. Tetrick, "Improving Cybersecurity Incident Response Team Effectiveness Using Teams-Based Research," George Mason University, Available from: https://ieeexplore.ieee.org/document/7180274, accessed May 2024.

[28] B. West, "Counterinsurgency lessons from Iraq," U.S.A. Army 5 May 2009, Available from: Counterinsurgency lessons from; https://www.army.mil/article/20621 /counterinsurgency_lessons_from_iraq , accessed 23 February 2024.

Comparison Between Existing Accident Models and Surrogate Safety Assessment Models (SSAM) on Unconventional Roundabouts, with Focused Applications of the Latter to some Real Study Cases

Antonio Pratelli
Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy
e-mail: antonio.pratelli@unipi.it

Lorenzo Brocchini
Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy
e-mail: lorenzo.brocchini@phd.unipi.it

Pietro Leandri
Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy
e-mail: pietro.leandri@unipi.it

Rosaria Aiello
Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy
e-mail: aiello.sara91@gmail.com

***Abstract*— This article has been divided into two main parts. The first and most substantial part describes the comparison between the expected number of accidents calculated through analytical models and the Surrogate Safety Assessment Model (SSAM) of the Federal Highway Administration (FHWA), regarding unconventional roundabouts. The novelty of this comparison lies precisely in the fact that the three roundabouts analysed fall into the category of so-called Unconventional Roundabouts, i.e., arrangements with "roundabout circulation", which do not fall within the types listed in the Italian Guidelines. In any case, apart from this latest innovation coupled with the small size of the sample observed, the present work can be considered an exploratory study with a view to further development. Returning to talk about roundabouts, it is possible to state that among the various types of accidents that may occur, those of the rear-end collision type occur more frequently, for which it was decided to use the formulas of the accident models relating to this type of conflict. In particular, the conflict type "Approach" for the Maycock & Hall model and the conflict type "Rear end" for the Arndt & Troutbeck model were taken into consideration. In addition to the application of analytical models, possible points of conflict (of the same category, i.e., "Rear end") were evaluated using dynamic simulation models. In particular, the dynamic simulation software Aimsun™ was used as a means to obtain the necessary inputs for the evaluation of the surrogate safety carried out through SSAM, a software application that reads the trajectory files generated by the simulation programs. The second part of the article instead focuses on the application of SSAM to two real case studies for which, thanks to the results obtained, it was possible to demonstrate the effectiveness of the proposed solutions. At the end of this work, both the conclusions on the comparison made and on the application of SSAM to real cases have been inserted.**

Keywords- Unconventional Roundabouts; Microsimulations with Aimsun; SSAM; Accidents Models; Real Case Studies.

I. INTRODUCTION

This article is an extended version of the conference paper "Comparison between Surrogate Safety Assessment Models (SSAM) and Accident Models on Unconventional Roundabouts" [1], presented at the Twelfth International Conference on Data Analytics in September 2023. In the first part, the comparison between the expected number of accidents calculated through analytical Accident Models and the Surrogate Safety Assessment Model (SSAM) of the Federal Highway Administration (FHWA), regarding unconventional roundabouts will be described. In the second part instead a focus on SSAM application, employed for two real case studies, will be illustrated.

Anyway, the whole research work starts from the idea of the authors to develop the work carried out by Vasconcelos et al. in the article "Validation of the Surrogate Safety Assessment Model for Assessment of Intersection Safety" [2]. In particular, the authors have decided to resume the research work carried out and extend it with their contribution, starting from their conclusion that the Surrogate Safety Assessment Model is a quite promising approach to assessing the safety of new facilities, innovative layouts and traffic regulation schemes. Then, the present work started from the fact that it is difficult to calculate the possible number of accidents in roundabouts with innovative layouts, because, unlike the conventional ones which are "geometrically identifiable", they have highly variable geometric parameters and therefore it is difficult to able to describe their road safety with a single model.

So, this research tried to describe the comparison between the Surrogate Safety Assessment Model (SSAM) of the Federal Highway Administration (FHWA) and the predicted number of accidents calculated through analytical models, regarding Unconventional Roundabouts. The extension of the work of Vasconcelos et al. [2] and therefore the novelties lie precisely in the fact that the three roundabouts analysed fall precisely into the category of so-called Unconventional Roundabouts, i.e., arrangements with "roundabout circulation", which do not fall within the types listed in the Italian Guidelines (Ministerial Decree 19-04-2006: "Functional and geometric rules for the construction of road intersections" [3]).

As said above, despite the innovations just highlighted, the significance of the study is mitigated by the limited size of the sample in question. Consequently, this work should be considered primarily as an exploratory effort, laying the foundation for future investigations aimed at refining and implementing a new possible study model. Thus, this preliminary study could serve as a springboard for deeper exploration and more comprehensive research for future advances in this area. In any case, although the present study represents just a preliminary exploration, its implications are already important, because they pave the way for future developments that have the potential to redefine the approach to computational modelling of Unconventional Roundabout safety.

Going back to talking about them, these roundabouts have shapes and dimensions that are out of the ordinary concept of roundabout intersection. As regards the accident models, it was decided to consider the formulas of the conflict type "Approach" for the Maycock & Hall [4] model and those of the conflict type "Rear end" for the Arndt & Troutbeck [5] model.

This choice is based on the fact that among the various types of accidents that can occur in roundabout intersections, rear-end collisions occur more frequently (literature the values vary from 20% to 25%). As far as the surrogate safety evaluation is concerned, it was carried out using SSAM (a software application that reads the trajectory files generated by the simulation programs) [6]. It was decided to use Aimsun™ as a dynamic microsimulation software, with which it was possible to obtain the ".trj files", i.e., the trajectory files, essential for calculating the possible points of conflict, which, by definition, are the points where two vehicles can potentially collide with each other at road intersections. Also, in this case, the points of conflict of the "Rear end" category have been taken into consideration.

Finally, to improve the visualization style of the points of conflict extrapolated from SSAM, it was decided to use the software Quantum Geographic Information System (QGIS); in this application, the files extrapolated from SSAM were inserted and geolocated.

The extended part of this work instead concerned a focus on SSAM application, employed for two real case studies. This idea was developed starting from one of the conclusions regarding the comparison between SSAM and Accident Models, i.e., that also for Unconventional Roundabouts there is a correspondence between the accident models and the calculation of the conflicts carried out with SSAM.

Thanks to this consideration, it was therefore possible to apply the SSAM model to two real cases. In detail, the first real case is nothing more than an Unconventional Roundabout (and, in particular, a Double Raindrop Roundabout) and the second, even more distinctive, is a Motorway Tollbooth.

The following sections will follow: a first more theoretical section which will deal with the Italian Unconventional Roundabouts with some examples that are taken into consideration; two sections concerning the existing roundabouts accident models and the SSAM approach from FHWA; a section, which will explain the comparison of the two approaches; the last section, followed then by the conclusions, will be focused on the application of SSAM approaches on the two real case studies.

II. Italian Unconventional Roundabouts

The subsections that follow will primarily deal with the theory of the so-called Unconventional Roundabouts, referring to the Italian Guidelines; and then move on to the three practical instances which were used to carry out the comparison between existing Accident Models and Surrogate Safety Assessment Models (SSAM).

A. Unconventional Roundabouts and Italian Guidelines

First of all, it is appropriate to specify what is meant by Unconventional Roundabouts [7] and why the authors decided to develop their research on them. In the Italian guidelines (Ministerial Decree 19-04-2006 [3]), there can be three basic types of roundabouts based on the diameter of the outer circumference: Conventional Roundabouts with an outer diameter between 40 and 50 m; Compact Roundabouts with outside diameters between 25 and 40 m; Mini Roundabouts with external diameter between 14 and 25 m. For arrangements with "roundabout circulation", which do not fall within the above typologies, we, therefore, speak of Unconventional Roundabouts and for them, the geometric dimensioning and verification must be adapted.

When we talk about Unconventional Roundabouts must be considered both the so-called "new generation roundabouts" (Raindrop or Double Raindrop Roundabouts; Turbo Roundabouts [8] [9]; Two-Geometry Roundabouts [10] [11]), which are currently being built to fulfil safety and performance objectives in cases where classic roundabouts are unable to work well; both the so-called "old roundabouts" which had dimensions and geometries suitable for when precedence was on the branches instead of on the ring (first generation roundabouts) [12]. In Italy, there are many Unconventional Roundabouts of both "typologies", both because in terms of space, there is the need to adopt solutions that are not conventional, and because for the moment there are always obsolete roundabouts on the national territory which have not been adapted and which are often poor in terms of security. Precisely for this last consideration, in this discussion, the authors have decided to take into consideration three Unconventional Roundabouts of the latter type and have decided to analyse them in terms of safety, also because from this point of view there are no in-depth studies for them.

A final introductory consideration concerns the type of accidents that the authors decided to analyse, i.e., rear-end collisions. They are the conflicts/accidents that occur on the entrance branches more frequently at "roundabout" intersections and for this reason, they were chosen as a study parameter.

B. Territorial Framework and O/D Matrices of the Three Identified Roundabouts

This short paragraph lists the three Unconventional Roundabouts analyzed by the authors.

All three roundabouts are situated in Italy, in the Tuscany region and are located in urban areas, therefore the speed referred to during the calculations is equal to 50 km/h [13]. In particular, in Fig. 1, Fig. 2 and Fig. 3, the three aerial images extracted from Google Earth are reported, where the progressive numbers of the branches of the roundabouts are also reported.

Reference is made to them for the reconstruction of the Origin/Destination (O/D) matrices, reported in turn in Table I, Table II and Table III.

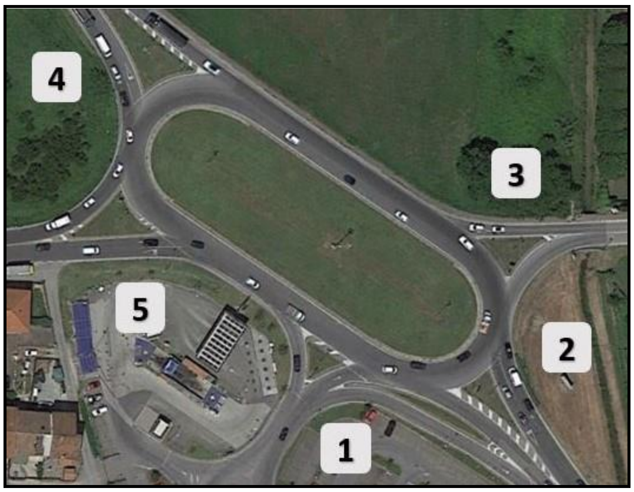

Figure 1. The territorial framework of the 1st Unconventional Roundabout located on SP61-Lucchese-Romana in Lucca, Tuscany, Italy (*source*: Google Earth Pro)

Figure 2. The territorial framework of the 2nd Unconventional Roundabout located on Viale Nazario Sauro in Livorno, Tuscany, Italy (*source*: Google Earth Pro)

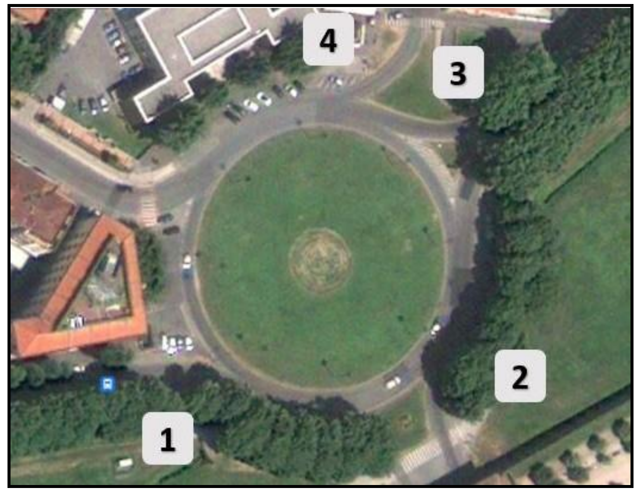

Figure 3. The territorial framework of the 3rd Unconventional Roundabout located on Porta Santa Maria in Lucca, Tuscany, Italy (*source*: Google Earth Pro)

TABLE I. O/D MATRIX OF THE 1ST UNCONVENTIONAL ROUNDABOUT

Roundabout 1 - SP61 Lucchese-Romana (Lucca, Tuscany, Italy)						
Matrice O/D	1	2	3	4	5	TOT
1	0	142	60	36	72	310
2	36	0	140	346	812	1334
3	44	204	0	114	76	438
4	58	320	56	0	280	714
5	58	794	184	372	0	1408
TOT	196	1460	440	868	1240	4204

TABLE II. O/D MATRIX OF THE 2ND UNCONVENTIONAL ROUNDABOUT

Roundabout 2 - Viale Nazario Sauro (Livorno, Tuscany, Italy)				
Matrice O/D	1	2	3	TOT
1	0	390	517	907
2	443	0	691	1134
3	476	541	0	1017
TOT	919	931	1208	3058

TABLE III. O/D MATRIX OF THE 3RD UNCONVENTIONAL ROUNDABOUT

Roundabout 3 - Porta Santa Maria (Lucca, Tuscany, Italy)					
Matrice O/D	1	2	3	4	TOT
1	181	299	1749	0	2229
2	253	0	195	0	448
3	951	52	12	0	1015
4	263	51	12	0	326
TOT	1648	402	1968	0	4018

These matrices were elaborated starting from the data surveys carried out on the three roundabouts through the use of Sony DCR-SX34 digital cameras, positioned at specific points of the intersections, during the peak periods of the week [14].

III. EXISTING ROUNDABOUTS ACCIDENT MODELS

Roundabouts, in general, are considered to be the safest road junctions as they have several advantages including reduction of points of conflict and lower movement and departure speeds. However, accidents can also occur on them and in particular, several studies state that the most common accident that can occur is a rear-end collision. To study the safety characteristics of the elements of the road system, there are several models for predicting accidents [15].

The authors have decided to use in this research two of the most used models, namely those of the Maycock & Hall model and the Arndt & Troutbeck model. They were chosen because they allow the number of accidents to be calculated taking into consideration both the traffic demand, the geometric characteristics of the intersection, and the dynamic ones (such as speed, for example). With these models, it is possible to calculate various types of accidents, but, as explained above, it was decided to use the formulas of the Conflicts Type "Approach" for the Maycock & Hall [4] model (1) and those of the Conflict Types "Rear end" for the Arndt & Troutbeck [5] model (2), which indicate precisely rear-end collisions. Both models make it possible to estimate the number of accidents over a period of time and therefore their unit of measurement is expressed in accidents/years [16].

The two formulas (1) and (2) used are therefore reported below, specifying that the coefficients of these formulas are the standard ones calibrated for conventional roundabouts. In fact, another of the interesting aspects of this research was precisely that of verifying whether these coefficients could also work for Unconventional Roundabouts. To answer this question, see section V.

$$A_2 = 0.0057 \times Q_e^{1.7} \times \exp(20C_e - 0.1e) \quad (1)$$

where:

- Q_e = entering flow, respectively (1000s of vehicles/day);
- C_e = entry curvature [Ce = 1/Re and Re = entry path radius for the shortest vehicle path (m)];
- e = entry width [m].

$$A_r = C_1 \times Q_a^x \times Q_c^y \times S_a^z + C_2 \quad (2)$$

where:

- Q_a = average annual daily traffic (AADT) on the approach;
- Q_c = various AADT flows on the circulating carriageway adjacent to the approach;
- S_a = 85th percentile speed on the approach curve (the potential relative speed between approaching vehicles) [km/h];
- $C_1 = 9.62 \times 10^{-11}$; $C_2 = 0$; x = 1; y = 0.5; z = 2. [5]

IV. SSAM Approach from FHWA

This concise section has been included to define what is meant by surrogate security assessment and how it is possible to carry out such an assessment.

After careful bibliographic research on the surrogate safety measures in safety evaluation and analysis [17], it is possible to affirm that, in any case, whatever safety analysis is a decisive aspect in the evaluation of design choices both for the new road system and for the adaptation of the existing road network.

In fact, several studies deal with safety assessment when any intersection is converted into a roundabout [18] and in addition to this, in the literature, there are also various insights regarding models that connect the parameter "safety of roundabouts" to the predicted speed in them (another fundamental parameter, for example in terms of efficiency estimation) [19].

So, to fulfil this, the Federal Highway Administration (FHWA) has developed and made available the Surrogate Safety Assessment Model (SSAM) program, through which it aims to offer designers, researchers and companies specializing in road design and construction a tool for assessing the safety of an intersection by estimating the frequency of conflicts [20] [21].

The concept of surrogate safety derives from the desire to develop alternative tools to the existing ones to evaluate the accident frequency of road infrastructure (among which mentioned the Empirical Bayesian analysis [22] or the Crash Modification Factors [23]).

While the so-called ordinary methods derive from statistical evaluations based on accidents that have occurred, the surrogate safety methods are instead based on factors that do not require years of accident statistics.

The SSAM program elaborates the trajectory files (.trj files) obtained in output from a dynamic simulation program (in the case of the present research it is decided to use the Aimsun™ program, but in general VISSIM™, TEXAS™, etc..). In detail, SSAM evaluates every single vehicle-vehicle interaction according to criteria with which it can establish whether there is a point of conflict and to which category it belongs. At the end of the computations, SSAM presents the results in tables, allowing the user to filter them according to the parameters of his choice.

As regards the classification of conflicts, the program contemplates four types: Rear end, Lane changing, Crossing and Unclassified.

To classify them, the program evaluates the crossing angle of the trajectories (Fig. 4), if this angle is less than 20° the conflict is of the Rear end type.

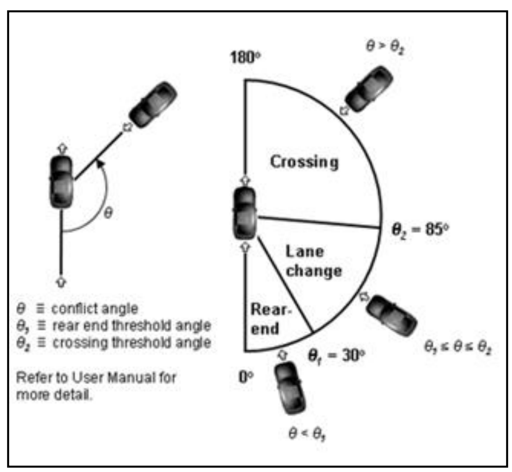

Figure 4. Conflict angle threshold (SSAM)

In the present research, the latter (Rear end) have been taken into consideration, since, as already explained, they are the ones that occur most frequently in roundabout intersections. Their unit of measurement is expressed in conflicts/day.

V. COMPARISON OF THE TWO APPROACHES

The following section presents the results of the first and most substantial part of the research, or rather, the comparison between the expected number of accidents calculated through analytical models and the Surrogate Safety Assessment Model (SSAM) of the Federal Highway Administration (FHWA) regarding the three unconventional roundabouts. First of all, a summary table (Table IV) of the calculations carried out is shown which serves to reconstruct the graphs on which most of the considerations will be made.

TABLE IV. SUMMARY TABLE OF THE CALCULATIONS MADE

Roundabout	Approach	Qe [veh/d]	Arndt & Troutbeck Rear-end [acc/y]	Maycock & Hall Approach [acc/y]	SSAM (TTC = 1.5 s) [conflicts/d]
1	1	3100	0.10	0.07	24
	2	13340	0.28	0.33	383
	3	4380	0.14	0.13	63
	4	7140	0.19	0.23	165
	5	14080	0.29	0.34	207
2	1	9070	0.16	0.15	120
	2	11340	0.20	0.37	203
	3	10170	0.16	0.27	119
3	1	22290	0.18	0.55	160
	2	4480	0.15	0.13	82
	3	10150	0.16	0.32	104
	4	3260	0.09	0.07	36

Furthermore, the authors considered it necessary to also report the explanatory images of the surrogate safety assessment. In detail, the following images (Fig. 5, Fig. 6 and Fig. 7) show an extract of the QGIS software of the three roundabouts, where the points of conflict have been inserted, georeferenced (with TTC = 1.5 s) extracted from the SSAM software after processing the ".trj file", which in turn was obtained from the Aimsun™ simulation software. The Time to Collision (TTC) is one of the SSAM software parameters and expresses the minimum collision time [24]. It can range from an infinite maximum value, when two vehicles never meet, to a minimum value of 0 seconds when an accident occurs. Various studies have been conducted to identify a threshold value of the TTC, such as to separate major accidents from minor and negligible or without consequences accidents [25]. This value, depending on the study, was identified as a fixed value or as the result of a function dependent on the speed or deceleration of the vehicles. The authors have decided to keep the default value of the SSAM program which assumes the value TTC = 1.5 s.

Figure 5. Number of Conflicts obtained by SSAM software and reported on QGIS of 1st Unconventional Roundabout

Figure 6. Number of Conflicts obtained by SSAM software and reported on QGIS of 2nd Unconventional Roundabout

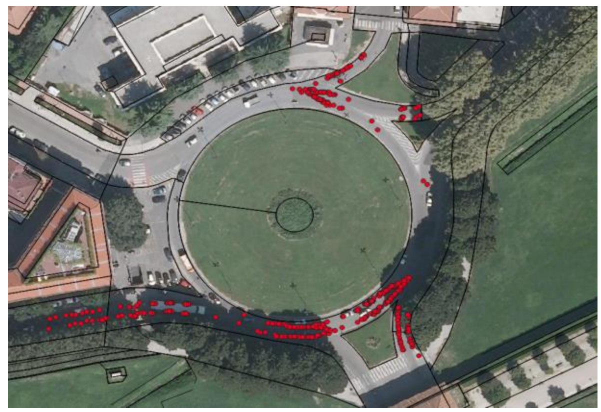

Figure 7. Number of Conflicts obtained by SSAM software and reported on QGIS of 3rd Unconventional Roundabout

Below are the graphs (Fig. 8, Fig. 9 and Fig. 10) which summarize most of the research results. In particular, each graph refers to one of the three roundabouts and is structured as follows: the Q_e (entrance vehicular flow) expressed in vehicles per day is shown on the abscissa axis; while there are two different y axes. The left y-axis is incident models (Arndt & Troutbeck / Maycock & Hall) and is expressed in accidents per year, while the right is the SSAM results and is expressed in conflicts per day.

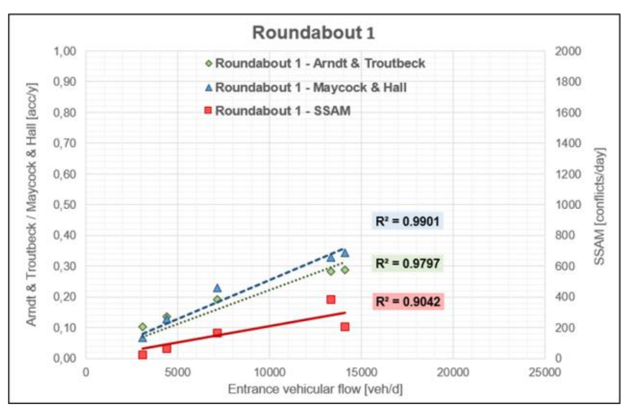

Figure 8. Graph of Results for the 1st Unconventional Roundabout

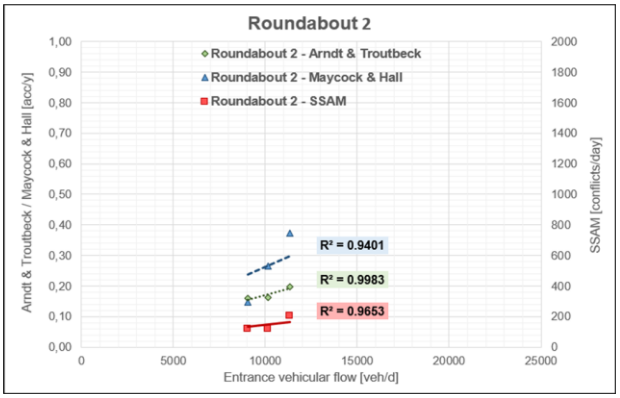

Figure 9. Graph of Results for the 2nd Unconventional Roundabout

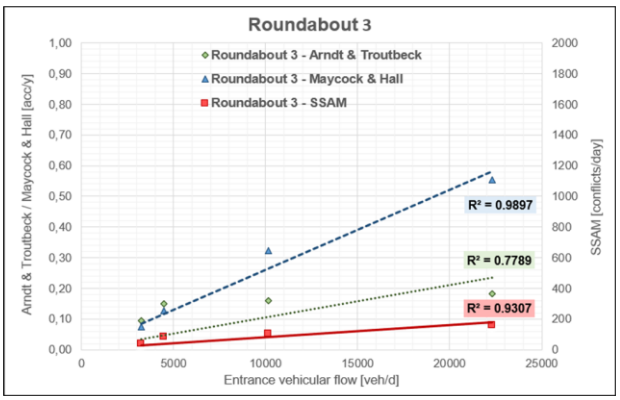

Figure 10. Graph of Results for the 3rd Unconventional Roundabout

On the graphs, as many points have been reported as there are entrance arms of the roundabout in question and a linear trend line passing through the origin (0; 0) has then been created for them.

After that, the authors decided to calculate the coefficient of determination R^2 for each trend line. It is a statistical value that allows us to understand whether a linear regression model can be used to make predictions. Its value is always between 0 and 1, or between 0% and 100% if you want to express it in percentage terms. $R^2 = 0$ indicates a model whose predictor variables do not explain the variability of y around its mean at all. $R^2 = 1$ indicates a model whose independent variables fully explain the variability of y around its mean; that is, knowing the values of the independent variables one can predict exactly what the value of y will be. Clearly, the values 0 and 1 are limit values, what emerges is that the greater the value of R^2, the more the model has high predictive power, i.e., the better the ability of the explanatory variables to predict the values of the dependent variable. Usually, we talk about high R^2 values, when they are higher than 0.7.

At this point, after having explained the type of graphs used and the reference values, it is possible to go into detail on the considerations relating to the actual results. For all the graphs, i.e., for all the roundabouts, the R^2 values are generally excellent (they are always higher than 0.9, except for one case), both as regards the accident models and as regards the values of the conflicts obtained with SSAM.

This is an excellent result as the three roundabouts to which the models have been applied are Unconventional Roundabouts, i.e., "different" intersections from the ones on which the models have been calibrated. Therefore, as a first result, it is certainly possible to state that the accident models used (Arndt & Troutbeck / Maycock & Hall), which are already valid and validated for conventional roundabouts, can also be used for Unconventional Roundabouts, using the same formulations and the same coefficients.

Also, about the SSAM results, the R^2 values are always higher than 0.9 and despite the different scales it is possible to state that the trend of the trend lines of the points deriving from SSAM is very similar to that relating to the accident models.

This is therefore another excellent result that the authors have arrived at, namely that even for Unconventional Roundabouts there is a correspondence between the accident models and the calculation of the conflicts carried out with SSAM.

Finally, the authors also noted a further fact regarding Fig. 10, i.e., the graph referring to roundabout number 3. The trend line of the Arndt & Troutbeck model has an R^2 that is always acceptable, but clearly lower than all the others (0.7789).

The explanation that the authors came up with is the following: roundabout number 3, in addition to being of an unconventional type, is also atypical from the point of view of the approaches, since, as can be seen from the territorial framework (Fig. 3) and the corresponding O/D matrix (Table III), the approach 4 is formed only by the input branch and not the output branch.

This, together with the particular geometry of the roundabout, has led to a high difference between the incoming flow rate Q_e and the circulating flow rate Q_c of the adjacent approach 1 (this difference is underlined in Table V). So, another result that the authors have reached is the consideration that the model of Arndt & Troutbeck does not adapt perfectly to Unconventional Roundabouts in which there is, for some branches, a high difference between the incoming flows and circulating flows.

TABLE V. EXTRACT FROM THE CALCULATION TABLE, WHERE THE DIFFERENCE BETWEEN QE AND QC CAN BE SEEN

Roundabout	Approach	Qe [veh/d]	Qc [veh/d]	Delta (Qe-Qc/Qc)
3	1	22290	1150	18.38
	2	4480	19540	0.77
	3	10150	4340	1.34
	4	3260	14490	0.78

A final comparison was also made for the three Unconventional Roundabouts as a whole.

In fact, a last graph (Fig. 11), of the same typology as the previous ones, was constructed however by taking into consideration the roundabouts as a whole and no longer approach by approach. In this way, it was possible to compare the three roundabouts on a single graph and this led to the following consideration.

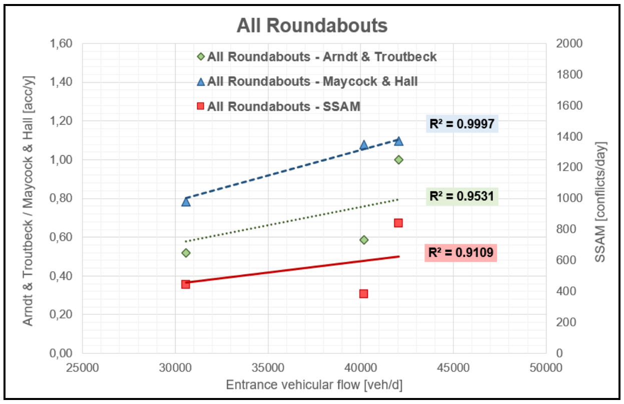

Figure 11. Graph of Results for the three Unconventional Roundabouts together

The values of R^2 are excellent and also the roundabout 3 which had a deficit on the Arndt & Troutbeck model due to the difference between the incoming flows and the circulating flows at one of the approaches, if it is considered as a whole, it is possible to homogenize with the other results.

VI. Focused Application of SSAM Approach to Real Case Studies

This last paragraph before the conclusions of the entire work, deals with the second part of the research and focuses on the SSAM application used for two real case studies for which, thanks to the results obtained, it was possible to demonstrate the effectiveness of the proposed solutions. This idea was developed starting from one of the conclusions set out in the previous paragraph, where the comparison was made between SSAM and Accident Models, i.e., that even for Unconventional Roundabouts there is a correspondence between the two methods. Thanks to this consideration, it was therefore, possible to use the SSAM model for two real cases, for which the accident models were not very suitable or difficult to apply. It should also be said that all this is further strengthened by bibliographic research, in which it was found that there already exist several studies on the application of SSAM to real cases [26] [27]. Starting from the considerations just made, the two real cases that have been analysed are the following: an Unconventional Roundabout (and in particular a Double Raindrop Roundabout); a Motorway Tollbooth (and in particular the intermediate section which is located between two successive motorway toll booths where various weaving manoeuvres take place). For both real case studies, two solutions will be illustrated. The so-called "Initial Solution" will concern either the current state of the intersection (in the case of the motorway toll booth) or a possible solution proposed for the intersection itself, which however was not found to be safe and efficient enough (in the case of the unconventional roundabout); and the so-called "Project Solution" which instead will concern a proposed design hypothesis for the intersection. For both solutions, a dynamic simulation was carried out with the aforementioned Aimsun™ software which made it possible to obtain the necessary inputs (trajectory files) for the surrogate safety assessment carried out via SSAM. There will now follow two paragraphs relating to the two real cases in which they have been inserted: the territorial framework of the intersection, the traffic status of the intersection, the number of conflicts obtained with SSAM and reported on QGIS for both the "Initial Solution" and the "Project Solution", a final comparison between the two solutions.

A. Real Case Study 1: Unconventional Roundabout in Lucca, Tuscany, Italy

The first real case study that will be covered is an Unconventional Roundabout located in Lucca, Tuscany, Italy. The roundabout falls into the category of Unconventional Roundabouts as both the shape of its "Initial Solution" and that of its "Project Solution" are certainly outside conventional geometric standards. If we wanted to classify it, we could include it among the Raindrop or Double Raindrop Roundabouts. This type of work is also supported by the fact that studies have already been done regarding the application of SSAM to real cases of Unconventional Roundabouts, such as those relating to the Separated Central Island (SCI) Roundabouts [28]. Having made this brief introduction, it is now possible to move on to illustrate the real case study 1 starting from the images of the territorial framework of the "Initial Solution" and the "Project Solution" (Fig. 12 and Fig. 13) and continuing with the O/D matrices always relating to the two solutions (Table VI and Table VII).

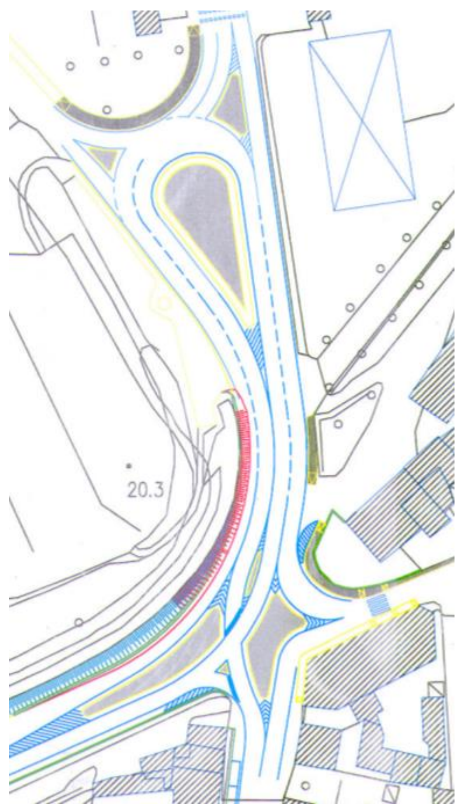

Figure 12. Territorial framework of the Unconventional Roundabout located in Lucca, Tuscany, Italy ("Initial Solution")

The "Initial Solution" (Fig. 12) is composed of a very complex at-grade intersection in the southern part and a raindrop roundabout in the northern part.

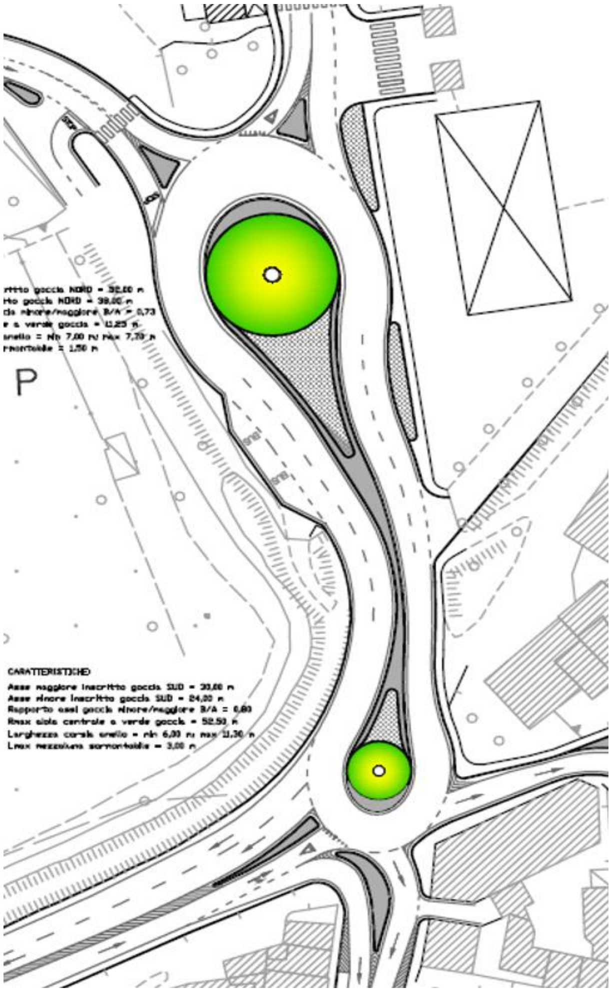

Figure 13. Territorial framework of the Unconventional Roundabout located in Lucca, Tuscany, Italy ("Project Solution")

The "Project Solution" (Fig. 13) is composed of a double raindrop roundabout. The two O/D matrices (Table VI and Table VII) refer to the arms of the intersections and are numbered from 1 to 4 starting from the North and continuing counter clockwise (therefore: N = 1, W = 2, S = 3 and E = 4).

TABLE VI. O/D MATRIX OF THE UNCONVENTIONAL ROUNDABOUT ("INITIAL SOLUTION")

O/D Matrix	1	2	3	4	Qe
1	0	439	1160	0	1599
2	482	0	66	0	548
3	0	0	0	0	0
4	502	488	25	0	1015
Qu	984	927	1257	0	3162

TABLE VII. O/D MATRIX OF THE 3RD UNCONVENTIONAL ROUNDABOUT ("PROJECT SOLUTION")

O/D Matrix	1	2	3	4	Qe
1	0	327	864	1164	2355
2	605	0	17	185	807
3	0	0	0	0	0
4	0	0	0	0	0
Qu	605	327	881	1349	3162

These matrices were elaborated starting from the data surveys carried out on the intersection, during the peak periods of the week.

At this point, as was done for paragraph V, the authors deemed it necessary to report the explanatory images of the surrogate safety evaluation of the Unconventional Roundabout of Lucca, both in its "Initial Solution" and "Project Solution" configuration. In detail, the following images (Fig. 14 and Fig. 15) show an extract from the QGIS software of the two solutions, where the conflict points were inserted, georeferenced (with TTC = 1.5 s) extracted from the SSAM software after having processed the ".trj file", which in turn was obtained from the Aimsun™ simulation software.

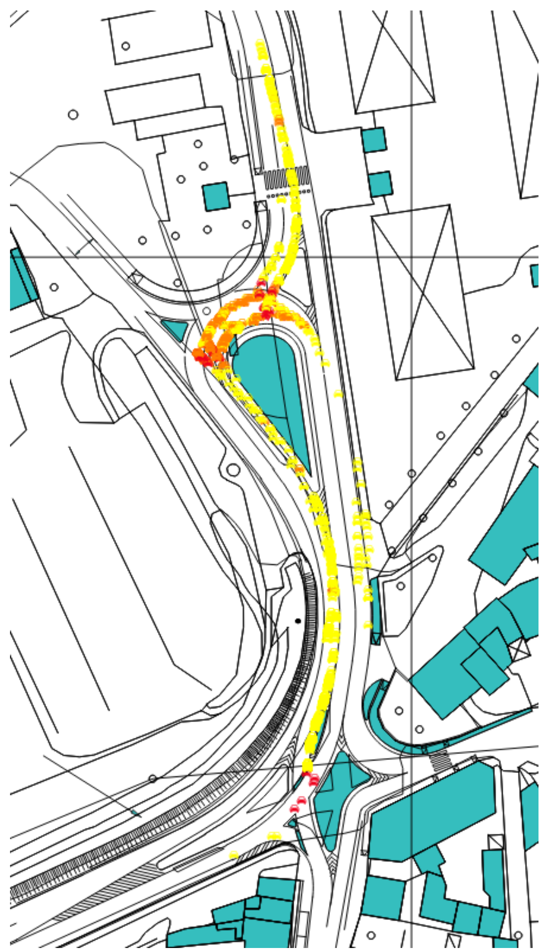

Figure 14. Number of Conflicts obtained by SSAM software and reported on QGIS of the Unconventional Roundabout ("Initial Solution")

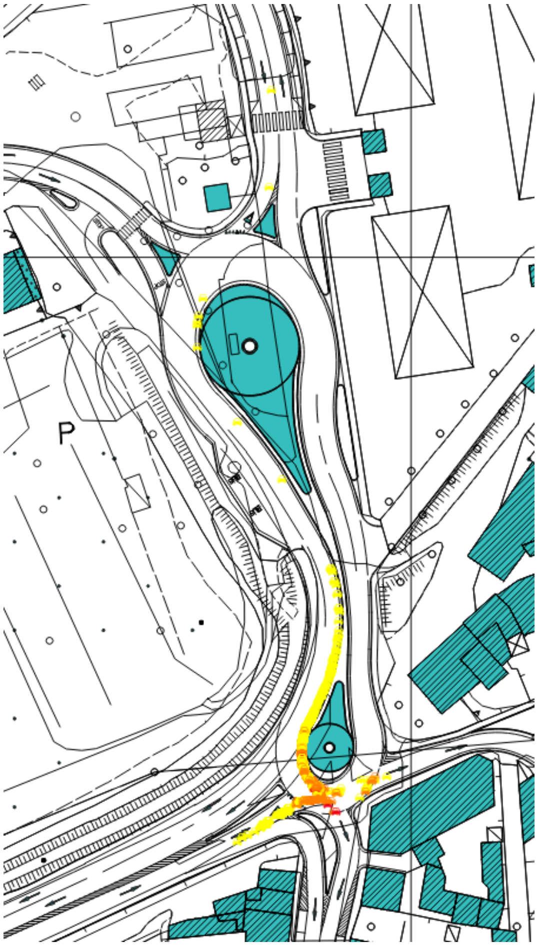

Figure 15. Number of Conflicts obtained by SSAM software and reported on QGIS of the Unconventional Roundabout ("Project Solution")

It is specified that in this case, you can also notice the difference between the various types of conflicts (listed and explained in Fig. 4) and in detail: in red there are Crossing type conflicts, in orange there are Lane change type conflicts and finally those of the Rear end type in yellow.

Before concluding the paragraph, it is also appropriate to insert the summary table (Table VIII) of the number of conflicts calculated with SSAM of the two solutions which served to compare them and demonstrate that, thanks to the results obtained, the "Project Solution" was better from the point of view of safety.

TABLE VIII. SUMMARY TABLE OF THE COMPARISON

SSAM Calculation					
TTC = 1.5 s					
"Initial Solution"		"Project Solution"		Comparison %	
Unclassified	0	Unclassified	0	Unclassified	-
Crossing	69	Crossing	46	Crossing	-33%
Lane Change	383	Lane Change	805	Lane Change	110%
Rear End	1015	Rear End	996	Rear End	-2%
Total	1467	Total	1847	Total	26%
TTC = 0 s					
"Initial Solution"		"Project Solution"		Comparison %	
Unclassified	0	Unclassified	0	Unclassified	-
Crossing	34	Crossing	0	Crossing	-100%
Lane Change	202	Lane Change	22	Lane Change	-89%
Rear End	32	Rear End	30	Rear End	-6%
Total	268	Total	52	Total	-81%

From this table, it is possible to immediately notice that regarding the TTC = 0 s (which is the theoretical minimum value to be assigned to evaluate the conflicts that "certainly" will occur) there is a very high decrease for each type of conflict, up to a total decrease of -81%. However, regarding the TTC = 1.5 s (default value), one fundamental thing can be noted, namely that even if there is an increase in the "lane change" type conflict points, there is a clear decrease in the crossing type conflicts which are the most dangerous (with an angle of incidence between 85° and 180°, see Fig. 4).

Therefore, thanks to this type of analysis, it is possible to state that from the point of view of security analysis, the "Project Solution" is better.

B. Real Case Study 2: Motorway Tollbooth in Lucca, Tuscany, Italy

The second real case study that will be illustrated is a highway toll booth located in Lucca, Tuscany, Italy. First of all, is important to affirm that also this type of work is supported by the fact that previous research has already been carried out regarding the micro-simulation of real cases of motorway toll booths [29], even if without the specific application of SSAM.

Therefore, this in-depth study constitutes a further innovative aspect of the research; that said, it is now possible to move on to illustrate the real case study. To be more precise, the simulations that were carried out concerned the two motorway toll booths in the city of Lucca (West and East) which are located at a very close distance, as is illustrated in the territorial framework (Fig. 16).

Figure 16. The territorial framework of the Motorway Tollbooth in Lucca, Tuscany, Italy
(*source*: Google Earth Pro)

In fact, the study began from the consideration that due to the close distance between the entrance from the Lucca West motorway toll booth and the exit from the Lucca East motorway toll booth, several weaving manoeuvres occur (dangerous manoeuvres especially at high speed). Consequently, the project was aimed at moving the Lucca East toll booth and transforming it into an exchange car park.

Without going into the details of the work, what interests the authors is the fact that one of the main factors that made it possible to demonstrate that this design idea was correct, were precisely the dynamic simulations and the SSAM application in the two situations (the "Initial Situation" with the toll booths at the current state and the "Project Situation" with the moving of the Lucca East toll booth and its transformation into an exchange parking lot). The results of the application of SSAM are therefore reported below.

As already mentioned in Fig. 16 the territorial framework has been inserted, while in the subsequent figures (Fig. 17 and Fig. 18) there are explanatory images of the surrogate safety assessment in the intermediate road sections between the two toll booths, where takes place the weaving manoeuvres. In particular, the following two images show the extracts from the QGIS software of the two situations, where the conflict points were inserted, georeferenced (with TTC = 1.5 s) extracted from the SSAM software after having processed the ". trj file", which in turn was obtained from the Aimsun simulation software.

The figures, as already mentioned, refer to the "Initial Situation" corresponding to the reality in which there are two motorway toll booths very close to each other, and to the "Project Situation" in which the Lucca East toll booth has been transformed into an interchange car park.

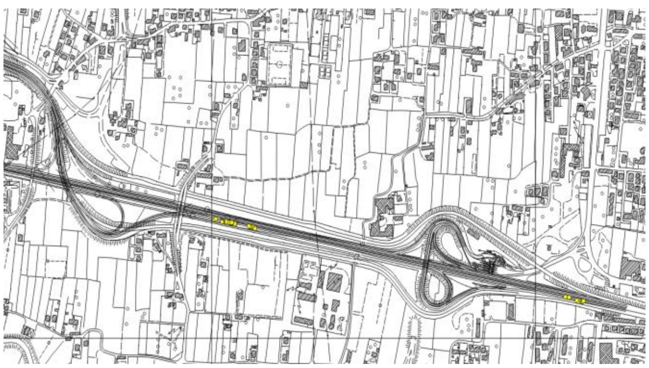

Figure 17. Number of Conflicts obtained by SSAM software and reported on QGIS of the Motorway Tollbooth ("Initial Situation")

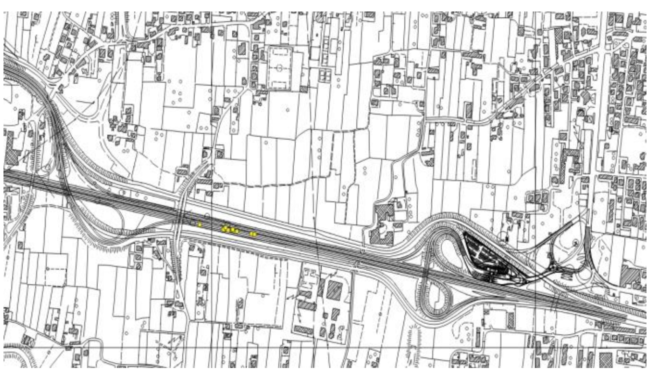

Figure 18. Number of Conflicts obtained by SSAM software and reported on QGIS of the Motorway Tollbooth ("Project Situation")

Before concluding the paragraph, as was done for the previous paragraph, it is also appropriate to insert the summary table (Table IX) of the number of conflicts calculated with SSAM of the two situations which served to compare them and demonstrate that, thanks to the results obtained, the "Project Situation" was better from the point of view of safety.

TABLE IX. SUMMARY TABLE OF THE COMPARISON

SSAM Calculation					
TTC = 1.5 s					
"Initial Situation"		"Project Situation"		Comparison %	
Unclassified	0	Unclassified	0	Unclassified	0%
Crossing	0	Crossing	0	Crossing	0%
Lane Change	0	Lane Change	0	Lane Change	0%
Rear End	30	Rear End	15	Rear End	-50%
Total	30	Total	15	Total	-50%

From this table it is possible to notice that at the Lucca West motorway toll booth, there is a reduction in the number of possible conflicts due to the movement of vehicular flows, directed towards the eastern part of the city of Lucca, towards the new toll booth moved further forward; which also involves a reduction in weaving manoeuvres along the stretch. In addition, the Lucca East motorway toll booth, transformed into a motorway interchange car park, has observed a cancellation of the possible conflicts in entry and exit to and from it, due to the change of destination of the same and consequently to the lower vehicle flows circulating there. Concluding, in terms of the number of possible conflicts that can occur on the stretch in question, an overall reduction of 50% is observed. Therefore, thanks to this type of analysis, it is possible to state that from the point of view of security analysis, the "Project Situation" is better.

VII. CONCLUSIONS AND FUTURE RESEARCH WORK

First of all, it is worth remembering that this article is an extended version of the conference paper "Comparison between Surrogate Safety Assessment Models (SSAM) and Accident Models on Unconventional Roundabouts" [1], presented at the Twelfth International Conference on Data Analytics in September 2023.

In the first and most in-depth part, this article describes the comparison between the Federal Highway Administration (FHWA) Surrogate Safety Assessment Model (SSAM) and the predicted number of accidents calculated using the Arndt & Troutbeck and Maycock & Hall analytical models, as concern the Unconventional Roundabouts [30] [31]. In the second part instead, in which most of the new contents were inserted, a focus on the SSAM application, employed for two real case studies, has been illustrated.

The conclusions relating to the first part will now be illustrated first. Three Unconventional Roundabouts located on the Italian territory that have different shapes and sizes from the regulatory standards were analysed. Other works and articles have been published regarding the comparison between the models mentioned, however, the novelty of this research proposed by the authors lies precisely in the different base data, i.e., the Unconventional Roundabouts. The type of accident and conflict chosen for the comparison made is that of rear-end collisions, as it is the most common present on roundabout intersections. In the sections of the article, various initial considerations follow one another which deepen the concepts of Unconventional Roundabouts, surrogate safety analysis models (SSAM) and accident models; up to section V where the results of the entire research were clearly explained.

Summarizing these results, the authors found that: 1) the accident models used (Arndt & Troutbeck / Maycock & Hall) already valid and validated for conventional roundabouts, can also be used for Unconventional Roundabouts, using the same formulations and the same coefficients also because a certain correspondence was also found between them in terms of the number of accidents per year; 2) also for Unconventional Roundabouts there is a correspondence between the accident models and the calculation of the conflicts carried out with SSAM; 3) Arndt & Troutbeck model is not perfectly suited to Unconventional Roundabouts in which there is, for one or more branches, too high a difference between incoming flows and circulating flows.

As previously said, the extended part of this work instead concerned a focus on SSAM application, employed for studying two real case studies. This idea was developed starting from the conclusion 2) illustrated above, i.e., that also for Unconventional Roundabouts there is a correspondence between the accident models and the calculation of the conflicts carried out with SSAM. Thanks to this consideration, it was therefore possible to apply the SSAM model to two real cases for which a safety analysis was required. In detail, the first real case was an Unconventional Roundabout (classified as Raindrop or Double Raindrop Roundabout) located in Lucca, Tuscany, Italy; was associated with the Unconventional Roundabouts because its shape in both the "Initial Solution" and the "Design Solution" was certainly outside conventional geometric standards. Instead, the second real case was a highway tollbooth located in Lucca, Tuscany, Italy; to be more specific, the research work that was carried out concerned the two motorway toll booths of Lucca West and Lucca East which are located at very close distance.

The conclusions that can be drawn from the study of these two real cases are the same, that is, thanks to this type of analysis it is possible to state that from the point of view of security analysis, one of the two solutions (and in these two specific cases, the project ones) is better than the other.

Speaking even more generally and therefore taking up the entire article, it is possible to state that the Surrogate Safety Assessment Model (SSAM) is a very powerful tool that can be used on various occasions both for research and practical purposes. The Accident Models however remain usable on all those occasions in which the starting conditions exist to be able to apply them (such as for conventional roundabouts), but in all those cases in which these models have not been validated and/or calibrated, SSAM remains one of the best solutions for evaluating intersection safety.

As previously mentioned in the introduction, it is important to clarify that this paper presents an exploratory study and, as such, the authors do not aim to propose any specific model. The primary objective of this paper is to describe empirical evidence derived from a small sample of Unconventional Roundabouts; this aspect is emphasized throughout the entirety of the paper.

Nevertheless, despite being only a preliminary exploration, the implications of the present study are significant as they lay the groundwork for future developments that could potentially redefine the approach to computational modelling of Unconventional Roundabout safety.

About that, before concluding the work, the authors decided to also propose some ideas for the possible future development of this research. First of all, this work can certainly be expanded by analysing further case studies and thus obtaining more points to use on the graphs obtained. Furthermore, the accident models utilised were used in the first analysis without the recalibration for the Unconventional Roundabouts; therefore another next step could be proper to go and search for the actual accident data and thus verify whether the parameters used can be further improved and better recalibrated for Unconventional Roundabouts (it is emphasized that however, as explained in section V, the accident models used, can already be used also for Unconventional Roundabouts, given the statistical results obtained by the authors).

References

[1] A. Pratelli, L. Brocchini, P. Leandri, and R. Aiello, “Comparison between Surrogate Safety Assessment Models (SSAM) and Accident Models on Unconventional Roundabouts”, Proceedings of the Twelfth International Conference on Data Analytics, IARIA, ThinkMind, Porto, pp. 7-12, 2023.

[2] L. Vasconcelos, L. Neto, A. M. Seco, and A. B. Silva, “Validation of the Surrogate Safety Assessment Model for Assessment of Intersection Safety”, Transportation Research Board, No. 2432, pp. 1-9, Washington, D.C., 2014.

[3] Italian Ministry of Infrastructures & Transport, “Norme funzionali e geometriche per la costruzione delle intersezioni stradali”, DM n. 1699 of 19/04/2006, Rome, 2006.

[4] G. Maycock and R. D. Hall, “Accidents at four-arm roundabouts”, PTRC Summer Annual Conference, Brighton, United Kingdom, 1984.

[5] O. K. Arndt and R. J. Troutbeck, “Relationship Between Roundabout Geometry and Accident Rates”, Transportation research circular, 1998.

[6] M. S. Ghanim and K. Shaaban, “A Case Study for Surrogate Safety Assessment Model in Predicting Real-Life Conflicts”, Arabian Journal for Science and Engineering, Vol. 44 pp. 4225-4231, 2019.

[7] T. Tollazzi, “Alternative Types of Roundabouts. An informational guide”. Springer, Berlin, Germany, 2015.

[8] A. B. Silva, L. Vasconcelos, and S. Santos, “Moving from Conventional Roundabouts to Turbo-Roundabouts”, EWGT2013 – 16th Meeting of the EURO Working Group on Transportation. Procedia - Social and Behavioral Sciences, Vol. 111, pp. 137-146, 2014.

[9] G. Tesoriere, T. Campisi, A. Canale, and T. Zgrablić, “The Surrogate Safety Appraisal of the Unconventional Elliptical and Turbo Roundabouts”, Journal of Advanced Transportation, pp. 1-9, 2018.

[10] A. Pratelli, R. R. Souleyrette, and L. Brocchini, “Two-Geometry Roundabouts: Design Principles”, Transportation Research Procedia, Vol. 64, pp. 299-307, 2022.

[11] A. Pratelli and L. Brocchini, “Two-Geometry Roundabouts: Estimation of Capacity”, Transportation Research Procedia, Vol. 64, pp. 232-239, 2022.

[12] A. R. Alozi and M. Hussein, “Multi-criteria comparative assessment of unconventional roundabout designs”, International Journal of Transportation Science and Technology, Vol. 11, pp. 158-173, 2022.

[13] D. Ciampa, M. Diomedi, F. Giglio, S. Olita, U. Petruccelli, and C. Restaino, "Effectiveness of unconventional roundabouts in the design of suburban intersections", European Trasnport, Vol. 80, 2020.

[14] A. Pratelli, P. Sechi, and R. R. Souleyrette, "Upgrading Traffic Circles to Modern Roundabouts to improve Safety and Efficiency – Case Studies from Italy", Promet – Traffic&Transportation, Vol. 30, No. 2, pp. 217-229, 2018.

[15] A. P. Tarko, "Use of crash surrogates and exceedance statistics to estimate road safety", Accident; Analysis and P revention, Vol. 45, pp. 230-240, 2012.

[16] S. Daniels, T. Brijs, E. Nuyts, and G. Wets, "Extended prediction models for crashes at roundabouts", Safety Science - SAF SCI, Vol. 49, pp. 198-207, 2011.

[17] D. Singh and P. Das, "A Review on Surrogate Safety Measures in Safety Evaluation and Analysis", 6th Conference of the Transportation Research Group of India, Lecture Notes in Civil Engineering, Vol. 273, pp. 113-129, 2021.

[18] B. N. Persaud, R. A. Retting, P. E. Garder, and D. Lord, "Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational Before-After Study", Transportation Research Record, Vol. 1751, pp. 1-8, 2001.

[19] Y. Chen, B. Persaud, E. Sacchi, and M. Bassani, "Investigation of models for relating roundabout safety to predicted speed", Accident Analysis and Prevention, Vol. 50, pp. 196-203, 2013.

[20] D. Gettman, "Surrogate Safety Measures from Traffic Simulation Models", Transportation Research Record Journal of the Transportation Research Board, No. 1840, 2003.

[21] O. Giuffrè, A. Grana, M. L. Tumminello, T. Giuffrè, and S. Trubia, "Surrogate Measures of Safety at Roundabouts in AIMSUN and VISSIM Environment", 15th Scientific and Technical Conference Transport, Katowice, Poland, 2018.

[22] M. de Lapparent, "Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas", Accident Analysis and Prevention, Vol. 38, pp. 260-268, 2006.

[23] B. Persaud, T. Saleem, M. E. Granados, T. Rajeswaran, "Can crash modification factors be estimated from surrogate measures of safety?", Advances in Transportation Studies, Vol. 1 pp. 149-162, 2018.

[24] A. Abdelhalim and M. Abbas, "An Assessment of Safety-Based Driver Behavior Modeling in Microscopic Simulation Utilizing Real-Time Vehicle Trajectories", 2022.

[25] A. Pratelli, L. Brocchini, and N. Francesconi, "Estimating and Updating Gap Acceptance Parameters for Hcm6th Roundabout Capacity Model Applications", WIT Transactions on Ecology and the Environment, Vol. 253, pp. 477-486, 2021.

[26] D. Muley, M. Ghanim, and M. Kharbeche, "Prediction of Traffic Conflicts at Signalized Intersections using SSAM", Procedia Computer Science, Vol. 130, pp. 255-262, 2018.

[27] M. S. Ghanim and K. Shaaban, "A Case Study for Surrogate Safety Assessment Model in Predicting Real-Life Conflicts", Arabian Journal for Science and Engineering, Vol. 44, pp. 4225-4231, 2019.

[28] M. Hasanvand, A. S. A. Nasiri, O. Rahmani, K. Shaaban, and H. Samadi, "A Conflict-Based Safety Diagnosis of SCI Roundabouts Using a Surrogate Safety Measure Model", Sustainability, Vol. 15, 2023.

[29] G. Aksoy, H. B. Celikoglu, and E. Gedizlioglu, "Analysis of toll queues by micro-simulation: results from a case study in Istanbul", Procedia – Social and Behavioral Sciences, Vol. 111, pp. 614-623, 2014.

[30] TRB, "Roundabouts: An Informational Guide - Second Edition". Washington, D.C.: The National, 2010.

[31] TRB, "Highway Capacity Manual 6th Edition: A Guide for Multimodal Mobility Analysis", Washington, D.C.: The National, 2016.

Streamlining AI: Techniques for Efficient Machine Learning Model Selection

Daniel Schönle
IDACUS Insitute
Furtwangen University
Furtwangen, Germany
schonledanielhfu@gmail.com

Christoph Reich
IDACUS Insitute
Furtwangen University
Furtwangen, Germany
christoph.reich@hs-furtwangen.de

Djaffar Ould Abdeslam
Institut IRIMAS
Université de Haute Alsace
Mulhouse, France
djafar.ould-abdeslam@uha.fr

***Abstract*—As machine learning (ML) systems become more ubiquitous, their resource requirements and associated financial burdens increase, highlighting the need to optimise energy consumption and costs to meet stakeholder expectations. While quality metrics for predictive ML models are well established, efficiency metrics are less commonly addressed. We present a comprehensive framework for evaluating efficiency metrics that facilitates the comparison of different types of efficiency. A novel efficiency metric approach (Compact Efficiency Metric) is proposed that considers resource usage, computational effort, and runtime in addition to prediction quality. Implementations for specific focus areas have been developed, such as the Quality-Focused Compact Efficiency Metric (QCO). This work also introduces a Pareto-based methodology for selecting ML models with an emphasis on efficiency. The QCO metric has undergone rigorous testing to validate its applicability, plausibility, and ability to adjust for variations in dataset size and hosting environment performance. This QCO metric has been applied to two datasets and calculated for a wide range of ML models. In particular, our metric identified an efficient solution when determining the optimal sequence length for transformer-based models. The results from Pareto-based selection were congruent with those derived from the QCO metric, providing a viable approach for pre-selecting preferred solutions. The framework enables stakeholders to make informed decisions concerning the utilisation and design of ML models, thereby ensuring environmental responsibility and cost-effectiveness.**

***Index Terms*—machine learning; nlp; efficiency; metric; software performance; automl.**

I. Introduction

In past decades, the imperative for computational efficiency was determined by the constraints of then-available technology, compelling the optimal use of limited hardware resources. With advancements in computing capabilities, particularly in machine learning (ML), focus has gradually shifted towards augmenting the quality of predictions, often at the expense of operational efficiency. The advent of large language models (LLMs) has been pivotal, marking a transformative phase in computational tasks and signalling their growing dominance in human-computer interactions. This shift coincides with the integration of ML in various environments, including edge computing, where resource limitations necessitate a critical reconsideration of computational approaches from perspectives of green computing and sustainability. Utilising resource-intensive techniques, such as transformer-based embeddings or LLM, introduces financial and environmental challenges, emphasising the necessity for enhanced computational efficiency. In response to these challenges, the oblique objective of this publication is to improve the efficiency of ML operations. By introducing a set of comprehensive metrics, this work aims to measure the efficiency of ML models, thereby facilitating their sustainable and economically viable deployment across different platforms [2].

The focus of ML research has been on improving model quality, for which a number of metrics are available. Research on effective ML lacks standardised and comprehensive efficiency metrics. To ensure reproducibility and facilitate comparison of results, best practices in ML research typically include detailed descriptions of experiments, including datasets, pre-processing steps, machine learning techniques, hyperparameters, and hardware setups [3]. The lack of dataset-agnostic procedures and the absence of 'golden standard' datasets pose challenges in achieving accurate reproducibility and fair comparisons in natural language processing (NLP) [4]. Furthermore, assessing the impact of innovations in ML process steps, such as improved pre-processing, on prediction quality is complicated by the influence of other ML steps.

The delicate balance between complexity and outcome is often overlooked in research efforts to utilise all available resources to reduce time-to-solution. Evaluating complexity in time and space is subordinated to finding the best model or process. Numrich stated [5]: "Increasing productivity by minimising the total time-to solution is a somewhat ill-defined statement of the problem. We propose an alternative statement: at each moment in time, use the resources available optimally to accomplish a mission within imposed constraints." In the context of ML research, it is important to establish metrics that address efficiency concerns alongside prediction quality.

We propose to fill the existing gap by introducing novel metrics for measuring the efficiency of machine learning models.

Note: This paper is a revised and extended version of [1].

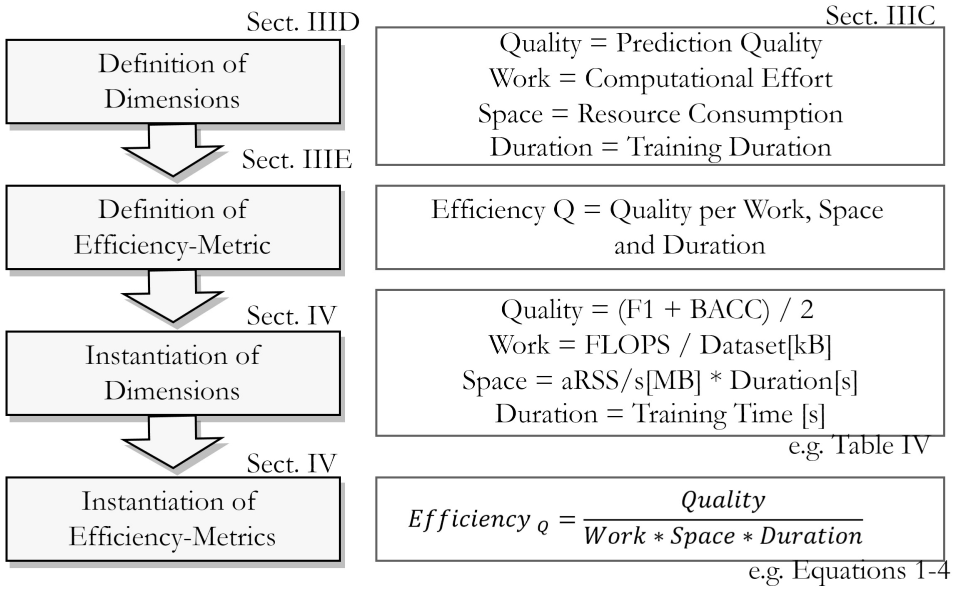

Fig. 1. Development of Quality Focused Efficiency Metric.

By incorporating resource consumption, computational effort, and runtime considerations into our efficiency metrics, we aim to provide a holistic perspective on the actual efficiency of ML models. We demonstrate the process of defining a quality-focused efficiency metric (Figure 1) and present the *Q*uality *CO*mpact (QCO) Efficiency Metric, detailed in Equations (4) and (5).

We acknowledge the significance of dataset-agnostic evaluation and propose solutions to overcome this challenge. Furthermore, we demonstrate the benefits of our metric for evaluating hyperparameter tuning. To facilitate the measurements, we have developed a tool to monitor the ML computing processes. Additionally, we introduce a Pareto-based approach to simplify selecting the most suitable ML setup. The aim is to enable researchers and practitioners to make informed decisions that prioritize prediction quality and efficiency, thus advancing the field of machine learning towards sustainable, green, and economically feasible solutions.

A. Summary of Contributions

The contributions of this research are as follows:

Efficiency Definition Following Physical Units. Defines machine learning efficiency using a framework analogous to physical units, providing a standardized approach to quantify and compare machine learning efficiency.

Definition of Efficiency Dimensions: Establishes key metrics such as quality, speed, and resource utilization.

Foundational Efficiency Metric: Introduces a basic efficiency model.

Focused Efficiency Metrics: Extends the Efficiency Metric to address specific aspects like computational cost or quality, allowing for targeted optimizations.

Definition of Instantiation Process: Applies the efficiency metrics on use cases, including selecting appropriate measurements, data transformation techniques, and determining validity ranges.

Exemplary Instantiation for QCO: Demonstrates how to implement an efficiency metric specifically focusing on quality.

Capturing Tool Observe: Develops a user-friendly tool to automatically capture data needed to evaluate the efficiency of machine learning algorithms, simplifying performance analysis.

Evaluation of Efficiency Metric: Discusses methods to assess the validity of efficiency metrics by comparing them to expert and Pareto solutions.

B. Comparison of Approaches

Table I compares traditional efficiency approaches and the solutions provided by the Compact Efficiency Metric, with particular emphasis on the advantages of the former. The comparison demonstrates the enhanced flexibility and applicability of the Compact Efficiency Metric across various computational contexts. It addresses the limitations of traditional methods by providing a structured, measurable framework for evaluating machine learning efficiency.

C. Outline

The structure of the paper is as follows: Section II (State of the Art) covers the research on efficiency types, metrics in ML, and the Pareto approach. In Section III, the efficiency metric is presented by elaborating on its objectives, followed by the theoretical foundations of *efficiency dimensions* and concepts, and finally, the definition of the efficiency metrics. In the subsequent Section IV, the metric for quality-focused efficiency is defined, adhering to a specified protocol. The score equations for QCO_F are presented, accompanied by a brief explanation of its usage. The Evaluation Section V uses

TABLE I
COMPARISON OF APPROACHES

Approach	Downside	Compact Efficiency Metric
Computational Cost	Complexity based on theoretical calculation	Empirical measurements
Computational Cost for Deep Learning	Model parameter based theoretical calculation	Empirical measurements
Resource Efficiency	Objective is a two-sided optimization of software and hardware	Resource consumption is emphasized, aiming to balance multiple efficiency dimensions
Efficiency Comparison	Restricted to comparing methods, not scoring	Provides comparable scores with defined validity ranges
Efficiency Metrics	Optimized for high-performance computing components with specific measurements	Measurements applicable on standard hardware and software
Pareto Efficiency	Multi-dimension optimization without dimension weights	Allows for weights to be assigned to each efficiency dimension

two experiments to assess the performance of the metric. The results obtained are discussed in detail in Section VI, leading to the presentation of the conclusion (Section VII).

II. State of the Art

This section provides an overview of the current state of the art and serves as a background chapter. Approaches that address computational cost as a method, with the aim of predicting both computational and financial costs, are considered first. Next, resource efficiency approaches are reviewed, aiming to find algorithms that operate cost-effectively. Additionally, generic approaches to efficiency metrics are examined. Finally, the concept of Pareto efficiency is introduced.

A. Computational Cost

Computational Cost or efficiency is based on the computational effort. Most statistical ML algorithms can be addressed, and their time complexity or space requirements can be calculated. For example, the time complexity of gradient descent is $O(ndk)$, where d is the number of features and n is the number of rows. In the context of transformer-based approaches, the number of operations for multi-head attention can be calculated as $n^2d + nd^2$, where n is the sequence length and d is the depth [6]. Translating these statistical calculations into real training times is challenging due to numerous optimisations of modern CPUs and GPUs that change the type of computation and the number of operations [7][8][9]. The approach defines work and duration dimensions based on actual measurements.

Computational Cost for Deep Learning is specific to deep learning due to its reliance on complex neural network architectures, which complicates the direct computation of complexity. Several approaches attempt to predict complexity, such as the model proposed by Li et al. [10], which introduces two classes of prediction models for distributed SGD. The use of profiling information in this approach is similar to the presented method, but it has limited validity for deep learning optimised with distributed SGD.

Resource Efficiency is essential for deep learning, where hardware requirements differ from statistical ML and constantly evolve. The research aims to adapt deep learning to specific hardware. Yang et al. [11] developed a method to bridge this gap, focusing on computing the model locally near the sensor. In HPC, research such as Performance Metrics based on computational action (Numrich [5]) optimises the use of hardware. Resource efficiency focuses primarily on the optimal hardware usage of specific algorithms, ignoring algorithm complexity or runtime. The efficiency definition addresses this aspect to provide comprehensive statements about the entire ML task.

B. Efficiency

Efficiency Comparison plays a role in evaluating novel approaches. For instance, Thomson et al. [12] present an optimisation for machine learning-based compilers that focuses on process speedup while overlooking the impact on resource consumption. Fischer et al. [13] propose a framework for evaluating the energy efficiency of ML without considering prediction performance. Kumar, Goyal & Varma [14] develop ML with a small footprint and compare efficiency based on model size, prediction quality, prediction time and prediction energy. Discussions of the novel approach primarily revolve around individual measurements, lacking an overall efficiency comparison. In contrast, Huang et al. [15] discuss the selection of an object detection architecture in terms of efficiency, defining it as a speed/memory/accuracy trade-off and evaluating it through two-dimensional trade-off curves. The proposed efficiency metric would provide a balanced and meaningful score for evaluating [14] and [15].

Efficiency metrics were 'invented' for HPC research, which deals with highly scaled hardware systems and highly specialised applications, making efficiency statements easier to derive and crucial. The difficulties have been recognised and discussed from an early stage [16] - to philosophical considerations [17]. Numrich of Cray Research developed an approach based on physical laws [18] [5], which inspired the proposed metric based on dimensions reflecting components of a physical law.

C. Pareto efficiency

In the field of machine learning, Pareto efficiency has emerged as a key criterion for evaluating and optimising algorithms under the constraints of multiple competing objectives.

Pareto efficiency delineates an optimal state where no single objective can be improved without simultaneously degrading another, thus embodying the essence of trade-offs inherent in decision-making processes.

Pareto efficiency is a critical factor in multi-objective optimization (MOO) [19] [20], where ML models are optimized across various dimensions, including accuracy, computational complexity, and energy consumption [21] [22]. This requires a departure from traditional single-objective optimization paradigms towards more nuanced approaches that can navigate the complex trade-offs among competing objectives. Several algorithms have been proposed to address MOO in ML. These include evolutionary algorithms [23] [24], swarm intelligence [25], and gradient-based techniques [26], each offering different mechanisms for approximating the Pareto front, which is the set of all Pareto efficient solutions.

Pareto efficiency has become a critical criterion for evaluating and optimising algorithms in the evolving ML landscape, where multiple competing objectives must be considered [27]. Pareto efficiency is a concept from economics that describes an optimal state in which it is impossible to improve one objective without degrading another, thus embodying the essence of trade-offs inherent in decision-making processes [28]. This chapter explains the applications of Pareto efficiency in machine learning, including its significance, evaluation methods, and implications for algorithm design and evaluation.

Recent advancements have led to the integration of Pareto efficiency with deep learning, particularly in areas such as neural architecture search (NAS) [29]. The objective is to discover optimal network architectures that balance performance with resource constraints. Studies have employed Pareto-based approaches to navigate the search space efficiently, identifying architectures that offer optimal trade-offs between accuracy and computational cost [30].

Furthermore, Pareto efficiency can be applied to optimisation and model evaluation and selection [31]. In this context, Pareto fronts serve as a valuable tool for evaluating the performance of various models, allowing practitioners to make informed decisions based on a comprehensive understanding of the trade-offs involved. This approach has proven advantageous in fields where performance is multifaceted and cannot be captured by a single metric. For instance, in recommender systems, accuracy, diversity, and novelty may all be of concern [32].

The investigation of Pareto efficiency in machine learning also poses fundamental questions regarding the nature of optimality and the objectives of optimization itself. It urges the community to reassess current practices and create new theoretical frameworks and algorithms to more accurately depict and navigate the intricate trade-off landscapes typical of real-world issues. In conclusion, integrating Pareto efficiency into machine learning research and practice represents a significant paradigm shift, promoting a more comprehensive and nuanced approach to optimization. As the field progresses, further exploration of Pareto-based methods is expected to provide significant insights into the design, evaluation, and application of machine learning systems, ultimately advancing the frontier of what is achievable in the domain.

III. Efficiency Metric Proposal

This proposal encompasses two integral parts: the development of abstract efficiency metrics and the definition of the quality-focused efficiency metric. The objectives, limitations, and use cases of efficiency in machine learning (ML) are introduced, and basic efficiency types based on trade-off relationships are established. Drawing inspiration from the laws of physics, efficiency is defined using *efficiency dimensions* for quality, work, space, load, and duration of the ML procedure, with each dimension comprising measurements of the ML process. Two types of metrics are presented: the *efficiency vector*, which provides insight into the raw strengths and weaknesses of the ML process in terms of efficiency, and the *focused efficiency scores*, which are designed for ease of interpretation. A defined procedure is employed to adjust dimensional weights and perform sophisticated measurement smoothing to enhance the significance of scores. For example, the metric equation for quality-focused efficiency is outlined, and a metric definition protocol is proposed to achieve metric validity. This protocol is applied to define the quality-focused efficiency metric, including the definition of the equation for score calculation, selecting appropriate measurements, smoothing measurement values, and elaborating dimension weights.

TABLE II
Limitations of *Q*uality *CO*mpact (QCO) Metric

Aspect	Category or Requirement	Evaluated	Compensation
Dataset	Dataset Independent	Labelled Text Samples	Dataset Size[1]
ML-Task	Complete or partial Pipeline	Text Classification Pipeline, NLP-Tasks[4]	No[2]
ML-Techniques	Technique Independent	Statistical and Deep Learning Techniques	No[3]
System Environment	Access to Measurements of Duration, Calculation Steps, Resource Consumption of ML-Task	Virtualised Host	No[3]
Host-Setup	Host-Setup Technique Independent	Non-HPC, Non-GPU	Compute Speed[1]

(1) Efficiency remains consistent regardless of value.
(2) Scores from different tasks are not comparable.
(3) Provides comparable efficiency scores per ML technique.
(4) untested

A. Objectives & Limitations

The methodology aims to ensure applicability across various machine learning techniques. It is designed to function on standard hosting setups, ensuring that measurements reflect balanced efficiency irrespective of differing host or system environments. Table II lists the requirements and the evaluated selection for each relevant aspect. The 'Compensation' column provides an overview of the metric compensation for variability, including dataset size and host setup performance. Measurements are conducted during the training phase of the classification model. Valid ranges for measurements must be aligned to the expected values. The validity ranges of the QCO Instance for the proof of concept are listed in Table III. Smoothing techniques are employed to standardise the validity of the data to ensure these measurements fall within comparable ranges.

TABLE III
VALIDITY RANGE FOR QCO INSTANCE

Aspect	Validity
Dataset	Size <256 MB
Training Duration	<48h
Calculation Amount	63P FLOPS, 800M Minor Page Faults
Resource Consumption	RAM <128GB

The application of the ML process is contingent upon the establishment of objectives and constraints. The following use cases are subject to specific efficiency requirements.

1) Effects of changes in the ML process: The effects of different techniques, such as pre-processing techniques, need to be measured [33].
2) Select ML technique by efficiency: Identify the ML technique that achieves high classification quality while minimising the use of computational resources [11].
3) ML technique for limited resources or private data: Select a Whitebox ML technique suitable for local model training [34].
4) Parameter optimisation: Effectiveness as a cost function in the optimisation of hyperparameters or setups [35].
5) Performance comparison: Compare the performance of an ML technique on different host setups to evaluate ML efficiency [36].
6) Predicting computational costs: Predicting the cost by predicting computational effectiveness of an ML technique in a production setup [37].

B. Dimensions

The efficiency approach may vary between applications but relies on similar elements. All concepts consider the trade-off between model performance and resource consumption during training or inference. The efficiency metrics focus on the following key components:

Accuracy or Performance. The efficiency of a machine learning model depends on its accuracy or performance. Several standard metrics can be employed to assess this, including accuracy, precision, recall, F1 score and area under the ROC curve (AUROC), depending on the task at hand.

Resource Utilisation. The efficiency should be evaluated regarding the resources consumed during the training or inference process. These resources include computational resources such as CPU, GPU, or memory usage. An efficient model should aim to minimise the utilisation of resources.

Relative Resource Utilisation. The load imposed on the host by the machine learning process provides a means of measuring the relative utilisation of hardware resources. A higher load indicates a more efficient use of available resources, as fewer resources are left unused.

Computational Effort. The efficiency of ML processes is contingent upon the complexity of the ML process itself, as well as the amount of computation required to train the model or to compute a result for inference. In order to enhance the efficiency of ML processes, it is necessary to minimise the computational effort.

Training Duration. The definition of efficiency encompasses the time required to train the machine learning model. Faster training times may be advantageous, particularly in instances where models must be trained frequently or where time constraints exist.

Inference Latency. In the context of models deployed in real-time or interactive applications, the time taken to make predictions or perform inference is critical. Low inference latency or fast response times can be important efficiency metrics in such cases.

In the context of cost-effectiveness in machine learning research, different dimensions or base units are considered. The need to define base units, such as distance and power, which can be used to define efficiency, has been discussed by Numrich [38]. The CO-Metric uses abstract dimensions that allow for customisation through flexible adaptation. The proposed efficiency metric uses the following *efficiency dimensions*, with a description of valid measurements:

Quality. (Or Performance) The machine learning model should achieve the desired level of accuracy as a performance indicator for addressing the given task or problem. This accuracy can be measured using appropriate evaluation metrics tailored to the specific task, including accuracy, precision, recall, F1-Score, or AUROC. The scores should compensate for any unbalanced datasets [6].

Work. (Or Computational Effort, Computational Complexity) The number of computational operations, such as matrix multiplications, gradient computations, data transformations, and the usage of computational cache (e.g., CPU L1-Cache). The theoretical amount of work can be calculated by applying the theory of computational complexity. The actual workload differs due to optimisation at the software and hardware level [7]–[9]. The measurement shall count generated and processed compute steps; optionally data transfers through memory and network. Computational steps can be counted directly (floating point operations or instructions) [6] or indirectly by measuring side-effects of computation, e.g., memory management activity.

Load. (Or Relative Resource Consumption) The relative host usage metric reflects the $degree$ to which all available host resources are utilised. This encompasses the relative usage of compute units (CPU and GPU cores), as well as relative memory usage. Additionally, it encompasses information per taining to load-related memory management events, such as major page faults.

Space. (Or Absolute Resource Consumption, Space Complexity) The $amount$ of data resources, such as memory and storage, required by the machine learning process. Memory space consumption is quantified by resource usage on the host system. This includes main memory usage and allocation, such as virtual memory allocation, resident set size, working set size or stack size.

Duration. (Or Time Requirements) Time-related measurements encompass metrics such as training duration and inference latency. These metrics quantify the time required to complete machine learning procedures and the duration spent on processing units.

Other non-dimension specific measures incorporate dataset characteristics, including the number of samples and the dataset size. Sample attributes like the number of sentences, words, and linguistic text properties are also considered for specialised metrics.

C. Efficiency

Efficiency (cost-effectiveness) is defined as the achievement of a high level of performance or accuracy while optimising the utilisation of resources and minimising associated costs. The objective is to balance the model's effectiveness (performance) and the costs or resources required to achieve that effectiveness. The CO-Metric approach encompasses three concepts of cost-effectiveness:

Solution Efficiency. Efficiency is defined as the balance between solution achievement and cost. Solutions are focused on quality, and costs include the efforts done and resources consumed. Every aspect is provided by one or multiple efficiency dimensions. Solution efficiency with a quality focus describes the computational effort used to achieve prediction quality. This reflects the efficiency of the model, i.e., the algorithm and its implementation. Efficiency increases by doing less work in less time and achieving higher prediction quality. Other focuses include achieving low latency of ML inference.

Resource Efficiency. Efficiency is defined as the degree to which resources are used. Resource efficiency is the capability of the ML procedure to utilise all available resources. It is increased by adapting to the host setup using more existing resources. This is important for designing hardware for specific ML techniques and adapting ML algorithms to specific hardware [39].

Synthetic Efficiency. Efficiency can be employed as a metric for the assessment of specific performance attributes. This may include the analysis of text quality indicators, as exemplified by the work of [40], or comparing performance outcomes [41].

Efficiency rules are defined based on the efficiency objectives:

1 Solution Efficiency
- 1.1 The more quality is achieved in less time, work and effort, the higher the ML quality efficiency.
- 1.2 The less time it takes to achieve more quality, the higher the ML-Speed-Efficiency.
- 1.3 The less work required for more quality, the higher the ML-Work-Efficiency.

2 Resource efficiency
- 2.1 The more load is used for more quality, less duration and less work; the higher the ML resource efficiency.

3 Synthetic efficiency
- 3.1 The less computational work is necessary per data chunk, the higher the ML model efficiency.

In addition to the efficiency objectives, there are two opposed requirements for handling ML efficiency results: interpretability and applicability. The greater the amount of information a metric provides, the greater the need for interpretation. This approach provides metrics at two levels of complexity. (i) Efficiency is determined as a single scalar by the at-a-glance metric (compact metric score) while supporting weights for each dimension. (ii) The efficiency vector metric represents uninterpreted values per dimension.

D. Compact Efficiency Metrics

Machine learning (ML) efficiency exhibits variability contingent upon the specific application. Metrics have been proposed for particular purposes and categorised according to their level of complexity. The group of compact metrics employs a subset of dimensions that contribute to calculating an efficiency score, which is determined with respect to the dominant dimension. The compact efficiency metric (CO) is specified in the Definition 1.

Definition 1: It exists a compact efficiency score CO of a ML procedure M for focused dimensions F with a focus weight α and unfocused dimensions U, defined in Equation (1); based on efficiency dimensions (D) quality q, work w, space s, load l and duration d with specific dimension weights β, detailed by Equation (2).

Quality Focused COmpact Efficiency Metric (QCO). A compact metric to reflect quality-focused efficiency. A score describes the best solution with a predefined high relevance of the quality dimension and low relevance of the work and duration dimensions. Relevant dimensions: Quality, Work, Space, Duration. Dominant dimension: Quality.

The QCO-score for an ML process M is derived from Equations (1) and (2) for the Quality-Focus, as stated by Equation (3). The quality dimension is represented by q, which measures the quality or performance of the machine learning model. w represents the dimension of computational effort, which quantifies the computational operations or effort required for the machine learning tasks. The resource consumption dimension s quantifies the system resources required during the execution of the model. d represents the dimension of duration, which measures the time or duration required to train the model. The weight per dimension β is employed to adjust the importance of dimensions, while α represents

$$[F]CO(M) = (F \times \alpha) \times U \tag{1}$$

$$[F]CO(M) = (q_M \times \beta_q) \times (w_M \times \beta_w) \times (s_M \times \beta_s) \times (l_M \times \beta_l) \times (d_M \times \beta_d) \times \psi \tag{2}$$

$$\text{where } D = \{r \in \mathbb{R} \,|\, r > 1\} \; and \; \{q, w, s, l, d \in D\}$$
$$F \subseteq D \quad \text{and} \quad U = D \backslash F$$
$$\psi = \text{Score-Compensation}$$

the additional weight of the focus dimension, both derived from expert knowledge of the use case. The compensation factor ψ is introduced to optimise the readability of the score, where $1 > \psi \geq 0.1$. The dominance of quality q is reflected in the numerator, so efficiency is defined as the quotient of quality divided by work w, space s and duration d (terms in the denominator). It is assumed that the dimensions will become increasingly important in proportion to their current size. Consequently, the weights β of the dimensions and the focus weight α are treated as exponents, with the respective dimension as the base.

$$QCO(M) = \frac{q^{\alpha * \beta_q}}{(w^{\beta_w} + s^{\beta_s} + d^{\beta_d})} * \psi \tag{3}$$

Resource Focused Compact Efficiency Metric (RCO). Compact metric to reflect resource-oriented efficiency. A score describes the best solution with a predefined high relevance of the relative load usage and a low relevance of the quality dimension. Relevant dimensions: Load, Quality, Work, Duration. Dominant dimension: Load.

Inference Focused Compact Efficiency Metric (ICO). Compact metric to reflect resource-oriented efficiency. A score describes the best solution with a predefined high relevance of duration, low relevance of the quality dimension and lowest relevance of work. Relevant dimensions: Quality, Work, Duration. Dominant dimension: Duration.

Algorithmic Focused Compact Efficiency Metric (ACO). Compact metric to reflect resource-oriented efficiency. A score describes the best solution with predefined high relevance of work and duration, low relevance of duration, quality, and dataset dimension. Relevant dimensions: Quality, Work, Duration, Dataset. Dominant dimension: Work.

E. Efficiency Vector Metric (EV)

The CO metrics condense efficiency information into a score. To provide information on the dimension-specific performance, the EV metric reveals the dimension scores of the CO metric. The EV is available per CO as a vector, to describe the efficiency in the vector space of the specific CO. For QCO the QEV is represented by a vector in a *Quality-Work-Space-Duration* space.

TABLE IV
INSTANTIATION PROTOCOL.

Step	Objective
1	Select Efficiency Metric
2	Define Validity Requirements
3	Setup and Conduct Experiment
4	Define Dimensions
5	Analyse measurements
6	Assign measurements to Dimensions
7	Define Validity Ranges
8	Normalisation of Measurement-Values
9	Determine Dimensional Weights
10	Define Score compensation factor
11	Define Score Equation

F. Pareto

In the field of machine learning, it is important to select the most efficient configurations to streamline usage. This can be achieved by considering efficiency metrics and making preliminary selections. The selection process should consider the requirements of the machine learning workflow, which reflect diverse priorities such as speed, resource utilization, and outcome quality. Using a Pareto-based approach to select the most efficient configurations aligns well with the given criteria. This methodology identifies configurations that offer an optimal trade-off among competing objectives, ensuring that the selected setups are not only efficient but also tailored to the specific demands of the machine learning process.

IV. COMPACT METRIC INSTANTIATION

To operationalize the proposed efficiency metric, the abstract definitions must be instantiated. The instantiation is akin to the object instantiation of a class in programming. Attributes like the efficiency dimensions are set, and the metric can then be applied to measurements to receive scores. These attributes are defined based on empirical evidence, valid within specified ranges and tailored for particular use cases. This paper focuses on a text classification use case aimed at optimizing classification quality for two distinct datasets. The instantiation stages for a CO-Metric are listed in Table IV. For reasons of brevity, the instantiation is limited to the QCO metric.

A. Use Case

In this demonstration of the instantiation process Use Case 1 was selected, the aim is to identify the most efficient machine learning method for a set of text classification tasks, prioritising efficiency - a measure of quality per unit of work, space and time - over pure classification quality. The Quality COmpact (QCO) score is employed to assess this efficiency. The use case involves two different text classification tasks: spam classification, using the dataset of Almeida et al. [42], and sentiment analysis of movie reviews, based on the dataset of Maas et al. [43]. Vectorisation methods include traditional TF-IDF and advanced word embeddings using DistilBERT. Classifiers selected include Support Vector Machines, Naïve Bayes, Gradient Descent, Random Forest and a transformer-based method using a fine-tuned DistilBERT model for both vectorisation and classification tasks.

TABLE V
VALIDITY REQUIREMENTS

Aspect	Count	Variables	Optional
Dataset	>=2	Size, Sample Count	Sample Length, Language
Vectorization	>=2	Algorithm	Dictionary Size, Model Size
Classifier	>=4	Algorithm, Classifier Tech.	Hyperparameters
Host-Setup	>=2	Hardware Conf., Operating System	Software Version

B. The QCO-Metric Instance

The dimensions and the QCO score are instantiated according to the protocol given in Table IV. The required validity for different datasets and ML procedures results in the empirical variance requirements presented in Table V. To gain comparison validity among host-setups, four different computing environments were set up (No. 1-4 as shown in Table VI).

TABLE VI
HOST-SETUPS

No.	Type	CPU-Model	Clock	Threads	RAM
1	Virtualised	AMD Ryzen 7 5800U	1.9	8	16
2	BareMetal	Intel Core i5-6200U	2.3	4	8
3	BareMetal	Intel Core i7-7700	3.6	16	32
4	Virtualised	Intel Xeon Gold 6230	2.1	4	8
5	Virtualised	AMD EPYC 7742	2.2	16	16

[Clock in GHz, RAM in GB.]
OS: Linux, Language: Python3,
Libraries: Scikit-learn [44], DistilBERT [45], torch [46], pandas [47].

C. Observe Tool

To facilitate a reusable instrument for gauging efficiency, an observation tool was developed. This tool is designed to monitor the performance metrics of a Linux process. Initially, it places the target process in a suspended state, thereby preparing the environment for the commencement of measurements. Subsequently, it activates three distinct measurement tools (Figure 2) and resumes the operation of the process under scrutiny. Upon the completion of the process, the observation tool automatically terminates the measurement utilities.

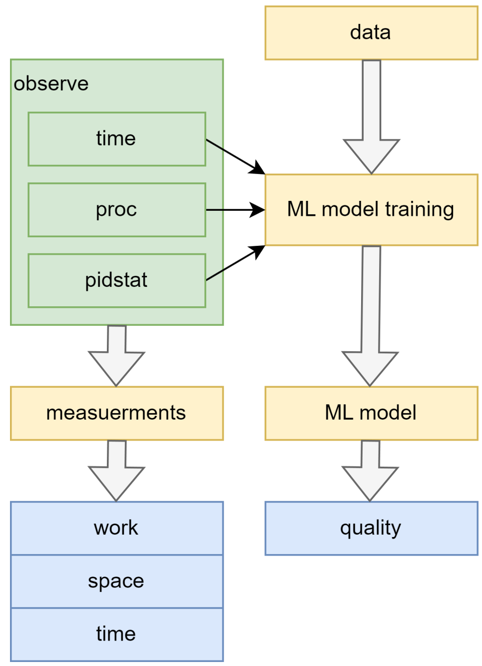

Fig. 2. Observe Tool

The measurements were provided by a set of Linux tools:

- `time`. Basic process measurement (CPU, Memory).
- `pidstat`. Advanced process measurement (CPU, Memory, IO-Usage).
- `perf`. Performance counter capture. (CPU, Memory).

The acquired dataset necessitates transformation for analytical purposes. The utility `time` generates output in a textual format, which requires conversion into a comma-separated values (CSV) format to facilitate subsequent data processing and analysis. Similarly, the `perf` utility collects performance metrics in a multi-table format, necessitating transformation into CSV format to enable efficient data manipulation and interpretation. Furthermore, the `pidstat` utility produces continuous output, which must be processed to extract minimal, maximal, and mean values over the specified duration. This transformation is essential for summarising performance characteristics and facilitating a comprehensive analysis of system behaviour under varying conditions.

Quality scores were computed separately from the ML procedure. Measurements were grouped for resource domain, e.g., memory consumption or computational work on CPU. The groups were filtered by correlation, the heatmap (Figure 3) shows Pearson Correlation Coefficients for selected measurements. `perf` was not supported on all host setups due to missing performance counters and conflicts with power

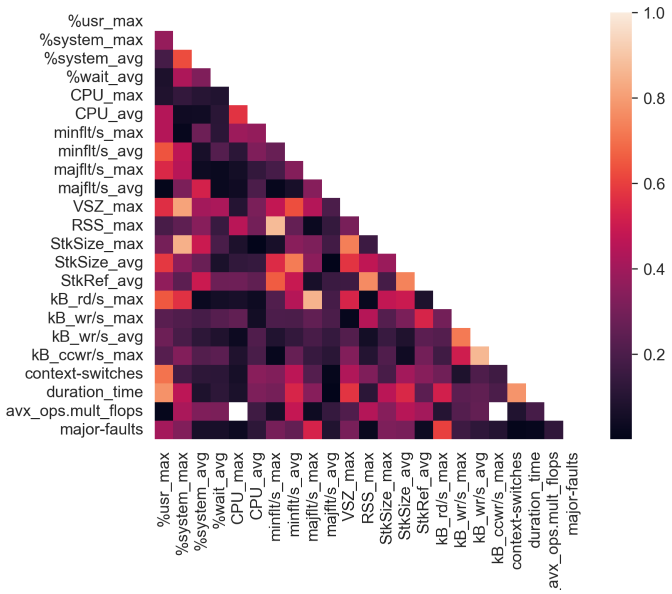

Fig. 3. Pearson Correlation Coefficients of Empirical Measurement Values.

TABLE VII
MEASUREMENTS

TYPE	DIM	IMP	TRANS	DEP	RNG
F1-Score	Quality	10	None		0-1
Bal.-Acc.	Quality	10	None		0-1
FLOPS	Work[CPU]	10	Log 63P	Data	0-63P
MinorPF	Work[CPU]	5	Log 800M	Data	0-800M
RSS (avg)	Space[Mem]	10	Log 128G	Time	0-128G
CPU Time [ns]	Duration	10	Log 172T		0-172T
Data Size	-	10	Log 256M		0-256M

TYPE=Measurement; DIM=Dimension; IMP=Weight; TRANS=Data transformation; DEP=Dependency; RNG=Range of validity

saving methods. Two sets of measurements have to be set up, which results in two QCO flavours: Floating Point Operation (QCO_F) based and Minor Page Fault (QCO_F) based. The selected measurements are listed in Table VII.

Range definition is essential for normalisation. The valid ranges for this QCO-Instance are listed in Table III. Normalisation is necessary due to the use of different units and data types in calculations. Monotonic data transformations on dimension values result in a range between 0 and 2, based on the maximum values per dimension. Consequently, valid ranges are not related to measurement ranges. The definition of valid measurement ranges, as shown in Table III, facilitates data transformations on measurement values. After transformation, values are placed on a closed scale with a slightly decreased distribution (Figure 4).

The dimension-equations are defined by interpreting the dependencies of the measurements (Table VII). Especially the dependency on duration and data size has been considered.

The dimension weight is used to adjust the importance of the focused metrics. The importance of quality is based on domain knowledge: Quality is about two times more important than work, space, and duration which delivers $\beta_Q = 6$. Readability compensation ψ is set to 10.

QCO is defined for each set of measures, resulting in Equations (4) and (5).

TABLE VIII
QCO DIMENSION INSTANCES

Dimension	QCO_F	QCO_P(*)
Quality	(F1 + BACC) / 2	(F1 + BACC) / 2
Work	FLOPS / Dataset[kB]	Minor PF / Dataset[kB]
Space	aRSS/s[MB] * Duration[s]	aRSS[MB] * Duration[s]
Duration	Time on CPU [ns]	Time on CPU [ns]

(*) FLOPS-Measurement was not available on all hosts.

D. QCO Metric Usage

1) Select QCO type according to available measurements. If CPU-Performance-Counters are available QCOF, otherwise QCOP. Respect expected validity ranges (Table III.
2) Perform training on a dataset subset while capturing measurements according to Table VIII.
3) Calculate efficiency by Equations (4) and (5).

E. QCO Score Calculation

The Quality-Focused Score is calculated for FLOPS-based-score as QCO_F (4) and QCO_P for Page-Fault-based score (5).

F. Pareto Implementation

The implementation of the Pareto principle has been used to address three critical dimensions: quality, space, and workload efficiency. To improve the utility and manageability of the Pareto set, a compression technique has been employed. This approach involves removing points within the set that exhibit a high degree of similarity, with a predefined similarity threshold set at 50% of the value range. This methodological adjustment ensures a streamlined and representative Pareto set, making it easier to identify optimal solutions across the specified dimensions.

V. EVALUATION

The applicability, plausibility and balance of the proposed metric is assessed in a comprehensive evaluation.

A. Experiments

In Experiment 1, preliminary studies were conducted to design the text classification pipeline. This included the selection and configuration of appropriate measurement tools, as well as the implementation of hyperparameter adjustments to set up text classification algorithms.

In Experiment 2, Use Case 1 was adapted to binary classification tasks, which were performed by different vectorization and classifier technologies. Two datasets are selected for Experiment 2, both with moderate text length; SMS Spam Classification (25.000 samples) [42] and Movie Survey Classification (7.805 samples) [43]. The experiments were

$$QCO_F(M) = \frac{((\frac{F1+BACC}{2})^6)}{(\log_{63P} FLOPS/DS[kB] + \log_{128G} RSS[MB] * log_{864M} D[s] + \log_{172T} TOC[ns]} * 10 \tag{4}$$

$$QCO_F(M) = \frac{((\frac{F1+BACC}{2})^6)}{(\log_{800M} MPF/DS[kB] + \log_{128G} RSS[MB] * log_{864M} D[s] + \log_{172T} TOC[ns]} * 10 \tag{5}$$

where $F1$ = F1-Score, $BACC$ = Balanced Accuracy Score,
$FLOPS$ = Floating Point Ops., MPF = Minor Page Faults,
DS = Dataset-Size, RSS = Resident Set Size,
D = Duration, TOC = Time on CPU

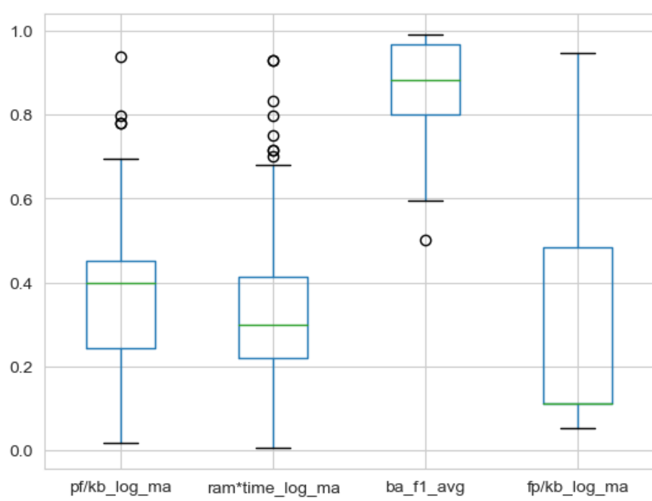

Fig. 4. Spread and Skewness per Dimension after logarithmic smoothing.

run on host 1 (Table VI) in two virtual hosts with different virtualisation technologies. The results of experiment 2 are shown in Table IX. To compare the QCOF and QCOP metrics in Experiment 2, two sets of QCO had to be created as some FLOPS measurements were not available (QCO_1 & QCO_2*).

In Experiment 3, the metric was further evaluated by applying it to an optimisation problem similar to Use Case 4. The objective was hyper-parameter optimisation with efficiency as the cost function. The ML process involved fine-tuning a transformer model (DistilBERT [45]), word embedding and text classification. The experiment aimed to find the most efficient value for the Maximum Sequence Length (MSL) for the SMS spam detection task [42], which was run on host 5 (Table VI).

B. Applicability

Experiment 2 shows surprising results that can be explained by runtime conditions such as schedulers, competing processes and caching techniques. The experiment is not designed to make general statements about specific combinations of vectorization or classification methods. Consequently, the following statements apply only to this experiment, which does not preclude testing the usefulness of the efficiency metric. The word embedding method is, on average, superior to the TFIDF in terms of quality, but there are classifiers (NB, GD) that can compensate for the quality disadvantage and, in some cases, achieve the highest efficiency. This is due to the low workload. The transformer method requires significantly more work. It achieves high quality, but also takes the longest time. The Random Forest (RF) classifier has a low efficiency because it requires a lot of computation and time to achieve good quality. The Support Vector Machine (only linear kernel) classifier benefits most from the word embeddings and, therefore, achieves good efficiency. When comparing the combinations in terms of the time-to-work ratio (WO-Focus), the worst ratio (1.28) is found for IMDB/TFIDF/SVM and the best for SMS/TFIDF/NB with 0.39. This leads to the conclusion that the measurement of time does not reflect the amount of work.

In Experiment 3, both QCO_F and QCO_P were successfully computed (see Table X). The most efficient MSL configuration consisted of 512 tokens, resulting in a high classification quality and moderate duration. On the other hand, the configuration with 126 tokens showed an increased workload and duration. The fastest result was obtained with an MSL of 256 tokens.

C. Plausibility

QCO was successfully generated for all ML methods in Experiment 2. A comparative assessment of QCO based on expert rankings is used for evaluation. Domain experts ranked the dimensions, listed in Table IX Column Rank-EXP. Comparing expert and QCO rankings, a minimal deviation from the expert rank was observed for high-quality ML methods, but the deviation increased with decreasing quality. This variance can be attributed to the expert's specific weighting of quality

TABLE IX
QCO EVALUATION RESULTS

				$EV Metric$					QCO Metrics		Rankings				
DAT	VECT	CLF	DUR	QUA	$TIME$	SPA	WO_P	WO_F	QCO_P	QCO_F	EXP_1	QCO_{1P}	EXP_2	QCO_{2P}	QCO_{2F}
SMS	TFIDF	NB	00:00:34	0.968	0.407	0.260	1.043	0.691	1.260	1.726	1	1	1	1	1
SMS	TFIDF	GD	00:02:36	0.972	0.583	0.371	1.043	0.631	1.190	1.678	2	2	2	2	2
IMDB	TFIDF	GD	00:00:19	0.885	0.339	0.222	0.616	0.610	1.145	1.153	3	3	3	3	3
SMS	DIST-T	DIST-T	00:01:34	0.982	0.525	0.384	1.249		1.099		6	4			
SMS	BERT	SVM	00:11:29	0.972	0.755	0.510	1.143	1.465	1.018	0.852	4	5	5	4	4
IMDB	DIST-T	DIST-T	02:38:32	0.982	1.058	0.795	1.058		0.970		5	6			
SMS	BERT	GD	02:04:00	0.978	1.030	0.696	1.140	1.637	0.956	0.752	8	7	4	5	6
IMDB	DISTIL	DISTIL	11:31:52	0.978	1.228	0.923	1.136		0.848		7	8			
IMDB	TFIDF	NB	00:01:18	0.845	0.504	0.330	0.618	0.581	0.766	0.797	9	9	6	6	5
SMS	BERT	NB	01:58:30	0.932	1.024	0.692	1.141	1.638	0.715	0.563	11	10	8	7	8
SMS	DISTIL	DISTIL	01:39:58	0.983	1.005	0.750	1.845		0.697		12	11			
SMS	BERT	RF	01:09:50	0.916	0.963	0.651	1.140	1.639	0.660	0.516	10	12	7	8	9
IMDB	TFIDF	RF	00:00:58	0.809	0.469	0.309	0.617	0.646	0.604	0.586	15	13	9	9	7
SMS	TFIDF	RF	00:03:02	0.787	0.601	0.376	1.043	0.931	0.335	0.363	16	14	12	10	10
IMDB	TFIDF	SVM	00:20:20	0.714	0.821	0.554	0.638	0.812	0.223	0.194	13	15	11	11	11
SMS	TFIDF	SVM	00:00:35	0.652	0.409	0.264	1.048	1.092	0.117	0.113	14	16	10	12	12

Columns: Dataset, Vectorizer, Classifier, Duration, Quality, Time, Space, WO_F= Work (FLOPS), WO_P = Work (Minor Page Faults), QCO Metrics, Rankings by Domain EXPerts, or QCO,
Abbrev.: SMS = SMS Spam Dataset [42], IMDB = IMDB Dataset [43], BERT = BERT word embedding, DIST-T = finetuned DistilBERT word embedding (PyTorch) & classification, DISTIL = finetuned DistilBERT word embedding (TensorFlow + keras) & classification, GD = Gradient Descent, SVM = Support Vector Machine, NB = Naïve Bayes, RF = Random Forest

TABLE X
EFFICIENCY OF DISTILBERT

	Measurements		Dimensions				Scores	
SL	Duration	F1	Q	W	S	T	QCO_F	QCO_P
128	09:08:50	0.76	0.64	3.02	0.45	0.78	0.159	0.164
256	00:27:02	0.78	0.64	2.86	0.45	0.71	0.171	0.172
512	00:50:13	0.78	0.65	2.94	0.47	0.72	0.183	0.174

Text Classification Efficiency with DistilBERT with different maximum Sequence Length (SL). Smoothed Dimensions: Quality, Work, Space and Time. Efficiency Scores Quality-Focused based on FLOPS (QCO_F) and Minor Page Faults (QCO_P)

relevance, which is particularly evident in the DistilBERT setups.

D. Balance & Compensation

QCO achieved a balance of aspects through compensation (Table III). The results of Experiment 2 showed no anomalies for different datasets; even ML processes with large datasets achieved high efficiency. Moreover, significant differences in speed and computational complexity were observed for comparable efficiency, suggesting a balance in these aspects. Due to the small number of hosts available for evaluation, the balance on host setups could not be verified.

E. Pareto Application

Pareto sets were successfully established for both datasets, as delineated in Table XI, optimizing for maximization of quality while minimizing work and space requirements. The initial composition of these sets included two distinct points for the IMDB and seven for the SMS Spam datasets. Subsequent refinement involved the elimination of points exhibiting significant similarity, with the threshold set at 50% of the value range, leading to a reduction in the SMS Spam Dataset's Pareto set (Figure 6), while the IMDB Dataset's set remained unchanged (Figure 5). The optimal Pareto front was characterized by the most favourable values within the set, namely the highest quality alongside the minimum for both work and space. The distance to this Pareto front was computed to evaluate the proximity of solutions to the optimal trade-off.

Analysis of the results for the IMDB Dataset [43] revealed that the fastest solution was achieved through TFIDF Vectorization combined with Gradient Descent, whereas the best quality solution was attributed to a fine-tuned DistilBERT model. For the SMS Spam Dataset [42], three top-quality performers emerged: TFIDF combined with Naïve Bayes, BERT with Gradient Descent, and a fine-tuned DistilBERT. A balanced solution, offering a compromise between quality and computational efficiency, was identified as TFIDF combined with Random-Forest. These findings highlight the efficacy of employing a Pareto-based approach to identify optimal solutions that cater to varying optimization dimensions across different datasets.

VI. DISCUSSION

This study proposes an efficiency metrics framework for machine learning techniques that addresses different aspects of cost-effectiveness, resource utilisation and model performance. The approach is intended to be adaptable and applicable to a variety of ML techniques and host setups.

The objectives of the efficiency metrics framework have been defined with the intention of addressing different real-world scenarios and use cases. The proposed efficiency metrics provide information that can be used to identify the optimal ML technique and hyperparameters, select ML techniques for

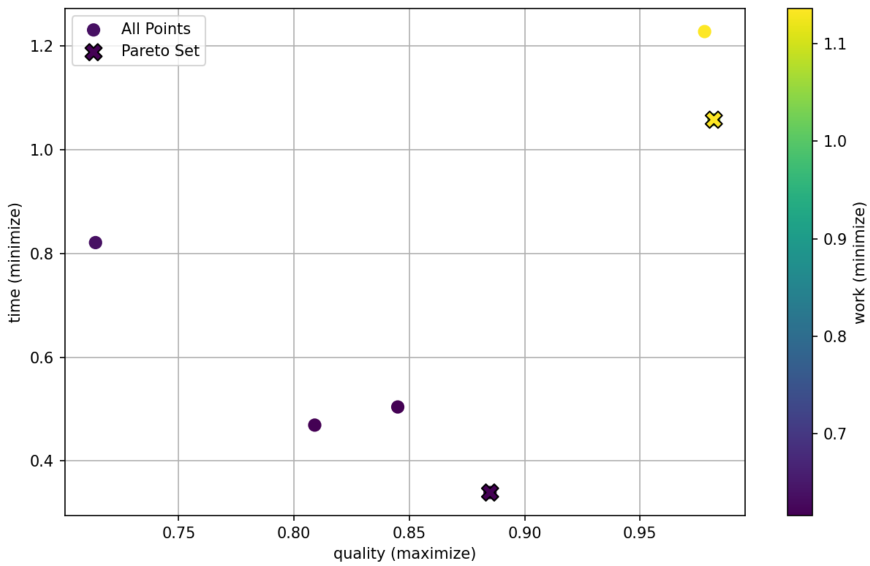

Fig. 5. Pareto results IMDB Dataset [43]

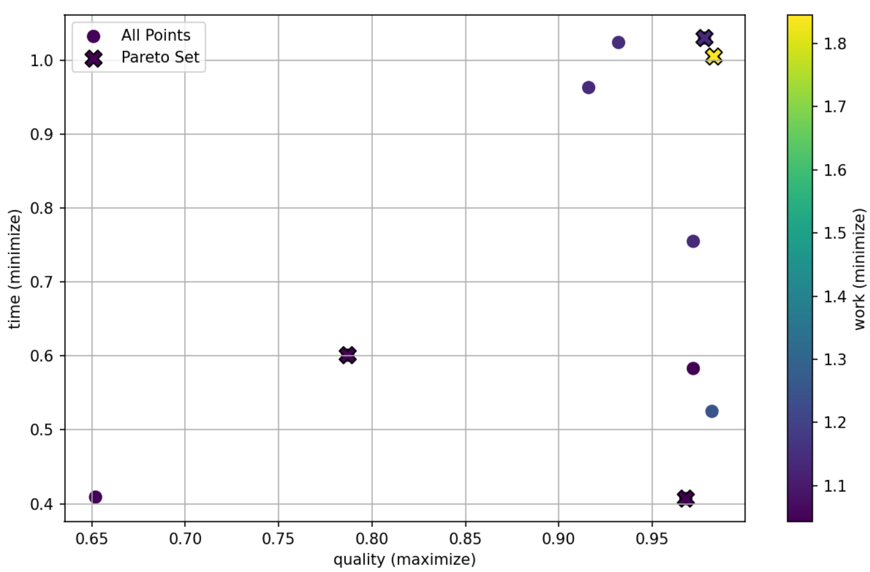

Fig. 6. Pareto results for SMS Spam Dataset [42]

TABLE XI
PARETO SETS

DAT	VECT	CLF	DUR	QUA	Time	SPA	Work	QCOP	Distance
SMS	TFIDF	NB	00:00:34	0.97	0.41	0.26	1.04	1.26	0.02
IMDB	TFIDF	GD	00:00:19	0.89	0.34	0.22	0.62	1.15	0.10
IMDB	DISTIL	DISTIL	02:38:32	0.98	1.06	0.80	1.06	0.97	0.84
SMS	BERT	GD	02:04:00	0.98	1.03	0.70	1.14	0.96	0.63
SMS	DISTIL	DISTIL	01:39:58	0.98	1.01	0.75	1.85	0.70	1.00
SMS	TFIDF	RF	00:03:02	0.79	0.60	0.38	1.04	0.34	0.28

Columns: Dataset, Vectorizer, Classifier, Quality, Time, Space, Work, $QCOP$ Metric, Distance to Pareto Front,
Abbrev: SMS = SMS Spam Dataset [42], IDB = IMDB Dataset [43], BERT = BERT word embedding, DIST = finetuned DistilBERT word embedding (PyTorch) & classification, DISTIL = finetuned DistilBERT word embedding (TensorFlow + keras) & classification, GD = Gradient Descent, SVM = Support Vector Machine, NB = Naïve Bayes, RF = Random Forest

limited resources or private data, compare classification performance across different host setups, and estimate computational cost.

The metrics framework introduces several dimensions that collectively capture the efficiency of machine learning techniques in achieving the aforementioned goals. These dimensions include quality, work, load, space and duration, each of which contributes to the overall efficiency score. The dimensions are intended to assess different aspects of machine learning performance and resource use, allowing for a comprehensive evaluation.

A key advantage of the proposed framework is its adaptability to different ML techniques and tasks. The dimensions and metrics can be adjusted based on specific use cases and requirements, ensuring relevance and accuracy in different contexts. This adaptability makes the metric framework suitable for a wide range of applications, from small-scale experiments to large-scale production systems.

The efficiency metrics introduced in the framework, such as QCO, FCO, ICO and ACO, provide different perspectives on efficiency. These compact metrics provide a clear, at-a-glance view of efficiency, making it easier for researchers and practitioners to evaluate and compare different ML techniques. In addition, the Efficiency Vector (EV) metric provides detailed information about the performance of ML techniques on individual dimensions, providing insights for further analysis and improvement.

The process of instantiating the efficiency metrics necessitates empirical investigation to ensure that the metric definitions are concrete and applicable to specific ML experiments. The validity of metric instantiation is emphasised, and the size of the experiment plays an important role in achieving reliable results. By conducting experiments on different datasets and host setups, the metric instantiation gains credibility and comparability.

The development of the observational tool provides a pragmatic approach to incorporating efficiency measurements into the machine learning (ML) toolchain. The implementation of an out-of-the-box efficiency measurement tool represents a significant contribution in several ways. First, it provides a tangible benefit to practitioners and researchers in the field of machine learning by facilitating the direct assessment of computational efficiency. This tool enables users to benchmark and optimise the performance of ML algorithms and systems, thereby improving the overall efficiency of the ML development lifecycle. Secondly, by providing a standardised method for measuring efficiency, it contributes to the reproducibility and comparability of experimental results, which are fundamental to the scientific rigour of machine learning research. Finally, the tool's accessibility and ease of integration into existing ML toolchains encourages its adoption, thereby broadening its impact on the field through improved performance evaluation practices.

In summary, the efficiency metrics framework developed in this study presents a robust approach to quantifying and comparing the efficacy of machine learning methods in terms of both performance and resource consumption. Its flexible and comprehensive nature makes it an invaluable resource for the machine learning community, aiding in the strategic decision-making process regarding the deployment of machine learning models and the allocation of computational resources. The continued refinement and application of this framework are poised to significantly influence the efficiency and sustainability of future machine learning endeavours.

VII. CONCLUSION AND FUTURE WORK

The successful computation and evaluation of efficiency scores represent a step forward in improving the effectiveness of machine learning research. By incorporating sophisticated dimensions that reflect measurement interdependencies - such as FLOPS relative to data volume or memory usage relative to duration - we could effectively balance the metric to account for variations in dataset size and host setup. A lack of sufficient samples challenged the evaluation of the QCO dimensions, which limited the robustness of our findings. Nevertheless, the FLOPS-based evaluation demonstrated consistency, and our innovative use of a minor page fault measure to extend the work dimension met with limited success. For future efforts, exploring efficiency dimensions that integrate ML-specific factors, such as model size, would be prudent to deepen our understanding of efficiency dynamics in different machine learning contexts. It is pertinent to highlight that the validation methods used for the proposed efficiency metric rely heavily on expert consensus. While this approach provides insightful feedback, the need for improved statistical validation is evident to strengthen the metric's credibility and structural integrity. Expert insight and the relevance of quality within specific applications particularly shape the effectiveness of machine learning methods. However, if quality is considered the sole criterion for development, it is important to recognise the potential for a significant escalation in complexity. Achieving a balance between different dimensions of efficiency is essential to ensure a pragmatic and reasoned strategy for optimising machine learning workflows.

A. Future

For further research, several key areas have been identified that require in-depth exploration to improve the understanding and implementation of machine learning efficiency metrics.

Comparability and Generalisability: This research has advanced the field by addressing comparability through the identification of dependencies within the data. The proposed compensations have validated the conceptual framework. However, a more detailed identification of these dependencies and a refinement of the compensation procedures remain crucial. Future studies should consider additional variables, such as those listed in the 'optional' column of Table V, or the complexity of the dataset to enhance the granularity and applicability of the results.

Applicability: The measurement tool presented in this study should be developed into a comprehensive library that simplifies the process of obtaining efficiency scores. This development would make the tool more accessible to a wider user base. In addition, creating solutions tailored for cloud-based computing environments will extend the utility of the tool and facilitate its widespread adoption in different computing environments.

Validation: Expanding the scope of experiments is essential for a deeper understanding of the biases present in current efficiency metrics. Comprehensive statistical investigations should be conducted to robustly validate these metrics, reducing reliance on expert opinion and strengthening the empirical foundations of efficiency measures.

Recurrent optimisation: Continuous improvement through recurrent optimisation processes is essential. Efficiency considerations should be routinely integrated into the development cycle of machine learning algorithms. This ongoing refinement will ensure that algorithms are not only effective but also optimised for efficiency and adapted to evolving computing environments and requirements.

AutoML: Integration of efficiency analysis within the AutoML framework to enhance automatic model selection and optimisation processes by incorporating efficiency metrics.

By addressing these areas, future research can significantly contribute to the sophistication and practicality of efficiency metrics in machine learning, paving the way for green machine learning.

ACKNOWLEDGMENTS

This Research was funded by the Federal Ministry of Education and Research of Germany in the framework of FiSK (Project-Number 16DHB4005).

REFERENCES

[1] D. Schönle, C. Reich, and D. O. Abdeslam, "Comparable Machine Learning Efficiency: Balanced Metrics for Natural Language Processing," in *GREEN 2023*, 2023.

[2] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia tools and applications*, vol. 82, no. 3, pp. 3713–3744, 2023.

[3] D. Rousseau and A. Ustyuzhanin, "Machine Learning scientific competitions and datasets," in *Artificial Intelligence for High Energy Physics*, World Scientific, 2022, pp. 765–812. DOI: 10.1142/9789811234033_0020.

[4] A. Ittoo and A. van den Bosch, "Text analytics in industry: Challenges, desiderata and trends," *Computers in Industry*, vol. 78, pp. 96–107, 2016. DOI: 10.1016/j.compind.2015.12.001.

[5] R. W. Numrich, "Performance metrics based on computational action," *The International Journal of High Performance Computing Applications*, vol. 18, no. 4, pp. 449–458, 2004.

[6] A. Vaswani *et al.*, "Tensor2Tensor for Neural Machine Translation," in *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, 2018, pp. 193–199.

[7] D. Ghimire, D. Kil, and S.-h. Kim, "A survey on efficient convolutional neural networks and hardware acceleration," *Electronics*, vol. 11, no. 6, p. 945, 2022. DOI: 10.3390/electronics11060945.

[8] G. Tzanos, C. Kachris, and D. Soudris, "Hardware acceleration on gaussian naive bayes machine learning algorithm," in *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2019, pp. 1–5. DOI: 10.1109/mocast.2019.8741875.

[9] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 15, pp. 4436–4449, 2019.

[10] Z. Li, M. Paolieri, L. Golubchik, S.-H. Lin, and W. Yan, "Predicting throughput of distributed stochastic gradient descent," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 11, pp. 2900–2912, 2022. DOI: 10.1109/tpds.2022.3151739.

[11] T.-J. Yang, Y.-H. Chen, J. Emer, and V. Sze, "A method to estimate the energy consumption of deep neural networks," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 2017, pp. 1916–1920. DOI: 10.1109/acssc.2017.8335698.

[12] J. Thomson, M. O'Boyle, G. Fursin, and B. Franke, "Reducing training time in a one-shot machine learning-based compiler," in *International workshop on languages and compilers for parallel computing*, 2009, pp. 399–407. DOI: 10.1007/978-3-642-13374-9_28.

[13] R. Fischer, M. Jakobs, S. Mücke, and K. Morik, "A Unified Framework for Assessing Energy Efficiency of Machine Learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2022, pp. 39–54. DOI: 10.1007/978-3-031-23618-1_3.

[14] A. Kumar, S. Goyal, and M. Varma, "Resource-efficient machine learning in 2 kb ram for the internet of things," in *International conference on machine learning*, 2017, pp. 1935–1944.

[15] J. Huang *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7310–7311. DOI: 10.1109/cvpr.2017.351.

[16] H. Westphal and S. Menge, "On the Supervisory Control of Distributed High-Performance Computing Systems in Engineering," *WIT Transactions on Information and Communication Technologies*, vol. 11, 1970.

[17] D. F. Snelling, "A philosophical perspective on performance measurement," in *Computer benchmarks*, 1993, pp. 97–103.

[18] R. W. Numrich, "Computational force, mass, and energy," *International Journal of Modern Physics C*, vol. 8, no. 03, pp. 437–457, 1997. DOI: 10.1142/s0129183197000370.

[19] K. Deb, K. Sindhya, and J. Hakanen, "Multi-objective optimization," in *Decision sciences*, CRC Press, 2016, pp. 161–200. DOI: 10.1201/9781315183176-4.

[20] Y. Cui, Z. Geng, Q. Zhu, and Y. Han, "Multi-objective optimization methods and application in energy saving," *Energy*, vol. 125, pp. 681–704, 2017.

[21] P. Ngatchou, A. Zarei, and A. El-Sharkawi, "Pareto multi objective optimization," in *Proceedings of the 13th international conference on, intelligent systems application to power systems*, 2005, pp. 84–91.

[22] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002. DOI: 10.1109/4235.996017.

[23] M. Kim, T. Hiroyasu, M. Miki, and S. Watanabe, "SPEA2+: Improving the performance of the strength Pareto evolutionary algorithm 2," in *Parallel Problem Solving from Nature-PPSN VIII: 8th International Conference, Birmingham, UK, September 18-22, 2004. Proceedings 8*, 2004, pp. 742–751. DOI: 10.1007/978-3-540-30217-9_75.

[24] C. A. C. Coello, *Evolutionary algorithms for solving multi-objective problems*. Springer, 2007.

[25] H. Han, W. Lu, L. Zhang, and J. Qiao, "Adaptive gradient multiobjective particle swarm optimization," *IEEE transactions on cybernetics*, vol. 48, no. 11, pp. 3067–3079, 2017. DOI: 10.1109/tcyb.2017.2756874.

[26] R.-J. Kuo and F. E. Zulvia, "Multi-objective cluster analysis using a gradient evolution algorithm," *Soft Computing*, vol. 24, no. 15, pp. 11 545–11 559, 2020. DOI: 10.1007/s00500-019-04620-0.

[27] B. C. Arnold, "Pareto distribution," *Wiley StatsRef: Statistics Reference Online*, pp. 1–10, 2014. DOI: 10.1002/9781118445112.stat01100.

[28] R. Sanders, "The Pareto principle: its use and abuse," *Journal of Services Marketing*, vol. 1, no. 2, pp. 37–40, 1987. DOI: 10.1108/eb024706.

[29] E. Lomurno, S. Samele, M. Matteucci, and D. Ardagna, "Pareto-optimal progressive neural architecture search," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2021, pp. 1726–1734. DOI: 10.1145/3449726.3463146.

[30] T. Elsken, J. H. Metzen, and F. Hutter, "Neural Architecture Search: A Survey," *Journal of Machine Learning Research*, 2019.

[31] M. Davtalab-Olyaie and M. Asgharian, "On Pareto-optimality in the cross-efficiency evaluation," *European Journal of Operational Research*, vol. 288, no. 1, pp. 247–257, 2021. DOI: 10.1016/j.ejor.2020.05.040.

[32] S. Vargas and P. Castells, "Rank and relevance in novelty and diversity metrics for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 109–116.

[33] D. Schönle, C. Reich, and D. O. Abdeslam, "Linguistic driven feature selection for text classification as stop word replacement," *Journal of Advances in Information Technology*, vol. 14, no. 4, pp. 796–802, 2023. DOI: 10.12720/jait.14.4.796-802.

[34] S. P. Bayerl *et al.*, "Offline model guard: Secure and private ML on mobile devices," in *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020, pp. 460–465. DOI: 10.23919/date48585.2020.9116560.

[35] M. Feurer and F. Hutter, "Hyperparameter optimization," *Automated machine learning: Methods, systems, challenges*, pp. 3–33, 2019. DOI: 10.1007/978-3-030-05318-5_1.

[36] S. González-Carvajal and E. C. Garrido-Merchán, "Comparing BERT against traditional machine learning text classification," *arXiv preprint arXiv:2005.13012*, 2020.

[37] C. Zhang, M. Yu, W. Wang, and F. Yan, "5MArk6: Exploiting Cloud Services for 5Cost-Effective6,5SLO-Aware6 Machine Learning Inference Serving," in *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, 2019, pp. 1049–1062.

[38] R. W. Numrich, L. Hochstein, and V. R. Basili, "A metric space for productivity measurement in software development," in *Proceedings of the second international workshop on Software engineering for high performance computing system applications*, 2005, pp. 13–16. DOI: 10.1145/1145319.1145324.

[39] M. Capra, B. Bussolino, A. Marchisio, G. Masera, M. Martina, and M. Shafique, "Hardware and software optimizations for accelerating deep neural networks: Survey of current trends, challenges, and the road ahead," *IEEE Access*, vol. 8, pp. 225 134–225 180, 2020. DOI: 10.1109/access.2020.3039858.

[40] C. Kiefer, "Quality Indicators for Text Data," 2019. DOI: 10.18420/btw2019-ws-15.

[41] E. M. Dharma, F. L. Gaol, H. Warnars, and B. Soewito, "The accuracy comparison among word2vec, glove, and fasttext towards convolution neural network (cnn) text classification," *J Theor Appl Inf Technol*, vol. 100, no. 2, p. 31, 2022.

[42] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering: new collection and results," in *Proceedings of the 11th ACM symposium on Document engineering*, 2011, pp. 259–262. DOI: 10.1145/2034691.2034742.

[43] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA: Association for Computational Linguistics, 2011, pp. 142–150.

[44] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[45] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[46] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A Matlab-like Environment for Machine Learning," in *BigLearn, NIPS Workshop*, 2011.

[47] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman, Eds., 2010, pp. 56–61.

Reducing Cybersickness from Virtual Reality: A Comprehensive Analysis

Monique Hanslo
University of South Africa
Pretoria, South Africa
nickyhanslo@gmail.com

Ridewaan Hanslo
Department of Informatics
University of Pretoria
Gauteng, South Africa
ridewaan.hanslo@up.ac.za

***Abstract*—Virtual reality, or VR, has several applications in teaching, entertainment, and business. When utilizing virtual reality, one may get "cybersickness," a simulated sickness. The usefulness of VR devices is severely hampered by Cybersickness (CS). Reducing the unpleasant feeling of CS is crucial to making the most of VR as a medium. A satisfying virtual reality experience results from a combination of technology, software, and user characteristics. There is a lack of comprehensive knowledge about the causes of Cybersickness, the methods for evaluating the degree of Cybersickness, and the variables that influence CS in a virtual reality setting. This research attempts to fill the gap by examining the causes of Cybersickness, how to evaluate it, and identifying and characterizing its contributing variables. A thorough analysis of the literature revealed 21 variables that influence VR CS. Furthermore, a taxonomy of contributing elements to Cybersickness was created so that academics and VR developers could assess them.**

Keywords-virtual reality; simulation sickness; Cybersickness; factors; head-mounted display; comprehensive analysis.

I. Introduction

Due in part to the recent media attention VR has received, it has lately entered the common vernacular [1][2]. In short, virtual reality (VR) provides a "Virtual Environment" (VE) where users interact with a highly lifelike artificial environment composed mainly of three-dimensional computer-generated pictures, sounds, and haptic feedback. VR has been used in several industries, including architecture, construction, and healthcare. Nevertheless, customers are more interested in video games than other VR applications [3]. Head Mounted Displays (HMDs) are the leading technology for virtual environments. Unlike traditional displays, HMDs immerse the user in the virtual environment (VE) by blocking external visual inputs that may disrupt the experience. Such immersive experiences have also been linked to a detrimental side effect called Cybersickness (CS) [3].

Being in a virtual environment (VE) can cause an unpleasant set of symptoms known as "cybersickness," which can last for a few hours or even days [4]. Headache, nausea, and even vomiting are some symptoms [5]. Between 20% and 80% of people are thought to be impacted by CS in some capacity [4]. Even though the condition has long been recognized and studied, CS claims have increased with VR devices' growing popularity [3].

In the worst cases, patients cannot use VR equipment because of the severity of their symptoms. In one case, players complained of feeling sick; thus, developers were obliged to take VR elements out of their games [3]. The symptoms of CS might negatively impact the patient and impede the effectiveness of medical therapy. The user may find even modest symptoms uncomfortable and bothersome.

A. Problem Statement

VR allows users to envision a redesigned three-dimensional environment. That being said, complete sensory awareness is required for maximum effectiveness. Before VR is sufficiently adapted and understood, interactions and visual cues in the VE must be established effectively to be as realistic as feasible [5]. The degree to which the user feels fully involved in the environment may be used to gauge its efficacy [6]. Issues with usability, like CS symptoms, will lessen a user's sense of presence in a virtual environment. Users will not be able to feel the realism of a VE if they encounter difficulties utilizing the environment.

A thorough picture of all the aspects that might lead to CS is necessary for the knowledge base. This study first attempts to uncover the elements leading to CS during VR technology use because of this research gap. Subsequently, a conceptual model that integrates these discovered elements will be developed.

B. Research Objectives

The following are the objectives of this study:

1) To determine the causes of Cybersickness in the Virtual Reality environment.

2) To determine the severity of Cybersickness experienced, or susceptibility to it, before, during, or following a Virtual Reality session?

3) To determine the factors that contribute to Virtual Reality Cybersickness

C. Research Questions

This research will address the following research questions:

1) What are the causes of Cybersickness in the Virtual Reality environment?

2) How can the severity of Cybersickness experienced, or susceptibility to it, be assessed before, during, or following a VR session?

3) Which factors contribute to Cybersickness during the application of Virtual Reality technologies?

D. Significance of the Study

This research provides an overview of all the proposed variables contributing to CS in a virtual reality setting. It aims to offer practical advice to upcoming and established VR developers on how to lessen CS symptoms to enhance the VR experience. It will also serve as a resource for scholars investigating CS in VR. This will act as a guide for growing the research and connecting it to the factors used. Furthermore, by improving the user experience, it is anticipated that this research would benefit users who encounter CS in VR environments.

This is how the remainder of the paper is structured. The literature on the causes, theories, and measuring techniques of Cybersickness is presented in Section II. The research technique for the Systematic Literature Review (SLR) is presented in Section III. Section IV presents the SLR results, and Section V offers a commentary on the research findings. The work is concluded in Section VI, which also provides suggestions for more research.

II. LITERATURE REVIEW

A. Cybersickness Causes and Theories

1) Sensory Conflict Theory

CS is a disorder that is difficult to categorize since there are a variety of symptoms, and the illness's effects differ from person to person. Many theories about how it began [4][7]. The most discussed hypothesis in literature is the Sensory Conflict Theory (SCT). It contends that CS results from a conflict between the information provided by several senses. It has been demonstrated that common motion sickness signs and physiological modifications, such as car or seasickness, are relatively similar to CS [8]. Additionally, sensory conflict seems to have an impact on both. However, the sensory conflicts in a car and VR are very different.

When traveling by car, one might perceive acceleration, but their visual surroundings, the vehicle's interior, remain still. This causes motion sickness. According to the SCT, you can lessen the conflict by gazing out the window, bringing the vestibular and visual information back into alignment. In VR, the conflict is going in the opposite direction. While the vestibular sense detects no motion or is out of sync with the visuals, VR users perceive motion and accelerations through visual cues. This affects how CS is treated differently from traditional motion sickness.

2) Vection

Vection, which refers to the perception of motion through visual stimuli, has frequently been linked to Visually Induced Motion Sickness (VIMS) or CS [9][10]. However, other research shows vection can happen even when no sickness is present [9]. This shows there is more to the relationship between vection and CS than just a straightforward causal one. In their study, [10], which intended to explore this connection further, discovered that a shift in vection causes sickness. From the standpoint of sensory conflict, it makes sense that CS is more often caused by apparent visual acceleration than continuous visual motion. Conflict happens when one reason detects acceleration while the other does not since the vestibular system can only detect accelerations.

However, the findings of [11] are at odds with those of the study by [10]. The vection's strength or fluctuation did not significantly impact VIMS. It is posited that [11] may have yet to successfully create a high level of motion sickness, which might account for these conflicting results. Therefore, any potential difference in the ability to generate motion sickness between constant and variable vection may have yet to be able to achieve statistical significance. Humans acquire information about body motion through their vestibular system, which detects the rotational and translational accelerations of the head, in addition to visual data. Therefore, combined with the visual system, the vestibular system is a crucial tool for humans to notice when our body is moving and distinguish between object and self-motion [12][13]. When you start moving in VR with a joystick, something other than this multisensory integration may work better. There is a sensory conflict since you can feel vection. Still, the vestibular system doesn't send any signals of self-motion.

SCT is explained from a different angle by [14] as an issue of dynamic sensory reweighting. They contend that visual input typically has a more significant weight than vestibular input since multisensory integration favors the most reliable signals; the weight will move more to the visual side when you engage in VR more frequently. CS symptoms may then be reduced with repeated VR exposure [15]. At first, the vestibular system's weighting is higher than the more prominent visual cues, causing significant sensory conflict. However, this sensory weighting shifts with time, and the vestibular system is disregarded, reducing conflict. The Peripheral Visual Field (PVF) controls sensing motion, including vection. In contrast, the Central Visual Field (CVF) is primarily responsible for detecting and identifying objects with the highest density of cones. According to [16], motion in the perceived background causes more vection than motion in the front.

The foreground is the emphasis of the CVF. However, [7] noted that peripheral inputs frequently have a background interpretation. Therefore, the impression of self-motion is more likely to result from motion in the PVF. When placed in the CVF, a motion that traveled laterally increased illness but not vection, whereas a motion that traveled longitudinally (forwards or backward) increased sickness and vection when placed in the PVF.

3) Postural Instability

Postural instability, a notion that [17] first proposed, is another frequently discussed theory. They suggested that symptoms happen when you have not learned how to maintain yourself in that particular situation and are experiencing postural instability. When riding a roller coaster in VR while standing, you might be familiar with this sensation of instability. Various studies appear to contradict one another, with some offering evidence for the theory [18][19]. In contrast, others only discovered postural instability due to CS or found no causal relationship [20]. It still needs to be determined what the exact relationship with CS is. However, this idea offers a foundation for measuring CS objectively.

4) Rest-Frame Hypothesis

The Rest-Frame Hypothesis is another theory that has influenced a typical CS mitigation technique [21]. According to this theory, CS results from the inability to identify or select a stable reference frame, also known as the rest frame, to interpret relative movements, locations, and orientations. The nervous system chooses the rest of the frame from various reference frames and gives it spatial-perceptual data [21]. According to the theory, the cognitive conflict that results from being unable to identify a single rest frame compatible with a person's inertial and visual motion signals, rather than the sensory conflict, causes CS [16]. In other words, illness is more likely affected by how the user interprets what is moving and what is not based on the degree of competing cues.

Choosing a reference frame is typically an unconscious procedure for most people [21]. However, suppose you have been on a train waiting at a stop adjacent to another train. In that case, you may be familiar with the experience of not having a comfortable resting space. There may be some uncertainty over which train is genuinely moving for a brief period after the other train begins to move. This continues until you see a resting place other than the trains, such as the earth beneath them or other structures. Your mind can thus rationally conclude that the other train, not yours, was moving.

B. Cybersickness Measurement Methods

As covered in the section above, there are only a few well-established theories on CS. Similarly, subjective and objective approaches to assessing CS exist, categorized into physiological state, postural sway, and questionnaires.

1) Questionnaires

The Simulator Sickness Questionnaire (SSQ) is used in most articles. Even though this questionnaire was first developed for military simulators (like flight simulators), it is still the most well-known for CS in VR research. From none to severe, participants assess the severity of 16 symptoms on a 4-point scale. The results are divided into four scores: overall score, nausea, oculomotor, and disorientation. Several researchers have suggested alternatives because the SSQ's primary intent was not VR and was evaluated on highly skilled professionals [22][23]. Both the Cyber Sickness Questionnaire (CSQ) and the Virtual Reality Sickness Questionnaire (VRSQ) published by [22] and [24] can be seen as subgroups of the SSQ. Only nine symptoms remain when the nausea-related symptoms are excluded from the VRSQ.

According to [24], the oculomotor and disorientation components of illness in VR are more critical than the nausea component. They contend that the difference in nausea-inducing effects between VR and simulators is due to the absence of inertial motion. The SSQ, the French version of the SSQ, the VRSQ, and the CSQ were all subjected to a psychometric study [23]. Compared to the SSQ and its French equivalent, they discovered that the VRSQ and CSQ demonstrated more validity. It is conceivable to ask participants to complete an SSQ but then analyze the data using a VRSQ or CSQ because those questions are subsets of the SSQ. A significant drawback is that you may only use the surveys mentioned above before or after a VR experience, owing to their size. As a result, real-time data cannot be obtained using the SSQ, VRSQ, or CSQ.

The Fast Motion Sickness Measure (FMS) is a one-dimensional scale that ranges from zero to 20. This scale, which indicates no motion sickness (zero) to severe motion sickness (20), was developed by [25]. It is feasible to gauge the time of the motion sickness since participants vocally rate each minute. The FMS, SSQ, and sub-scores also show a substantial correlation in other research [25][26]. The Misery Scale (MISC) was developed by Wertheim et al. as an alternative to the FMS. The scale extends from zero (no symptoms) to ten (vomiting). In addition to verbal responses, a physical dial may also be used to record answers on a one-dimensional illness scale, as a [27] study showed.

It might be essential to know a participant's vulnerability to motion sickness in addition to measuring CS during or after a VR session. Participants' susceptibilities to CS can vary. Thus [28] updated the Motion Sickness Susceptibility Questionnaire (MSSQ) developed to gauge this. The participant's history of motion sickness is examined using the MSSQ. The Visually Induced Motion Sickness Susceptibility Questionnaire (VIMSSQ), which looks at prior encounters with symptoms rather than motion sickness in general, was created by [25] since this questionnaire was not designed for CS (or VIMS). [29] Also, due to its length, a condensed version of the VIMSSQ was suggested and examined in another research study.

2) Physiological State

Although questionnaires are the most popular way to detect CS, they have certain drawbacks. First, surveys interfere with the user's experience, making it impossible to track their illness in real time [30]. The fact that surveys are inherently subjective is another disadvantage. As a result, they only sometimes accurately gauge what they are attempting to perform. Researchers can assess the physiological status of the consumers to get past these issues. This is doable in real time and may offer a source of unbiased data.

The majority of the literature uses many physiological signals, not just one. [4] recommends using an electrocardiogram (ECG) and blood pressure in their review. Still, [31] recommends the most accurate assessment technique for galvanic skin response. Other potential techniques include eye tracking, heart rate, breathing, and cutaneous thermoregulatory vascular tone [2][30]. [8] discovered that autonomic arousal was primarily responsible for variations in heart rate and breathing. Measuring the physiological status makes it possible to create a closed-loop system, which is a significant advantage.

The user's present status might be assessed by sensors, which would subsequently apply the appropriate CS mitigation techniques. A method that can evaluate CS in real-time using physiological data was developed by many researchers using machine learning [20][32]. Based on physiological data, such as heart rate, breath rate, heart rate variability, and galvanic skin reaction, [32] created an entirely closed-loop system. Based on the determined amount of sickness, the field of view (FOV) reduction or Gaussian blurring was applied, which might lower the level of nausea. The degree of CS was determined by periodically evaluating the user's physiological data. The system's capacity to lessen CS was not put to the test.

Despite being objective, physiological evidence has not been able to displace the SSQ as the gold standard for assessing CS. Physiological outcomes have often been employed in research to support their conclusions rather than as the primary measurement technique. Additionally, the SSQ or other questionnaires frequently validate physiological measures. Therefore, their validity is dependent on arbitrary information.

3) Posturial Sway

Postural sway, a type of body movement, has yet to be included in several investigations as an impartial evaluation technique, even if the relationship between postural instability and CS still needs to be fully understood [33]. [34] showed that gait metrics may also be measured to determine CS. They recorded the necessary data using an inertial measurement unit on each foot. They then used a support vector machine, a machine learning model, to create a classifier for CS.

Using a balancing board to measure movements around the center of gravity is one method of documenting postural instability [18][35][36]. After analyzing their data, [35] identified the precise postural sway characteristics that might predict VIMS. According to the findings, those who reported feeling worse had more circular postures (as opposed to elliptical) and a higher frequency of forward/backward oscillations. According to each participant's postural sway, [36] trained a deep, short-term memory model that may forecast their likelihood of experiencing CS.

However, there are also sensors in users' Head Mounted Displays (HMD) that may capture postural sway. Head dispersion, or the change in roll and pitch, was tested by [37] and shown to be significantly connected to changes on the x- and y-axis around the center of gravity. Participants had to hold their heads motionless or stare straight ahead to assess head dispersion. The relationship between the location information from the HMD and CS was also examined by [38]. They found strong correlations between a few location factors and the SSQ scores, even though the data was pretty noisy. These findings imply that it may be feasible to design a system that collects the HMD's location data, calculates the user's level of CS in real time, and utilizes that information to modify the methods for reducing sickness.

III. METHODOLOGY

A. Introduction

This research uses an SLR, defined as "*a means of identifying, evaluating and interpreting all available research to a particular research question, or topic area, or phenomenon of interest*" [39]. Simply put, an SLR is a review of primary studies. This study follows the SLR guidelines by [39]: identifying sources, study selection, data extraction, data synthesis, and writing up the study as a report.

B. Search Terms used in selected databases

"Virtual Reality" AND ("cybersickness" OR "motion sickness" OR "simulator sickness") AND ("factors" OR "fail" OR "break down" OR "flounder" OR "blunder" OR "flop" OR "deteriorate" OR "challenge" OR "issue" OR "problem" OR "obstacle*" OR "success" OR "accomplish" OR "achieve" OR "advance" OR "progress*" OR "realisation" OR "triumph" OR "victory" OR "fruition" OR "attainment" OR "model" OR "method" OR "framework").

1) Source Selection

The following data sources were selected to perform the search:

- IEEE Xplore Digital Library
- Scopus
- ACM Digital Library
- Google Scholar

All of these databases are well-known research repositories in information technology. In addition, Google Scholar was employed to help locate sources via backward and forward citation searches.

2) Selection Criteria

The selection of research material for inclusion in this systematic review was based on this section's inclusion and exclusion criteria.

For a source to be included in the research, it had to meet the following criteria:

- Papers describe the factors that lead to Cybersickness in a VR setting.
- Papers containing at least three keywords in the title, abstract, or keywords were chosen.
- Journal articles, conference papers, book chapters, dissertations, and theses were considered.
- No limitations on the publication date.

A source is excluded from the research for the following reasons:

- Papers that don't discuss the factors that contribute to Cybersickness in a virtual reality setting.
- Non-English language academic papers.
- If the full text of the publication is not available.
- Duplicate papers meaning the same paper retrieved from different databases.

3) Process for conducting the review

The search string above was performed on the selected databases, returning 1231 articles. The Google Scholar citation search found an additional ten records. After that, 219 duplicate papers were removed. Screening by the title

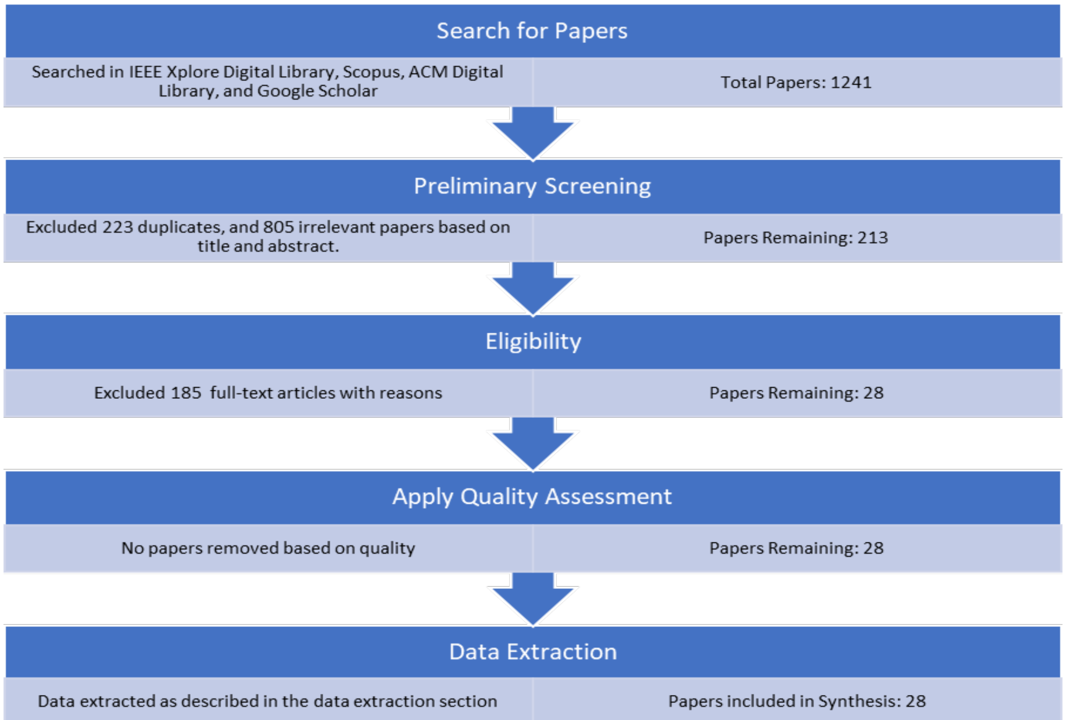

Figure 1. Summarized process for conducting the systematic review.

and abstract was conducted, leaving 213 full-text articles. These full-text articles were further assessed for eligibility, resulting in 28 remaining articles used for data extraction and synthesis (see Figure 1). The search was completed in August 2022.

4) Quality Assessment

The included papers were assessed using four quality assessment questions. The questions aimed to evaluate the quality aspects mentioned by [39]. These aspects are characterized as **objectivity** - if the research is free of bias; **reliability** - the accuracy and reliability of the research instruments used; **internal validity** - whether the research was well structured, so data was collected from suitable sources; and **external validity** - determines if the findings can be predicted for subsequent occasions.

Therefore, the following questions were devised to assess the quality of the selected literature:

Q1. Is Virtual Reality and Cybersickness factors the center of the discussion?

Q2. Does the research have a clear goal in mind?

Q3. Does the article follow a research process and describe the data analysis techniques?

Q4. Does the article report its findings based on evidence and argument?

These questions had three possible answers: Yes and No. Each response is given the following weighting: Yes = 1 and No = 0. The final score was noted and utilized as a scale from 0 to 4 to represent the overall quality of the chosen literature. The articles' outcomes and quality ratings are displayed in the results section.

C. Data Extraction

The data extraction was carried out on 28 papers included in the SLR. After that, a qualitative thematic analysis was conducted to synthesize the extracted data. Some of the article's content was highlighted in the paper while it was being read. These ideas/concepts, usually called codes, were carefully investigated to group them into common themes. All the pertinent information that helped answer the research question was extracted, including the citation, the journal article or conference title, the source database, year published and study type, article sub-concepts, and the central concept. Google Sheets were used to extract data for the thematic analysis.

IV. RESULTS

A. Search Results

The articles listed in the source selection section were looked at in four databases, which include Google Scholar. Figure 2 displays the percentage distribution. Most of the articles came from IEEE Explore (41.8%). Scopus accounted for 40.3% and ACM digital library 12.9%. 5% derived from the Google Scholar citation searches.

Most papers included many factors, while some focused on one specific factor. Table I lists these 21 factors and their sources.

B. Quality Evaluation of Articles

As mentioned earlier, four questions were used to assess the quality of the selected literature. Most papers were of good quality, with an average score of 3.75 out of 4. No paper scored below 3 (see Table II).

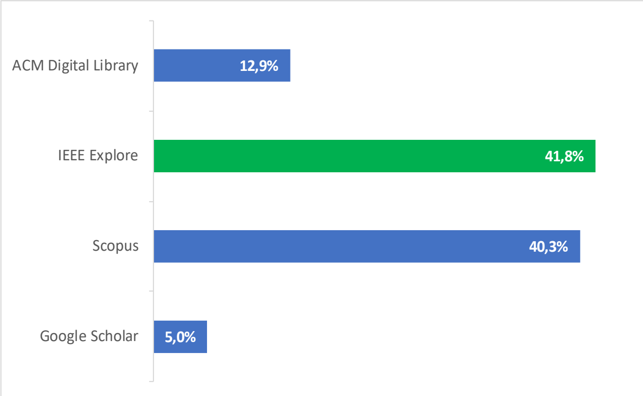

Figure 2. Database articles percentage distribution.

C. *Synthesis of Identified Factors*

A thematic analysis was conducted to identify the core themes and subthemes within the selected literature. The factors were categorized under subthemes and grouped under a theme. Initially, 42 factors contributed to CS in a VR environment. Upon examination of the definitions of each of these factors and the references made to them by the authors of the selected literature, 21 factors were merged into others, resulting in 21 final factors. The remaining 21 factors were further analyzed to identify any additional relationships to help categorize them. Categorizing the factors helps to understand the more significant themes and gives more profound insight. The synthesis using a thematic analysis went through 5 iterations, resulting in three themes, eight subthemes, and 21 factors. These three common themes were identified as User, Hardware, and Software. Table III lists the synthesized themes, subthemes, and the contributing CS factors. A taxonomy of the contributing factors to Cybersickness is shown in Figure 3.

V. DISCUSSION

This section of the research aims to answer the three research questions. The core SLR themes identified are 1) User, 2) Hardware, and 3) Software. Each of these themes has sub-themes that translate into factors. The factors under each theme and subtheme are discussed next, followed by an address to the research questions.

A. *Factors Contributing to CS*

1) User: There are differences in CS susceptibility at the user level. These factors include Age, Gender, Habituation, Duration, Environmental Conditions, Physical Health and Posture. Each of these factors is discussed below. These factors are grouped into Demographics, Experience, and Physical Attributes.

TABLE I. VIRTUAL REALITY CYBERSICKNESS FACTORS WITH SOURCES

Factor	Sources
Habituation	[40]-[43]
Duration	[5][33][40]-[42][44][45]
Environmental conditions	[46]
Physical Health	[5][41][47]-[49]
Posture	[5][45][50]
Gender	[5][33][40][42][46][47][51]-[53]
Age	[5][33][40][47][53][54]
Field of View	[33][40][41][44][55]
Flicker	[5][33][53][56]
Screen size	[40][56]
Head-mounted displays	[2][6][31][33][42][46][56]-[58]
Lag and Frame Rate	[44][59]
Method of movement	[56][60]
Calibration	[5]
Position Tracking error	[53]
Head motion	[45]
Playing position	[40]
Locomotion	[33][40][44][61]
Immersion	[33][40][62]
Sensory support	[62]
Graphic Realism	[33]

a) *Demographics:* The Demographics subtheme consists of factors of Age and Gender.

Age. According to the literature, younger persons are more resistant to simulation sickness [33]. After age 40, people's vestibular perception threshold, or the lowest signal recognized, decreases, rendering them more susceptible to simulation sickness [47]. [54] discovered changes in the postural balance between young and middle-aged test participants. Furthermore, postural balance deteriorates as people age, which can contribute to illness.

Gender. Females have consistently been found to be more susceptible than males to CS. With the usage of HMDs, CS may differ depending on gender. [52] investigated the influence of gender and technology and their possible contributions to simulation sickness. Using data from 223 people (108 men and 115 women), they investigated the degrees of simulation sickness concerning gender, sensory conflict, and advancements in VR technology. They concluded that women had a greater level of simulation sickness than males.

TABLE II. QUALITY EVALUATION ANSWERS

No	Citation	Q1	Q2	Q3	Q4	Score
1	[56]	Yes	Yes	Yes	Yes	4
2	[59]	Yes	Yes	Yes	Yes	4
3	[44]	Yes	Yes	Yes	Yes	4
4	[33]	Yes	Yes	Yes	Yes	4
5	[62]	Yes	No	Yes	Yes	3
6	[6]	Yes	Yes	Yes	Yes	4
7	[58]	Yes	Yes	Yes	Yes	4
8	[57]	Yes	Yes	Yes	Yes	4
9	[42]	Yes	Yes	Yes	Yes	4
10	[46]	Yes	Yes	Yes	Yes	4
11	[52]	Yes	Yes	Yes	Yes	4
12	[54]	Yes	No	Yes	Yes	3
13	[60]	Yes	Yes	Yes	Yes	4
14	[53]	Yes	Yes	Yes	Yes	4
15	[2]	Yes	Yes	Yes	Yes	4
16	[40]	Yes	No	Yes	Yes	3
17	[31]	Yes	Yes	Yes	Yes	4
18	[61]	Yes	Yes	Yes	Yes	4
19	[51]	Yes	Yes	Yes	Yes	4
20	[43]	Yes	Yes	Yes	Yes	4
21	[5]	Yes	No	Yes	Yes	3
22	[49]	Yes	No	Yes	Yes	3
23	[50]	Yes	Yes	Yes	Yes	4
24	[41]	Yes	Yes	Yes	Yes	4
25	[55]	Yes	Yes	Yes	Yes	4
26	[47]	Yes	No	Yes	Yes	3
27	[45]	Yes	No	Yes	Yes	3
28	[48]	Yes	Yes	Yes	Yes	4

[46] conducted many trials. They discovered that females were equally susceptible to motion sickness caused by an improper fit of the VR headgear to the inter-pupillary distance (the distance between the center of one's eyes). They also propose redesigned VR headsets with adjustable interpupillary distance to decrease CS in women.

b) *Experience:* The Experience subtheme consists of factors such as Habituation, Environmental Conditions, and Duration.

Habituation. According to [42], an increase in exposure time was directly related to the degree of unpleasant symptoms. Compared to non-susceptible individuals, those prone to motion sickness might suffer nearly double the severity. Users who feel nausea when riding carnival rides might expect to endure unpleasant sensations. Exposing a person to virtual surroundings briefly, halting the encounter before or during illness, and retrying in a day or two will assist the user in acclimatizing to the virtual world. Exposure to virtual settings regularly may reduce or eliminate simulation sickness.

Environmental Conditions. CS symptoms worsen in environments with high temperatures and inadequate ventilation. Good airflow and ventilation can help reduce nausea and aid recovery after dizziness [46].

Duration. Several studies have found that more than 10 minutes of VR exposure can cause nausea, and the longer the exposure period, the more severe the VR sickness [33][40][41][44]. According to these studies, the application should allow users to pause the experience for a rest and then resume it later. In contrast, an application might advise users to take breaks regularly to avoid unpleasant sensations [44].

c) *Physical Attributes:* The Physical Attributes subtheme consists of Physical Health and Posture factors.

TABLE III. SYNTHESIZED THEMES, SUBTHEMES, AND FACTORS

Themes	Subthemes	Factors
User	Experience	Habituation Duration Environmental Conditions
	Physical attributes	Physical Health Posture
	Demographics	Gender Age
Hardware	Device	Field of View Screen Size Flicker Head Mounted Displays Lag and Frame Rate
	Tracking	Method of Movement Calibration Position Tracking Error Head Motion
Software	Stabilizing information	Playing Position
	Environment	Locomotion
	Design	Immersion Sensory Support Graphic Realism

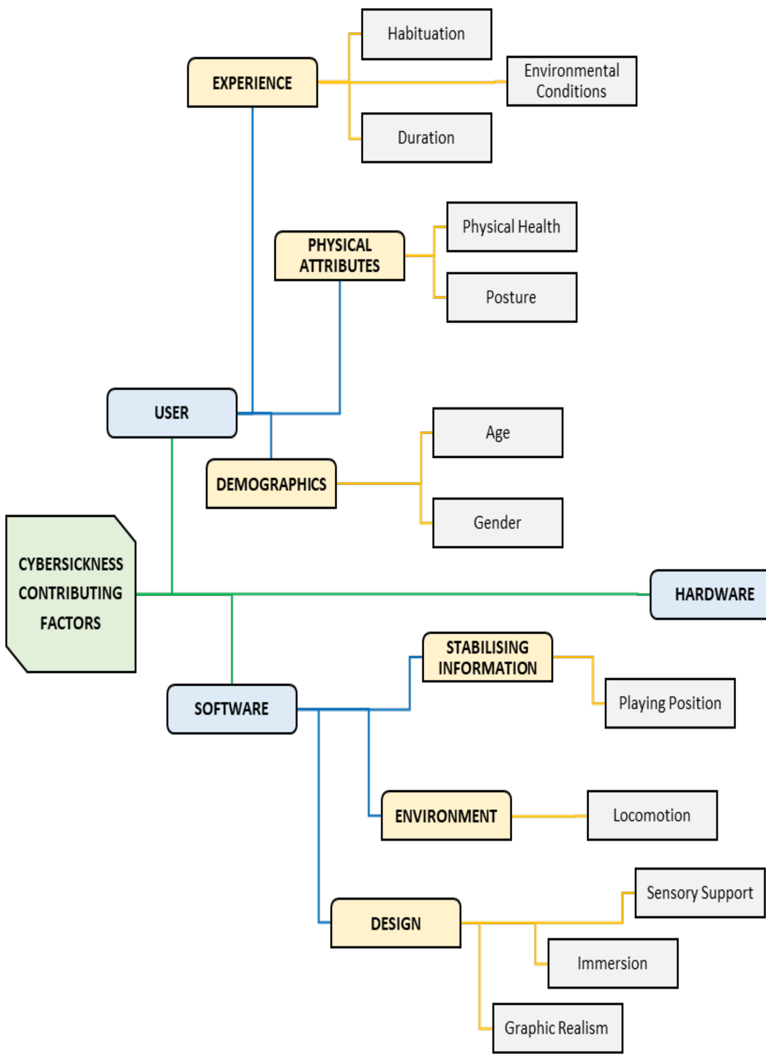

Figure 3. Taxonomy of Cybersickness contributing factors.

Physical Health. The user's senses must be at their peak to attain a heightened presence level. For the optimum VR experience, users should be physically fit and have a strong sense of balance. If a user has a hangover, cold, headache, tired, or is sleep deprived, it is best to avoid a virtual environment since their symptoms may aggravate [44].

Posture. Postural instability is a well-documented consequence of exposure to a Virtual Environment (VE). Postural stability is frequently assessed before and after VE exposure to detect changes in stability caused by the exposure. Less posturally stable individuals are more likely to get CS or suffer from more severe illness when compared to more posturally stable individuals [5][45][50].

2) Hardware

Some factors associated with hardware used in a VE can induce CS. These include HMDs, Flicker, Field of View (FOV), Lag and Frame Rate, Screen Size, Method of Movement, Calibration, Position Tracking Error, and Head Motion. These factors are grouped in the subthemes of Device and Tracking.

a) *Device:* The Device subtheme includes factors such as HMDs, Flicker, FOV, Lag and Frame Rate, and Screen Size.

Head-Mounted Displays (HMDs). When using HMDs, contrast, light, exposure length, and operating distance contribute to straining the visual system. When utilizing a stereoscopic HMD, such as EyePhone LX, in an immersive virtual world for 10 minutes, around 60% of respondents exhibited symptoms such as eye strain, nausea, and headache, while 20% reported a loss in binocular visual perception [58]. Similar symptoms were reported by 61% of participants following twenty minutes of exposure to immersive virtual material using a DVisor HMD [31].

Technical developments in VR display technology, such as Oculus VR DK1 and Oculus VR DK2, did not significantly reduce CS [52]. Sensory conflict, however, plays a vital role in developing nausea and other symptoms. Body movement, confusion caused by head movement, and poor optical design led to strain-induced ocular pain [1]. Recently, it was observed that using HMDs caused more motion sickness than stereoscopic desktop displays. Some users stated that they felt more immersed in an HMD. However, they could only sustain the experience for a short period.

Flicker. Flicker has been extensively researched. The literature [5][33][53][56] suggests that flicker should be avoided at all costs. In a VR scenario, flicker is the brightness fluctuation on video screens that can cause nausea. This oscillation is visually disturbing and affects the user's eye health. The user will likely see flicker around the screen's edges when using larger displays. Avoiding flicker is crucial for HMDs with brighter panels and a high refresh rate [33]. Several components of the visual presentation influence flicker perception. The most relevant to visual displays or VR systems are the refresh rate, brightness level, and field of vision [5]. To reduce flicker, the refresh rate must increase as the brightness level increases [53].

Field of View (FOV). The display's horizontal and vertical angular dimensions are known as the FOV [55]. CS is more common in VE situations with a wide FOV than those with a narrow FOV [40]. This is likely due to enhanced vection caused by higher peripheral retina stimulation from a broad FOV display [41]. A wide FOV also enhances the probability of detecting flicker [44]. This is because the peripheral visual system is more sensitive to flicker. To eliminate flicker, a broader FOV requires a quicker refresh rate [44].

Lag and Frame Rate. Latency is the time between the user's input and the visible response in a VE display. Frame rate measures how rapidly frames flow through the rendering process. A dip in frame rate might occur in a VR application with sophisticated visuals. Suppose the delay between user input and virtual content production is significant. In that case, there is a considerable risk of developing simulation sickness [44]. A suggested delay is 20 milliseconds; anything more substantial than 46 milliseconds might cause motion nausea. Companies such as Oculus, Sony, and Steam stress the significance of virtual content with low latency, responsiveness, and fast frame rates for greater virtual content quality [59].

Screen size. Vection is highest in peripherally moving visual flow fields [40]. As a result, huge displays pose an increased risk of motion sickness. With full-flow fields, virtually everyone will feel intense vection. Generally, the smaller the visual picture (or display), the lower the likelihood of CS [56]. Laboratory investigations have shown that the danger of vection is limited, with pictures reaching a viewing angle of fewer than 300 degrees [40]. A typical 17-inch computer screen, seen from a distance of 50 cm, contains 340 pixels and will not readily cause vection [40].

b) *Tracking:* The Tracking subtheme comprises the factor's Method of Movement, Calibration, Position Tracking Error, and Head Motion. These are discussed below.

Method of Movement. The VR user does not always have control over the character's motions. This lack of mobility can lead to significant problems. To satisfy sensory expectations, movement in a virtual world should be realistic. Inappropriate motions, such as quick tilting, rolling, and waveform, should be avoided. Gun sway, head bob, and moving up and down stairs are incorrect movements. According to [44], incorporating motions centered on leaps rather than continuous walks may help to reduce nausea. Uncontrolled user movement outputs should be restricted, such as flipping, falling, or zoom transitions [60].

Calibration. Because of variances in human physical traits, poor calibration exacerbates CS symptoms. Interpupillary distance, for example, the distance between the pupils' centers in both eyes, differs among persons [5]. Because stereoscopic displays require each eye to get a slightly offset image of the virtual world, this offset must be as near the user's interpupillary distance as feasible. Calibration failure might result in greater spatial and temporal distortions, setting the scene for CS due to distorted graphics [5]. As a result, each individual requires suitable calibration. [5] believes that the right size, appropriate focus, and perfect alignment will aid in treating CS.

Position Tracking Error. The VR system's position-tracking error informs the computer about the location of the user's head and, presumably, limbs in the VE [53]. The system uses this data to depict the user within the VE visually. If this information needs to be corrected, tracked items may appear in locations where they are not. If the tracked items are part of the user's body, the mismatch between where the graphical representation of the objects appears in the visual display and where the user believes they should appear may bother the user [53]. As a result, the illusion of the simulation may be broken, resulting in sickness-related symptoms, such as dizziness and loss of focus. Finally, location tracking mistakes might generate jitter or oscillations of portrayed body parts, disturbing users [53].

Head Motion. According to [45], adopting a supine posture results in a considerable reduction in CS. They ascribed this to limited head mobility. Head movements are known to be related to CS via Coriolis and pseudo-Coriolis stimulation pathways [45]. Coriolis stimulation occurs when the head is tilted away from the axis of rotation during actual body rotation [45]. When the head is inclined, apparent self-rotation is caused by visual cues, resulting in pseudo-Coriolis stimulation [45].

3) Software

The characteristics of the software in a VE may impact the probability of CS. The theme is divided into three subthemes: Stabilizing Information, Environment, and Design. Playing Position, Locomotion, Immersion, Sensory Support, and Graphic Realism are contributing factors.

a) *Stabilizing Information:* The stabilizing information subtheme consists of the Playing Position factor.

Playing Position. [45] revealed that a significant reduction in CS occurs when individuals assume a supine position, probably due to limited head mobility. In most circumstances, subjects are expected to be seated or standing within a VE [40]. Because of the lower demands on postural control, sitting patients would experience less illness, according to [40].

b) *Environment:* The Environment subtheme consists of the factor Locomotion.

Locomotion. A vital factor in VE discomfort is accelerated movement or speed. Sensory conflicts that cause discrepancies occur due to sudden increased or decreased acceleration. Therefore, increasing or decreasing acceleration slowly would result in a pleasant user experience [44]. Rapidly zoomed movements should also be avoided, such as when the visual cones move faster than expected when a user's view is zoomed in [61].

c) *Design:* The Design subtheme comprises Immersion, Sensory Support, and Graphic Realism.

Immersion. [6] studied the impact of virtual content type on simulation sickness. They noticed that the kind of video content, immersive vs. non-immersive, is critical for VE usability. Video content type influenced the contributor's sensitivity to simulation sickness and physiology. Their conclusion was based on the results of a Simulation Sickness Questionnaire (SSQ) and other physiological measures. The lowest SSQ score was recorded for non-immersive virtual content displayed on a television screen, while the highest scores were reported on an HMD with immersive content [6].

Sensory Support. A user might experience higher VR immersion and expect relevant vestibular information after exposure to strong illusions. The system can cause motion sickness if the VR system cannot provide suitable sensory input [62]. Therefore, designing a logical environment in which the players can focus and bind is essential. The user interface elements should be fixed rather than floating, creating an environment with a clear, steady horizon and reference points that users can focus on to minimize sickness. A world with imbalanced or changing backgrounds should be avoided. Designing a virtual world that supports human sensory systems is ideal [62].

Graphic Realism. [33] investigated the results of rendering realistic scenes. Participants who experienced realistic graphic content were prone to higher simulation sickness. The authors also suspect a sensory discrepancy between the vestibular and visual systems may cause more discomfort.

B. Answering the research questions

1) What are the causes of Cybersickness in the Virtual Reality environment?

A literature review was done in an attempt to understand the reasons why individuals become cybersick in a VR environment. The Sensory Conflict Hypothesis was the CS theory discovered to be the most often discussed in the literature. According to the hypothesis, illness results from an imbalance between two sensory systems, the vestibular and visual systems. Other research identifies postural instability or the absence of a rest frame, a fixed reference frame, contributing to CS [17]. However, experiencing motion sickness in VR can potentially lead to postural instability.

2) How can the severity of Cybersickness experienced, or susceptibility to it, be assessed before, during, or following a session?

A literature review was conducted to provide an answer to this question. According to the literature, several objective and subjective techniques can be used to gauge one's vulnerability to or degree of CS. Although the CSQ and VRSQ have shown superior validity for VR, according to the study of [23], the SSQ is still the most often used assessment technique. Examples of one-dimensional scales that let researchers quantify CS while participants are in VR are the FMS and MISC [25]. The MSSQ generally assesses prior experiences with motion sickness, whereas the VIMSSQ assesses susceptibility to CS [25].

In addition to surveys, the physiological condition reveals how much CS individuals feel. The advantage of physiological data collection is that it can be done throughout the VR experience and is a reliable source of factual information. Measuring the characteristics of gait or postural sway is another technique to obtain objective data. CS was shown to be connected with specific VR headset positional and rotational features [43].

3) Which factors contribute to Cybersickness during the application of Virtual Reality technologies?

A systematic review was conducted to answer this question. Systematic reviews deliver an orderly, clear means for gathering, synthesizing, and evaluating the results of studies on a specific topic or question [63]. The purpose of a systematic review is to minimize the bias linked with solitary studies and non-systematic reviews [63]. A thematic analysis was used to identify the core themes and factors within the selected literature.

Twenty-eight publications were included in the systematic review based on four carefully chosen databases. Twenty-one factors contributed to CS during the application of VR technologies. These factors are Age, Calibration, Duration, Environmental Conditions, Field of View, Flicker, Gender, Graphic Realism, Habituation, Head Motion, Head Mounted Displays, Immersion, Lag and Frame Rate, Locomotion, Method of Movement, Physical Health, Playing Position, Position Tracking Error, Posture, Screen Size, and Sensory Support. As a result, a conceptual model of the factors that lead to CS has been developed.

VI. CONCLUSION

This study's main objective was to find and characterize the variables that lead to Cybersickness (CS) in a virtual reality (VR) setting. Thematic analysis and an SLR were used to accomplish this. A model of the contributing elements to CS has been created to facilitate the investigation of CS in VR.

It became evident from doing this analysis that CS is a complex problem. There isn't a magic bullet answer available right now. Thankfully, a lot of options have previously been considered. Certain ones work better than others. By testing and learning more about the underlying mechanics of CS, we can move closer to a VR experience that may be devoid of it. Individual CS-inducing elements might be minimized if not completely removed.

The first limitation of this study is that it only looked at articles written in English. Thus, this analysis does not include information published in a language other than English that may be relevant to the research issue. Second, there's a chance that pertinent information from additional databases was overlooked because the SLR only used four data sources. Third, it is possible that the SLR's search criteria were not strict enough, leading to omitting important themes and variables.

With the established model, future scholars and practitioners may assess the conditions that give rise to CS in a virtual reality setting. The elements and topics of this study should be further investigated and supported or refuted by similar studies.

REFERENCES

[1] M. Hanslo and R. Hanslo, "Lowering the Effects of Virtual Reality Cybersickness: A Systematic Review". In the Fifteenth International Conference on Emerging Networks and Systems Intelligence. IARIA, pp. 5-14, 2023, ISBN: 978-1-68558-110-7

[2] G. Y. Menshikova, A. I. Kovalev, O. A. Klimova, and V. V. Barabanschikova, "The application of virtual reality technology to testing resistance to motion sickness". Psychology in Russia. State of the Art, 10(3), 2017.

[3] K. Berntsen, R. C. Palacios, and E. Herranz, "Virtual reality and its uses: a systematic literature review". In Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 435-439, 2016.

[4] L. Rebenitsch and C. Owen, "Review on cybersickness in applications and visual displays". Virtual Reality, 20, no. 2, pp. 101-125, 2016.

[5] S. Davis, K. Nesbitt, and E. Nalivaiko, "A systematic review of cybersickness". In Proceedings of the 2014 Conference on Interactive Entertainment, pp. 1-9, 2014.

[6] J. Guna et al., "Virtual reality sickness and challenges behind different technology and content settings". Mobile Networks and Applications, 25, no. 4, pp. 1436-1445, 2020.

[7] M. C. Howard and E. C. Van Zandt, "A meta-analysis of the virtual reality problem: Unequal effects of virtual reality sickness across individual differences". Virtual Reality, 25, no. 4, pp. 1221-1246, 2021.

[8] A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko, "Profiling subjective symptoms and autonomic changes associated with cybersickness". Autonomic Neuroscience, 203, pp. 41-50, 2017.

[9] R. H. So, W. T. Lo, and A. T. Ho, "Effects of navigation speed on motion sickness caused by an immersive virtual environment". Human Factors, 43 no. 3, pp. 452-461, 2001.

[10] F. Bonato, A. Bubka, S. Palmisano, D. Phillip, and G. Moreno, "Vection change exacerbates simulator sickness in virtual environments". Presence: Teleoperators and Virtual Environments, 17, no. 3, pp. 283-292, 2008.

[11] O. X. Kuiper, J. E. Bos, and C. Diels, "Vection does not necessitate visually induced motion sickness". Displays, 58, pp. 82-87, 2019.

[12] M. W. Greenlee et al., "Multisensory integration in self motion perception". Multisensory Research, 29, no. 6-7, pp. 525-556, 2016.

[13] A. M. Green and D. E. Angelaki, "Multisensory integration: resolving sensory ambiguities to build novel representations". Current Opinion in Neurobiology, 20, no. 3, pp. 353-360, 2010.

[14] M. Gallagher and E. R. Ferrè, "Cybersickness: a multisensory integration perspective". Multisensory Research, 31(7), pp. 645-674, 2018.

[15] N. Dużmańska, P. Strojny, and A. Strojny, "Can simulator sickness be avoided? A review on temporal aspects of simulator sickness". Frontiers in Psychology, 9, 2132, 2018.

[16] A. Delorme and C. Martin, "Roles of retinal periphery and depth periphery in linear vection and visual control of standing in humans". Canadian Journal of Psychology, 40, no.2, 176, 1986.

[17] G. E. Riccio and T. A. Stoffregen, "An ecological theory of motion sickness and postural instability". Ecological Psychology, 3, no. 3, pp. 195-240, 1991.

[18] D. Risi and S. Palmisano, "Effects of postural stability, active control, exposure duration and repeated exposures on HMD induced cybersickness". Displays, 60, pp. 9-17, 2019.

[19] S. Litleskare, "The relationship between postural stability and cybersickness: it's complicated–an experimental trial assessing practical implications of cybersickness etiology". Physiology & Behavior, 236, 113422, 2021.

[20] M. S. Dennison and M. D'Zmura, "Cybersickness without the wobble: Experimental results speak against postural instability theory". Applied Ergonomics, 58, pp. 215-223, 2017.

[21] E. Chang et al., "Effects of rest frames on cybersickness and oscillatory brain activity". In 2013 International winter workshop on brain-computer interface (BCI). IEEE, pp. 62-64, 2013.

[22] W. B. Stone III, Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of Cybersickness (Doctoral dissertation, Iowa State University), 2017.

[23] V. Sevinc and M. I. Berkman, "Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments". Applied Ergonomics, 82, 102958, 2020.

[24] H. K. Kim, J. Park, Y. Choi, and M. Choe, "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment". Applied Ergonomics, 69, pp. 66-73, 2018.

[25] B. Keshavarz and H. Hecht, "Validating an efficient method to quantify motion sickness". Human Factors, 53, no. 4, pp. 415-426, 2011.

[26] R. Reinhard et al., "The best way to assess visually induced motion sickness in a fixed-base driving simulator". Transportation Research part F: Traffic Psychology and Behaviour, 48, pp. 74-88, 2017.

[27] N. McHugh, S. Jung, S. Hoermann, and R. W. Lindeman, "Investigating a physical dial as a measurement tool for cybersickness in virtual reality". In Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology, pp. 1-5, 2019.

[28] J. F. Golding, "Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness". Brain Research Bulletin, 47, no. 5, pp. 507-516, 1998.

[29] J. F. Golding, A. Rafiq, and B. Keshavarz, "Predicting individual susceptibility to visually induced motion sickness by questionnaire". Frontiers in Virtual Reality, 2, 576871, 2021.

[30] E. Nalivaiko, S. L. Davis, K. L. Blackmore, A. Vakulin, and K. V. Nesbitt, "Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time". Physiology & Behavior, 151, pp. 583-590, 2015.

[31] P. Caserman, A. Garcia-Agundez, A. Gámez Zerban, and S. Göbel, "Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook". Virtual Reality, 25, no. 4, pp. 1153-1170, 2021.

[32] R. Islam, S. Ang, and J. Quarles, Cybersense: "A closed-loop framework to detect cybersickness severity and adaptively apply reduction techniques". In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 148-155, 2021.

[33] E. Chang, H. T. Kim, and B. Yoo, "Virtual reality sickness: a review of causes and measurements". International Journal of Human–Computer Interaction, 36, no. 17, pp. 1658-1682, 2020.

[34] T. Feigl et al., "Sick moves! motion parameters as indicators of simulator sickness". IEEE Transactions on Visualization and Computer Graphics, 25, no. 11, pp. 3146-3157, 2019.

[35] J. R. Chardonnet, M. A. Mirzaei, and F. Mérienne, "Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality". International Journal of Human–Computer Interaction, 33, no. 10, pp. 771-785, 2017.

[36] G. Wang and S. Ayoung, "User adaptation to cybersickness in virtual reality: a qualitative study". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, pp. 5-15, 2019. ISBN 978-1-7336325-0-8 Research Papers.

[37] K. Lim, J. Lee, K. Won, N. Kala, and T. Lee, "A novel method for VR sickness reduction based on dynamic field of view processing". Virtual Reality, 25, pp. 331-340, 2021.

[38] L. Rebenitsch and B. Quinby, "Cybersickness and postural sway using HMD orientation". In Virtual, Augmented and Mixed Reality. Multimodal Interaction: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21, Springer International Publishing, pp. 500-509, 2019.

[39] B. Kitchenham, Procedures for performing systematic reviews. Keele, UK, Keele University, 33, pp. 1-26, 2004.

[40] [40] L. Rebenitsch, "Managing cybersickness in virtual reality". XRDS: Crossroads, The ACM Magazine for Students, 22, no. 1, pp. 46-51, 2015.

[41] L. Rebenitsch and C. Owen, "Individual variation in susceptibility to cybersickness". In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 309-317, 2014.

[42] K. M. Stanney, K.S. Hale, I. Nahmens, and R. S. Kennedy, "What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience". Human Factors, 45, no. 3, pp. 504-520, 2003.

[43] L. Rebenitsch and B. Quinby, "Cybersickness and postural sway using hmd orientation". In International Conference on Human-Computer Interaction, Springer, Cham, pp. 500-509, 2019.

[44] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, and L. Valente, "Minimizing cyber sickness in head mounted display systems: design guidelines and applications". In 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), IEEE, pp. 1-6, 2017.

[45] G. Wang and A. Suh, "User adaptation to cybersickness in virtual reality: a qualitative study". In 27th European Conference on Information Systems (ECIS 2019): Information Systems for a Sharing Society. Association for Information Systems (AIS), 2019.

[46] K. Stanney, C. Fidopiastis, and L. Foster, "Virtual reality is sexist: but it does not have to be". Frontiers in Robotics and AI, 7, 4, 2020.

[47] M. Mousavi, Y. H. Jen, and S. N. B. Musa, "A review on cybersickness and usability in virtual environments". In Advanced Engineering Forum, vol. 10. Trans Tech Publications Ltd, pp. 34-39, 2013.

[48] S. Bruck and P. A. Watters, "The factor structure of cybersickness". Displays, 32, no. 4, pp. 153-158, 2011.

[49] A. K. Ng, C. H. Leung, L. K. Chan, and H.Y. Lau, "Human Factors Related to Cybersickness Tolerance in Virtual Environment". In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 528-532, 2022.

[50] S. Ang and J. Quarles, "You're in for a Bumpy Ride! Uneven Terrain Increases Cybersickness While Navigating with Head Mounted Displays". In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp. 428-435, 2022.

[51] K. Stanney, B. D. Lawson, B. Rokers, M. Dennison, C. Fidopiastis, T. Stoffregen, S. Weech, and J. M. Fulvio, "Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda". International Journal of Human–Computer Interaction, 36, no. 19, pp. 1783-1803, 2020.

[52] D. M. Shafer, C. P. Carbonara, and M. F. Korpi, "Modern virtual reality technology: cybersickness, sense of presence, and gender". Media Psychology Review, 11, no. 2, 1, 2017.

[53] J. J. LaViola Jr, "A discussion of cybersickness in virtual environments". ACM Sigchi Bulletin, 32, no. 1, pp. 47-56, 2000.

[54] P. Era, P. Sainio, S. Koskinen, P. Haavisto, M. Vaara, and A. Aromaa. "Postural balance in a random sample of 7,979 subjects aged 30 years and over". Gerontology, 52, no. 4, pp. 204-213, 2006.

[55] J. Teixeira and S. Palmisano, "Effects of dynamic field-of-view restriction on cybersickness and presence in HMD-based virtual reality". Virtual Reality, 25, no. 2, pp. 433-445, 2021.

[56] W. Bles and A. H. Wertheim, "Appropriate use of virtual environments to minimise motion sickness". Human Factors Research Inst Tno Soesterberg (Netherlands), 2001.

[57] D. Saredakis et al., "Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis". Frontiers in Human Neuroscience, 14, 96, 2020.

[58] M. Mon-Williams, J. P. Warm, and S. Rushton, "Binocular vision in a virtual world: visual deficits following the wearing of a head - mounted display". Ophthalmic and Physiological Optics, 13, no. 4, pp. 387-391, 1993.

[59] K. Raaen and I. Kjellmo, "Measuring latency in virtual reality systems". In International Conference on Entertainment Computing. Springer, Cham, pp. 457-462, 2015.

[60] S. Sharples, S. Cobb, A. Moody, and J. R. Wilson, "Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems". Displays, 29, no. 2, pp. 58-69, 2008.

[61] Y. Farmani and R. J. Teather, "Evaluating discrete viewpoint control to reduce cybersickness in virtual reality". Virtual Reality, 24, no. 4, pp. 645-664, 2020.

[62] Y. Y. Kim et al., "The application of biosignal feedback for reducing cybersickness from exposure to a virtual environment". Presence: Teleoperators and Virtual Environments, 17, no. 1, pp. 1-16, 2008.

[63] J. Jesson, L. Matheson, and F. M. Lacey, "Doing your systematic review-Traditional and systematic techniques". International Journal of STEM Education, 2020.

The Multidimensional Screen Model

Luciane Maria Fadel
Graduate Program in Knowledge Engineering and Management (PPGEGC)
Federal University of Santa Catarina
Florianópolis, SC, Brazil
e-mail: luciane.fadel@ufsc.br

Bolívar Teston de Escobar
Graduate Program in Design (PPGDesign)
Federal University of Paraná
Curitiba, PR, Brazil
e-mail: bolivarescobar@gmail.com

***Abstract*— The visual dimensions of the screen can be projected along the axes of aesthetics, use, or interaction. Each axis plays a crucial role in shaping the design interface practically. These practices are brought to life through various design techniques. A thorough exploration of these dimensions reveals the nuanced experience of the screen's multidimensionality, taking into account both the designer and the interactor realm. The result is a model that delineates three aspects of the multidimensionality of the screen: space, which represents the realm of the designer; meaning, which serves as a shared realm between the designer and the interactor; and performance, which pertains to the interactor's realm. The model has implications for the teaching and practice of design, as it supports the understanding, analysis, and design of the multidimensionality of the screen.**

Keywords-space; performance; meaning; screen; dimensions.

I. INTRODUCTION

Our society is increasingly reliant on screens for various purposes. This trend originated with screens displaying static images, such as painting and photography, capturing a singular moment of imagination. Over time, this evolved into screens featuring moving images, like those in cinemas, which extended the imaginative experience over a period. With the introduction of television into our homes, screens demanded more of our visual attention, akin to paintings and photographs hanging on the wall that clamoured for attention as the cinema. As we transitioned from one type of screen to another, our perception adapted, accepting and interpreting their images as technological advancements.

Images serve as bridges between the object, its representation on the screen and the viewer, intertwining dimensions of use and aesthetics [1] to enrich the interface's complexity.

According to [2], technology is a multifaceted entity encompassing technical knowledge and human attitudes. It becomes ingrained in our routines, offering convenience and enhancing our comfort.

The intricate nature of technology also manifests in interface design education, where one must navigate the practical application of technology in design and the artistry of its poetics or creative expression. Poetics constitute the principles of design that best define an object or work [3]. A central poetic of the digital interface is remediation: the representation of one media into another [4].

For instance, a digital calendar is expected to replicate its printed counterpart. The months appear in a tabular format, with days displayed within cells. Consequently, this paper suggests that perceiving the screen as flat is a direct outcome of the remediation process.

For [4], remediation progresses through a four-level evolution, wherein the representation of a new media diverges increasingly from its predecessor. Therefore, we argue that achieving each level necessitates comprehension of contemporary media and acknowledging its unique language and properties, ultimately giving rise to its poetics.

Another poetic aspect of digital media is its multidimensionality, which is grounded in the principle of numerical representation [5]. This principle enables new dimensions of use by leveraging the artefact to express various aesthetic elements and forms of interaction.

The poetic of multidimensionality, as explored by [6], is established through data density, a concept defined by [7] as the intense flow of information captured and sent by the interactor+artefact dynamic. Consequently, the screen mediates this data density across visible, perceived, or social dimensions. Reference [1] devised a "visual dimensions framework" based on these three dimensions. In addition, teaching poetics benefits from visualising each dimension, allowing students to design and explore the interconnectedness among dimensions to add depth to their creations.

Therefore, we argued in a previous paper [1] that the screen is not flat, as its depth develops by mastering remediation and data density manifested as aesthetic, use and interaction and draws the boundaries for the aesthetics of the screen. This paper develops each dimension regarding design techniques to support teaching and designing a screen that explores its multidimensionality. The result is a multidimensional screen model.

The method follows qualitative research, highlighting screen dimensions from the literature and dialoguing with teaching practice. The model is built by closely reading the framework. This practice enabled many observations about students' difficulties visualising screen dimensions.

The remainder of this paper is organised as follows: Sections II and III present the background review of the aesthetic, use, and interaction dimensions. Section IV extends the visual screen dimension to encompass the design techniques and introduces the multidimensional screen model.

Section V discusses the model. Finally, Section VI draws a brief conclusion.

II. Aesthetic dimension

The multiple dimensions of the screen become undeniable when establishing the possibilities of the interface design aesthetic. One option is simulating three-dimensional objects, i.e., the object is created in its three dimensions. In addition, layers of information, movement, and Information Design (ID) are visual dimensions. These dimensions are constructed through the lens of remediation and data density.

A. *The screen interface*

The interface can be understood as a mediating layer between the artefact and the interactor. The user interacts with the product through the physical or digital interface. Thus, a product can be complex to manipulate, and its use may require a layer of translation of its mechanics. For example, a typewriter presents itself to the interactor through a coating, which hides its gears and leaves enough in view to be used. Therefore, [8] associates design with the interface to link the user, the tool, and the action. Thus, it is likely that the more complex the object's engineering, the more critical the role of the interface as a tool facilitating use. This role becomes evident with digital interfaces, given the complexity of the artefact.

Reference [9] states that the interface is the software for the user, which means it does not matter if the algorithm is highly complex or has a layer of artificial intelligence. What the user perceives is the contact and control over the tool mediated by the interface.

Thus, digital interfaces have made this mediating layer visible (hypermediation), often because of the complexity of its use. By understanding this complexity, many designers seek to create invisible or transparent interfaces (immediacy). However, one of the main qualities of digital objects is their oscillation between hypermediation and immediacy.

This oscillation is also referred to as remediation by [4]. The authors argue that the opacity of the interface is necessary for interaction to occur, as the interactor needs to see the options to act on them (hypermediation). On the other hand, immersion happens when engaging with the content, and the interface becomes transparent (immediacy). Therefore, this oscillation is another poetic of interactive media and a dimension of the interface.

To decrease the oscillation, [10] advocates the narrativization of the interface. Lessening the oscillation can be accomplished by (1) narrativized 'look and feel' of the interface, (2) behavioural mimic and behavioural metaphors, (3) narrativized perspective, and finally, by building (4) bridges and mixed-reality interfaces.

The 'look and feel' incorporates narrative elements into the graphic representation. The aforementioned has to do with the visual identity of the artefact and how the imagery representation is expected to reinforce the project concept. For instance, feedback could be presented as illustrations, reinforcing the adopted narrative.

Also, interface elements can mimic behaviours or behavioural metaphors. For example, if an interface element demands an urgent response, its graphical representation can assume a hurried behaviour, such as getting agitated.

On the other hand, the narrative perspective acts on the depth dimension of the screen. That is, the screen's graphic design makes the z-axis of the spatial representation explicit. This representation is evident in in-game scenarios or environments where the interactor can simulate moving around.

Finally, data density can support the bridges and mixed-reality interfaces establishing digital and virtual connections. Augmented reality artefacts are excellent examples, as they apply new layers of dynamic data on top of the captured image of the place (Figure 1). Other bridges can be established using interactors' information to capture environmental information. Locative media are examples of this dynamic.

Thus, design techniques to support designing a multidimensional screen involve mastering remediation as oscillation and representing an older media into a new one. Mastering oscillation evolves with a narrativized interface, and when one of the concepts associated with the interaction, such as dialogue, transmission, tool use, optimal behaviour, embodiment, experience, and control, is clear for the interactor. Section III-C discusses these concepts. Mastering the representation of one media into another emerges, encouraging the creation of the meaning of the artefact in the use, language, genesis and ecology of the artefact [11]. Thus, design techniques, such as metaphors, affordance, user mental models, and Gestalt, can be applied to foster meaning-making.

Figure 1. Example of augmented reality artefact using Google translate App.

B. *Tri-dimensional objects*

Treating objects in three dimensions allows different renderings to simulate their spatiality, such as rotating the object or moving it in the screen space. So, it requires the object to be thought of in true 3D, which moves away from the printed media as it requires a 2D representation. In 3D, the design domain would approach the realm of sculpture because it would encompass elements of 3D representation such as body, weight, movement, and lines of action, among others, expanding design to volume treatment.

Figure 2 shows an interface of an app prototype that presents content on the skeletal system for medical students. The skeleton is presented in 3D and can be rotated and zoomed. This paper suggests that this control over the visualisation encourages exploring the object. 3D objects occupy the multidimensional space of the screen and offer many possibilities for representation. Just as the screen's surface allows it to be treated as a 3D surface, objects can also be designed within 3D dimensions.

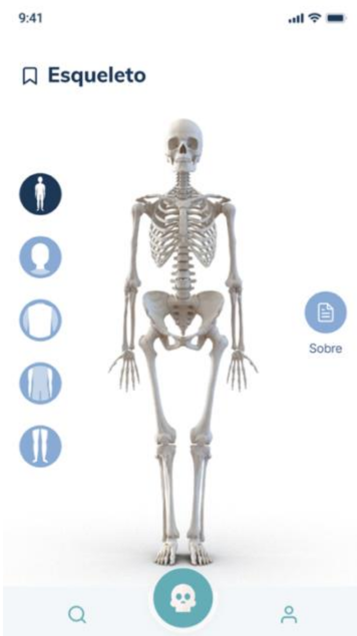

Figure 2. The Meduca App screenshot was developed by Boesing and Wagner (2020).

The calendar, for example, which is constantly translated into digital with firm reference to its printed predecessor, could be represented by a 3D object, such as a sphere. The spherical calendar allows movement to explore new possibilities of representation.

In addition to 3D representation, space simulation enables layers of movement and different forms of interaction. By treating the screen as a three-dimensional space, motion layers are created in the depth of this space, where objects can move around. For example, a disabled element can occupy a bottom layer of space and project to forward layers when enabled.

Moreover, the space can become active, posing as a design and communication element. As advocated by [12], the digital space, as a remediation of the medium, expands the possibilities of interaction as it becomes a meaningful dimension. Beautiful interfaces explore 3D dynamics by presenting 3D objects and a 3D scenario. The beauty is supported by a spectacle involvement that evolves through the excitement about the images and 3D forms.

The screen's rectangle format suggests the space as a two-dimensional space. But some examples, such as the Apple Watch®, bring new possibilities when the screen is designed in its three-dimensional space (Figure 3). The surface is considered spherical, which implies that the graphic elements can slide around the sphere, assuming different sizes when traversing it. They increase in the centre and decrease when approaching the edges. The treatment of the surface in 3D enables new attention arrangements, given primarily by size and position.

Figures 2 and 3 are examples of remediation. The former remediates a real skeleton, and the latter remediates the screen as a ball. Thus, remediation and space-medium remediation are techniques for designing 3D representation.

Figure 3. Apple Watch Interface.

C. *Information layers*

Two fronts provide an understanding of information layers: position and meaning. While the positioning layer defines different layers in different spatial positions (on any of the three axes of the screen), the meaning layer implies different degrees of importance built through Information Design using contrast, hierarchy, typography, composition, colour, and image. Figure 4 shows a login screen highlighting the input fields in a front layer. The background layer is out of focus to improve the distance between layers.

The positioning layer uses the spatial geometry of the screen to place the information layers. Spatial geometry implies the independence of the layers, both at the content and interaction levels.

One of the best examples of this arrangement of multiple layers on the same screen is Augmented Reality (AR) applications. AR presents a layer of dynamic information on the physical environment, whether captured by a camera or not. That is, its definition guarantees a multidimensional understanding. AR can happen in 3 arrangements: (1) through information projected on a physical space, such as films projected on buildings; (2) using an instrument to capture the physical space and, on the same screen, insert the dynamic information; and; (3) using glasses or lenses on which the information is projected while the ocular system captures the physical space [13].

AR is distinguished from a simple projection of a video onto a screen by considering the three characteristics that [13] attributes to AR:

- It combines the real and the virtual;
- It is interactive in real-time;
- It is registered in three dimensions.

The multidimensionality of the screen is explicit, given that the interactor is the one who builds it. This co-creation allows the interactor a certain degree of control, given the dynamism of the composite image.

In the composition of AR, one can have several layers of information organised by the distance between the object and the interactor, the screen's permanence, the interactor's importance, or any other design criterion. These criteria that are exposed by AR composition can be applied in other interface design projects. AR makes it easier to understand this multidimensional composition of information. Therefore, AR and Information Design can be considered design techniques to improve the screen depth. The independence between layers relies upon the space-medium remediation.

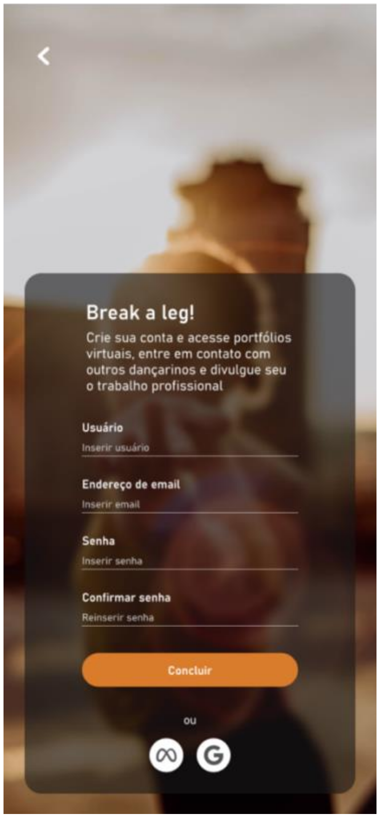

Figure 4. The Pulse App screenshot was developed by Bevilacqua and Paiva.

D. Movement

Movement is another screen dimension that can be understood using four approaches: (1) moving images and objects, (2) user physical interaction, (3) the movement of the interactor in the space, and (4) the movement of the device itself. Movement is the poetic aspect of digital media that attracts attention and reveals the wonders of the multidimensional screen. Movement can be applied to deliver feedback, transitions, and humour.

Given the principle of numerical representation, objects projected onto the screen can be created in true 3D, which allows the objects to be manipulated on all three axes, as discussed in Section II-B. The object movement through animations, micro animations, sliding in different directions, and appearance, among others, adds dynamism to the interface elements, providing feedback to the interactor. Feedback, such as loading animation, uses movement to inform the interactor without distraction.

Moving images, such as videos, graphic motion, or animations, are characteristic of media based on time. These complex media translate narratives into different dimensions, such as time, space, or sequential images. For example, the horizontal scrolling of gapsystudio.com remediates an art gallery as the interactor moves the 3D scenario. This remediation facilitates understanding the space sequentially and the studio's work as art. In addition, the transition from one page to another, between views or states, can be designed using a visual effect, such as fade, mask, or slide. Transition builds continuity that guides the attention from one object to the other, reinforcing the depth between objects.

Humour can be delivered through movement because the animated objects are perceived as alive and reactive and, therefore, they become partners in a shared experience. Emphasis on-screen depth is essential because the animated object shows personality. Micro-interactions shape humour using small movements that smooth interaction.

Interactor's movement occurs in physical space or/and on the screen, navigating among pages. As argued by [7] and [14], the former is supported by mobile technology with small screens. The coincidence of movement and the creation of spatial representations is called "performative cartography" [14]. This simultaneity highlights the screen depth, as the image is created as it is perceived (see Section III-B). The moving interactor also implies locative media, which explore the augmented space [7] by triggering an interaction. The three principles of performative cartography on locative media [14] can enhance the depth of the screen: tagging, plotting and stitching. Tagging adds metadata to objects or locations. Plotting places the tagged objects into a map while stitching integrates visual layers and digital cartography [14]. The multidimensionality of performative cartography requires multiple levels of the interface. At the same time, software communication tags the objects and the screen plots and stitches as a navigable space.

The device's movement brings new possibilities of embodied or haptic interaction. That is, the control of the screen can occur through actions with the device. For example, shaking the device can switch pages. It is also a powerful accessibility technique.

E. Information design (ID)

The design project introduces new dimensions by intertwining interaction and navigation within Information Design (ID). These layers incorporate various design elements that enhance the screen's depth. The ID plays a crucial role in processing information, enabling seamless interaction and navigation. For instance, when creating an interactive button, the ID is meticulously designed to provide clear guidance on the available actions.

Navigation Design shares a similar dynamic, defining pathways across digital pages, while Information Design focuses on delivering optimal solutions for user guidance. From metaphor to practical implementation, navigation signifies a connection with the artefact beyond the physical confines of interaction space. It can be argued that navigation is an interaction mode that gained popularity with hypertext systems in the 1980s [15]. These systems, featuring graphic-textual interfaces, are explored through user inputs like clicking or tapping, leading to the discovery of interconnected information segments. However, it's important to note that scholars like [16] point out that navigating information has historical roots in traditional artefacts such as books and printed materials. The visual cues in these materials, such as page numbers, headings and indexes, can be viewed as "navigational devices" between documents contents.

The metaphor of navigation emerges from the imperative to comprehend a spatial semantic system, as suggested by [17]. In intricate interfaces with information layers distributed horizontally and vertically, navigation becomes pivotal in shaping the experience. This process defines personal preferences, information retrieval patterns, and identities, underscoring navigation as a dialogue between the user and the system. The interaction reveals boundaries between the two entities, both technically and politically.

It is, therefore, possible to comprehend these navigable systems as mediators of the relationship between users and the world represented by the system's information. This relationship entails technological mediation, as understood in the post-phenomenological sense articulated by [18] and [19]. This concept involves an interpretation of technology based on how humans engage with it dynamically, where the human actor defines themselves in conjunction with their technological counterparts through the interactive process.

The poetics of navigation are intricately connected to movement. The resulting perception is so impactful that it gives rise to new ways of existence and interaction. The concept of cyberspace, initially depicted in the science fiction novel Neuromancer [20], embodies an interaction model that transcends mere positioning on a two-dimensional screen. Instead, it embraces the phenomenon of semantic exchange as described by [21]. Navigation can be seen as a performance within a specific time and space framework inherent to the experience of projecting oneself into abstract space.

Through "navigational practices" such as search systems, dialogical operations, or simply browsing smartphone and tablet screens, a responsive dimension emerges from the programming of artefacts. Consequently, a system can exhibit different states depending on its usage. Each usage presents a unique phenomenon, and each navigation action serves as a form of expression. As a result, interaction modalities such as clicking, scrolling, or zooming in on a map can be viewed as a technological manifestation of the human capacity to navigate. Navigation Design introduces multidimensionality into its conceptual framework by navigating information that may be distributed in depth (layers) complexity or exploring the special aspect that constitutes the object.

Interaction Design encompasses the entire process of designing an interactive object, which includes information and navigation design [22]. Moreover, it can be viewed as the design of the mechanisms that enable users to navigate effectively [23].

Hence, Navigation Design is responsible for mapping out the potential paths, while Interaction Design devises the mechanisms to empower interactors to act upon the interface. Information Design, in turn, conceptualizes these mechanisms.

III. USE AND INTERACTION DIMENSIONS

The prevalence of mobility has heightened the use of digital objects, accentuating the screen's significance. This attribute underscores the necessity for the usage to fall within the encompassed reception area for data transmission and reception. The interface is dynamically developed through the seamless integration of receiving, processing, and presenting data, a process referred to as performative cartography [14]. Thus, mobility and performative cartography emerge as integral dimensions of the screen.

A. Mobility

Mobility, defined as the utilization of digital products across various locations, is facilitated by individual Internet access technology and the compact size of artefacts like smartphones and tablets, allowing for their usage while users are on the move. This mobility has enhanced data density, transforming space into an active entity by gathering user data or providing locative data and information. The screen now functions as a gateway through which information about a location is relayed to the user. Locative media, including games and apps, have the potential to introduce novel levels of responsiveness influenced by spatial factors.

Reference [5] coined the term "augmented space" to describe this interactive space and proposed that this expansion should be viewed as a concept or a cultural and aesthetic practice. This reconceptualization broadens the scope of creative possibilities, transforming the screen into a multifaceted space. Within this complexity lies the notion of continuous monitoring, a process often overlooked by interactors but essential to consider within the design domain. Monitoring is an inherited aspect that can be disregarded or integrated by digital artefacts, requiring addressing it in the interface.

As a cultural practice, numerous objects have become intertwined with work and leisure routines, such as ubiquitous computing, artificial intelligence, augmented reality, and wearables. Despite this integration, aesthetics as an art form often remains disconnected from the interactor's presence and surroundings. Furthermore, these objects are typically envisioned in isolation, leading to a lack of consideration for their overall ecology. For example, the Internet of Things (IoT) features could be integrated into the digital artefact design to enhance the use of data and functionalities.

These are some of the challenges to consider within the mobility dimension. The cultural practice of performative cartography makes these challenges explicit.

B. Performative cartography

The simultaneous displacement of individuals in both the physical environment and on the screen is defined as "performative cartography" [14]. In this process, the interactor navigates the interface while the interface is being shaped. A practical example of this concept is the map displayed in Google Maps, which is dynamically generated based on the subject's real-world location (Figure 5).

The visualisation and image construction process co-occur in a creative endeavour that [14] describes as a 4D operation within a 3D space. To address the challenge of representation, the author proposes that the fourth dimension should focus on space-time rather than solely on time. The assertion that time and space are manifested during usage supports this suggestion. Consequently, the concept of performative cartography entails changes, distinctions, and a degree of unpredictability in the evolving movements.

Figure 5. Example of performative cartography using Google Maps.

C. Interaction

Due to interactivity, interaction with digital objects occurs in new dimensions in addition to the physical movement of the device. Interactors' experience with the screen is shaped by their actions and perceptions - how they engage and comprehend. This dynamic establishes a two-way communication process between the interface and the interactors. The interactivity of a narrative experience is categorized [24] into four modes: cognitive, functional, explicit, and meta-interactivity. Cognitive Interactivity [24] involves revisiting a text that challenges previous understanding. Functional Interactivity pertains to physicality, usability, and Information Design. Explicit Interactivity examines user action within the interface and the interaction itself. Meta-interactivity explores engagement with the text beyond the immediate experience, such as discussing it with others.

Interaction can be understood through various categories outlined by [25], encompassing concepts such as dialogue, transmission, tool use, optimal behaviour, embodiment, experience, and control. Each concept offers a unique perspective on the relationship between product and human, shaping the poetics of use and meaning within the screen space.

The interface engages in a dialogue with interactors, aiming for a seamless conversation where both parties understand how the interface functions and the responses they can provide. This mutual exchange underscores the importance of considering the mental model dimension in design.

Viewing interaction as a form of transmission requires a focus on the quality of the communication channel and the amount of information transmitted, with noise management being a crucial factor.

When interaction is seen as tool use [25], it influences how users interact with the system, emphasizing the mediation role of the interface and the user's engagement with the tool itself. This perspective requires considering the extension of the body and senses, as proposed by [26].

Optimal behaviour in interaction includes balancing performance and resource allocation (both human and technological) to achieve the best outcomes. Therefore, the time-space-statistical dimension [25] of the screen emerges.

Designing interaction as embodied requires situating its agents in a physical world. Reference [25] indicates that situating interaction involves intention, coupling, and context.

Conceiving interaction as experience means understanding how the interaction unfolds. It considers the technology's qualities and aesthetic, emotional, and holistic aspects, deepening the value attributed to the screen. Finally, the control in interaction design focuses on error management and system adjustments based on feedback to align actions with the desired outcomes.

As discussed in Section II-D, haptic interaction provides feedback through tactile sensations like vibration, enhancing accessibility. The advent of touch screens has introduced new interaction possibilities [14] eliminating the need for intermediary devices like a mouse and allowing for touch-based interactions that incorporate narrative elements, enriching the depth experience of the screen.

IV. The Multidimensional Screen Model

Reference [1] argued that the multidimensionality of the screen is an inherent characteristic of digital conceptualisation and introduced the screen dimensions framework. This framework examines the screen's dimensions through aesthetic, use and interaction lenses. (1) Aesthetics encompasses the graphic qualities of the interface, 3D representation, space as a medium, layers of information, movement and the design of information. (2) Use dimension comprises mobility and performative cartography. (3) Interaction is associated with narrative interactivity and the interaction concept itself.

Furthermore, the nature of the media involves remediation. Therefore, defining its poetics, including the screen dimension, is essential in creating new media. Section II outlines various design techniques that contribute to determining these poetics. This paper contributes to the screen dimension framework [1] by introducing design techniques tailored to each dimension (Figure 6).

The outcome expands the framework for the visual dimensions of the screen, which are structured around three

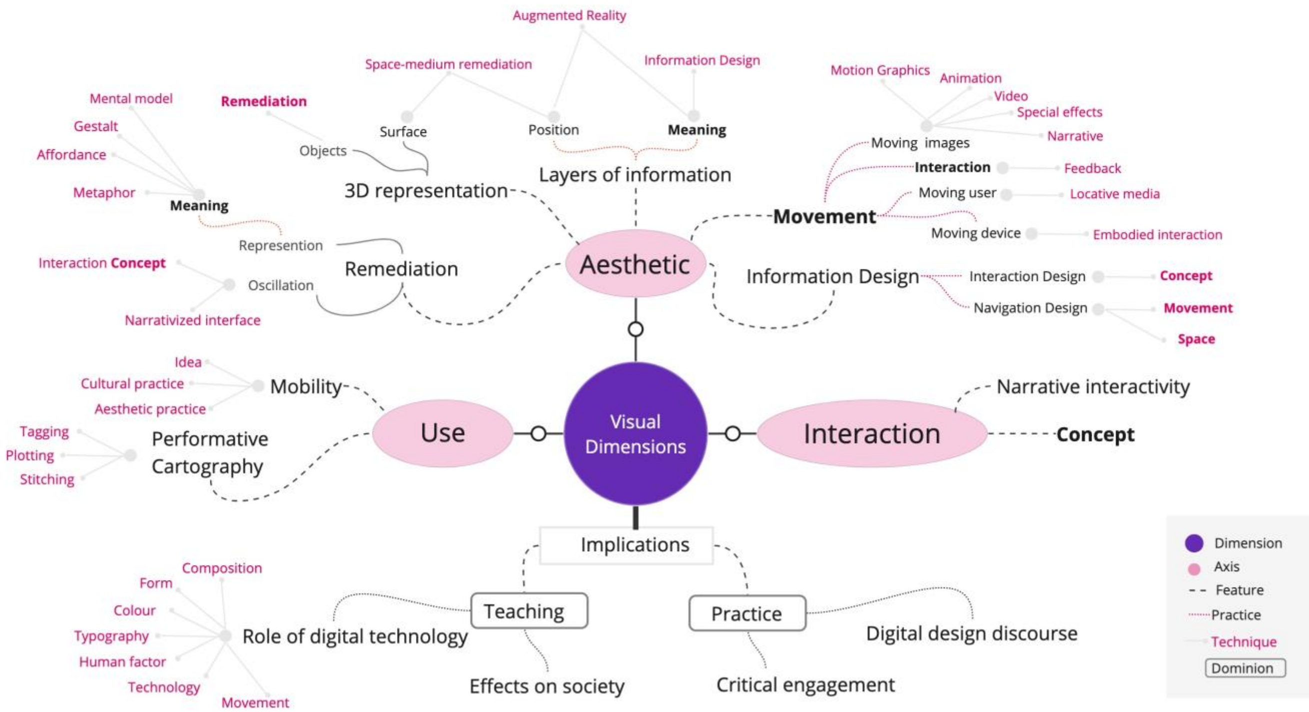

Figure 6. The expanded visual dimensions framework.

axes: aesthetics, use, and interaction. Subsequently, the paper discusses practices with qualities that strengthen the screen's multidimensionality. Lastly, it describes techniques to create a design capable of effectively representing these visual dimensions.

The close reading method was applied to establish a connection between the framework's elements and their respective realms. Several tables were created to compose a configuration that articulated the framework regarding the designer and the interactor realms. The final table highlighted elements such as narrative, remediation, interaction or movement in various configurations and positions. The dynamic interplay of these elements suggests three aspects of the screen's multidimensionality: space, meaning and performance.

Space encompasses the constituent elements of the interface and those linked to the materialisation of the design. Meaning pertains to the components contributing to the interactor's apprehension of the design. Finally, performance encompasses the aspects that manifest through the interactor's action. These aspects intersect, forming the axes of the visual dimension: aesthetic, use and interaction.

In this setup, aesthetics is situated at the intersection of space and meaning, as both aspects engage the senses. Use occurs at the intersection of meaning and performance, as it represents the realization of the artefact by the interactor. The artefact is fulfilled when it is put to use. Finally, interaction resides in the intersection between performance and space, where the materialisation becomes part of a dynamic reconfiguration with the interactor. Figure 7 shows the model

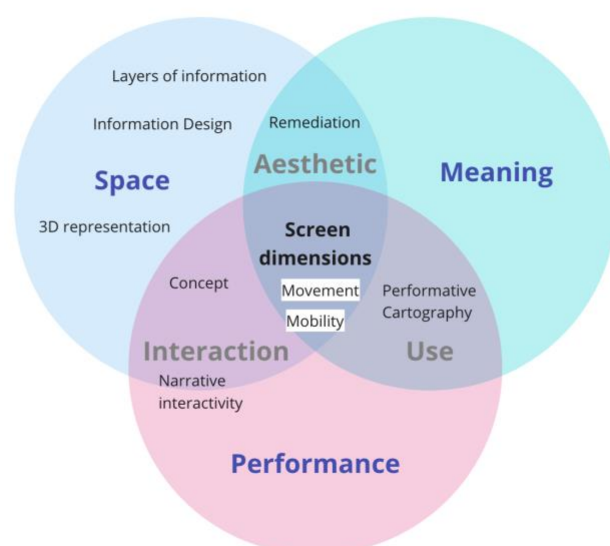

Figure 7. The multidimensional screen model.

that accomplishes the three aspects of the multidimensional screen.

V. Discussion

The model establishes three aspects of screen multidimensionality: space, meaning and performance. These aspects are designed to enhance the experience of screen multidimensionality. The space aspect encompasses the dimensions crafted by the designer using various design features, making it a tangible construction within the realm of the designer. Space plays a crucial role in the experience of multidimensionality as it often draws upon the three-dimensional metaphor of physical space, a standard visualization tool. The performance aspect of the screen's multidimensionality is experienced through the active participation of the interactor. The interactors generate the interface through their interactions, completing the object and adding a dynamic element to the experience. Meaning, on the other hand, is attributed by the interactor to the object, especially when it delivers aesthetic and functional possibilities envisioned by the designer thus supporting the interactor's performance. The model suggests that performance and meaning are more abstract than space, given the intense involvement of the interactor. This participation underscores that space belongs to the designer's realm. At the same time, performance is the interactor's realm, and meaning emerges as a co-creation between the designer and the interactor.

Moreover, the model can be applied in teaching and practising design to explore screen dimensions [1] and their respective aspects, offering a structured approach to deepen understanding and analysis. Therefore, the model is a teaching and design tool for effectively creating and analysing projected dimensions.

A. Teaching practice

Understanding the multidimensionality of the screen requires responsible teaching of interface design. This starts with acknowledging digital technology as a customary element in our lives. This approach prompts critical questions about technology's role in society and everyday interactions.

Reference [1] proposed addressing these questions, focusing on seven axes of design: composition, form, colour, typography, human factors, technology, and movement. Each axis contributes to the artefact's role in daily life, and design education should emphasize their interconnections to channel their societal impact effectively.

Composition involves organizing elements within the design, and spatial understanding can be expanded through meaningful remediation techniques like narratives and imagery. Form explores representing information in various geometric or organic forms, such as images, photographs, graphics, or illustrations. The students could be motivated to try different forms to express the same information and articulate the final alternative in terms of dimension.

Colour delves into the dynamic representation of the light on the screen. Light is dynamic and complex as it presents millions of possibilities. Thus, screen depth can be creative using light. Students should be motivated to create simple solutions, with light as the only element.

Typography explores the variability of font characteristics to enhance communication.

Human factors consider that people will interact with the object and bring all their previous and expected experiences into the new one. Thus, design benefits from human sciences, such as psychology, sociology, and anthropology.

Technology serves as the object's medium, requiring designers to master its properties and qualities for harmonious, effective, efficient, and beautiful work. Thus, the screen's dimensions are expected to contribute to this harmony.

The movement has become an axis for digital design, given the principle of numerical representation, which supports time-based media. Furthermore, the movement has been studied through the persistence of vision, which has shaped Gestalt studies since its publication by Wertheimer in 1912 [27]. Or even as an evolutionary priority over the perception of forms [28], recognising danger was decisive for survival. The perception of movement is a fundamental aspect of vision; even micro-movements can attract attention. Therefore, movement, as one of the axes of design, emphasises the depth of the screen, as commented on in Section II-D.

Accordingly, to address the role of digital technology, design teaching could expose different contributions of the screen depending on the type of artefact in focus. For each axis, the design elements and their contribution to the role of the screen would be related. This construction promotes a critical position and develops the analytical skill of the designer. The three screen dimensions of Aesthetics, Use, and Interaction could elaborate on other issues raised in this paper.

The study of aesthetics enhances understanding of the interface, identifying it as an active mediator. An active medium requires viewing the interface as a dynamic entity oscillating between opacity and transparency. In addition, the elements of the interface support and react to the interactor's actions, providing information and feedback. Furthermore, teaching 3D modelling encourages abstract thinking about spatial dimensions on the screen and its objects.

Design education delves into layers of information by its very essence. Objects (type, form, and function), action (passive or interactive), hyperlinks in depth, design choices such as gamification and metaphors, or even behaviour, such as movement, shape this essence.

Motion often takes a back seat in design projects. Therefore, the crucial need for design education to emphasize its incorporation into projects is notorious.

Teaching movement requires considering time and space and favours narrative constitution. Teaching narrative as a poetics of design requires treating narrativization of the interface, that is, treating the design elements as passive or active agents of the narrative. Concepts and elements of narrative will be revisited for this purpose.

Information Design is a consistent element in design projects, albeit assimilated into the specialities required in digital design, such as interaction design and navigation. Teaching digital Information Design underscores the interconnectedness and distinctions among these specialized areas. As seen in our brief description of the navigational aspects of the experience, there remains ample room to investigate the role of Information Design in the expressive and political outcomes that emerge from navigating informational spaces. An individual's identity within a specific context is intricately linked to their level of visibility, accessibility, and interaction. Instructing designers on what to permit or restrict in user terms forms the foundation of navigational interfaces.

The implications of dimensions in teaching about use and interaction lie in the recognition of mobility and performative cartography as requirements and properties of the object. Therefore, teaching can highlight such factors and discuss the axis of technology and its consequences on artefact's use, production, and creation.

The interaction dimension implies teaching interactivity through some biases such as narrative, embodied, and agency. These biases can broaden interaction treatment and incorporate new technology methods, presenting the potential for accessibility.

B. *Design practice*

In design practice, the multidimensional aspects of the screen shape a digital design discourse that emphasizes iterative and responsible construction and creation of artefacts. Designers engage with the space aspect through aesthetics and interaction, critically examining elements that enhance screen depth and their effect on reception by the interactor. This active and collaborative practice values diverse perspectives and knowledge sources, leading to innovative design solutions and systematic analysis processes.

The performance aspect identifies these two dimensions based on the active participation of the interactor. This performance clarifies the value of this participation, which can be understood as a project requirement and therefore, projectable.

Design practices are shaped by design methods. In the case of the Iterato method [23], the spatial aspect is materialised in the structure and sensory design phases but has been thought of since its conceptualisation. The screen dimensions are part of the artefact's design that constitutes, creates, materialises, and promotes the construction of meanings by the interactor. The aspect of meaning establishes this imbrication.

Overall, having a model for creating multidimensional screens provides visibility to these dimensions and requires a comprehensive understanding of the role and interactions between space, meaning, and performance in design and educational practices.

When confronted with existing models [29,30,31,32, 33], the Multidimensional Screen Model complements the knowledge framework for this practice. The complementation situates the design of the screen as the interconnection between the interactor's use, the design of aesthetics and the performance between the two. It also highlights the consequences of design teaching, which is rarely mentioned.

Reference [29] reviewed 45 articles focusing on the interface for mobile learning and identified four critical dimensions of the user interface: (a) design principles of mobile learning applications, (b) context of use, (c) hardware specifications, and (d) modelling language. The design principles refer to the craft of the interface, highlighting four elements: size of the elements, proximity, transition and minimalist design. The context of use centres on the design based on the users' context, especially their mobility. The hardware specifications require support collaborative work among users and adaptations to varying device screen sizes. The modelling language advocates for an object-oriented approach as the preferred modelling language for user interface design.

Reference [30] introduces a Schema that outlines the dimensions of the User-Product Experience. The schema, which underscores the interaction between concrete and abstract product dimensions, the user, the context of use and the temporality of the experience, is not just a theoretical construct. It is a practical tool based on user and product sensors and responses in a specific use context.

Reference [31] delves into the intricacies of designing augmented reality, highlighting the importance of human perception dimensions, product dimensions, context of use, and temporal factors. It underscores the need to separate geometric relationships (locales) from semantic relationships (contexts) for effective design. In virtual reality design, the focus shifts to examining input as semantic dimensions and virtual targets as dimensional outcomes, further illustrating the complexity of the design process [32].

Reference [33] creates a consumer experience using five dimensions organized into three categories. Effort and Usability: (1) browsing, searching, finding and (2) comprehending, consuming, interacting); Power and Usefulness: (3) creating meaningful content and interactions; Persuasion and Emotion: (4) responding to value — calls to action and (5) perception of brand.

Table I summarizes these models. Other references for designing interfaces focus on specific dimensions such as the cognitive load of the learner [34], cultural dimension [35], or personality dimension [36].

While these models acknowledge the key role of user, product and context properties in the experience (UX), the model's focus in this paper is not on UX itself but instead on the properties of the screen. This approach aims to enhance

TABLE I. PRODUCT AND USER EXPERIENCE MODELS

Authors	Focus	Dimensions	Features
(HAMZAH, PERSADA and HIDAYATULLAH, 2018)	Mobile learning applications	Design principles	Size Proximity Transition Minimalist design
		Context of use	Mobility
		Hardware specifications	Collaborative works Screen size
		Modelling language	Object-oriented approach
(BONGARD-BLANCHY and BOUCHARD, 2015)	Product design	Human perception	Physiological sensors Cognition and affect Percept-action loop
		Product	Sensors Functional behavioral Semantic sensorial properties
		Context of use	Situational Cultural Social dimensions
		Temporal dimension	Interaction sequences Long-term UX
(SCHMALSTIEG, FUHRMANN and HESINA, 2000)	Collaborative augmented reality	Geometric relationships (locales)	
		Semantic relationships (contexts)	
(YEO, KWOK, *et al.*, 2024) Metaphor-guided quadrant model	Virtual reality	Input dimensional semantics	2D-Dimensional Input 3D-Dimensional Input
		Dimensional outcomes	2D or Less 3D or More
(GODDARD, MCLEARY and GORNEY, 2008)	E-commerce	Browsing, searching, finding	Effort (Usability)
		Comprehending, consuming, interacting	
		Creating meaningful content and interactions	Power (Usefulness)
		Responding to value — calls to action	Persuasion (Emotion)
		Perception of brand	

understanding of the relationships between different dimensions and their potential impact on the user experience.

VI. CONCLUSION

Several researchers have examined the multidimensional nature of screens across various design axes, including composition, shapes, colour [37], typography [38], human factors [39], technology [40], and movement [41]. The emergence of virtual and augmented reality artefacts has stressed the need for research on other dimensions of the artefact [42].

Establishing the boundaries of different screen dimensions inspires investigations and draws attention to the complexity of the screen. This complexity goes far beyond the reach of this paper because it involves social, emotional, psychological, historiographic, and philosophical dimensions, among others.

This paper contributes to this field of research and practice by presenting a multidimensional screen model. The model locates the space, meaning and performance as aspects of the multidimensional screen experience. The design for this experience offered the aesthetic, use, and interaction dimensions of design techniques to support teaching and designing a screen that explores multidimensionality. Our experience teaching digital design pointed to great difficulty for students in giving depth to the screen. One issue is the visualisation of this depth. It is hoped the multidimensional screen model supports the visualisation and creation of these dimensions.

The taxonomy of these dimensions and their implications for user experience are promising avenues for future research, which the author intends to pursue. Further studies could explore the relationship between these dimensions and their impact on UX, providing a deeper understanding of the multidimensional screen model and its applications.

REFERENCES

[1] L. M. Fadel, “The Screen is not Flat,” in *IARIA Congress 2023: The 2023 IARIA Annual Congress on Frontiers in*

Science, Technology, Services, and Applications, Valencia, 2023.

[2] A. Cupani, "Types of technology and their cultural consequences," *Revista Dialectus,* pp. 82-95, 2020.

[3] D. Bordwell, Poetics of Cinema, New York: Routledge, 2007.

[4] J. D. Bolter and R. Grusin, Remediation, MIT Press paperback edition, 2000.

[5] L. Manovich, The Language of New Media, Cambridge: MIT Press, 2001.

[6] L. M. Fadel and A. Coelho, "Augmented reality in information design," In press, 2023.

[7] L. Manovich, "The Poetics of Augmented Space," *Visual Communication,* pp. 219-240, 2006.

[8] G. Bonsiepe, Gui Bonsiepe: Interface - An Approach to Design, Jan Van Eyck Akademie, 1999.

[9] D. Norman, Design of Everyday Things: Revised and Expanded, New York: Basic Books, 2013.

[10] J. Bizzocchi, M. B. Lin and J. Tanenbaum, "Game, narrative and the design of the interface," *International Journal of Art and Technology,* vol. 4, pp. 460-479, 2011.

[11] K. Krippendorff, The Semantic Turn: A New Foundation for Design, Boca Raton: CRM Press, 2005.

[12] L. M. Fadel and J. Bizzocchi, "Designing background as space medium remediation," *Estudos in Design (Online),* v. 17, 2019, pp. 5-22.

[13] R. T. Azuma, "A Survey of Augmented Reality," *In Presence: Teleoperators and Virtual Environments,* vol. 6, no. 4, 1997, pp. 355-385.

[14] N. Verhoeff, Mobile Screens The Visual Regime of Navigation, Amsterdam: Amsterdam University Press, 2012.

[15] J. Conklin. "Hypertext: An introduction and survey". *Computer*, v. 20, n. 09, 1987, pp. 17-41. doi.org/10.1109/MC.1987.1663693.

[16] J. Drucker. 2008. Graphic Devices: Narration and Navigation. Narrative, Vol. 16, No. 2, pp. 121-139. doi.org/10.1353/nar.0.0004.

[17] S. Padovani, D. Moura. Hypermedia Navigation: A User-Centred Approach. Rio de Janeiro: Modern Science, 2008.

[18] P.P. Verbeek. What things do: Philosophical reflections on technology, agency, and design. Pennsylvania: Penn State Press, 2005.

[19] D. Ihde. Postphenomenology and technoscience: The Peking University lectures. New York: Suny Press, 2009.

[20] W. Gibson. Neuromancer. New York: Ace Publishing, 1984.

[21] D. Ihde. Technology and the lifeworld: from garden to earth. Bloomington: Indiana University Press, 1990.

[22] Y. Rogers, H. Sharp and J. Preece, Interaction Design, Nova Jersey: Wiley, 2023.

[23] B. S. Gonçalves, L. Fadel, C. R. Batista and M. Woloszyn, "Iterato: a method for designing interactive digital objects," in *14º Brazilian Congress of Research and Development in Design*, Gramado, 2022.

[24] E. Zimmerman, "Narrative, Interactivity, Play and Games", in *First Person: New Media as Story, Performance, and Game*, Cambridge, MIT Press, 2004, pp. 154-164.

[25] K. Hornbæk and A. Oulasvirta, "What is interaction?," in *CHI 2017*, pp. 5040–5052, Denver, 2017, doi.org/10.1145/3025453.3025765.

[26] M. McLuhan, "The Playboy Interview: Marshall McLuhan," *Playboy Magazine*, pp. 53-74, 1969.

[27] R. Sekuler, "Motion Perception: A modern view on Wertheimer's 1912 monograph," in *On Perceived Motion and Figural Organization*, London, Lothar Spillmann, 2012, pp. 101-126.

[28] R. L. Gregory, Eye and Brain: The psychology of seeing, New York: McGraw-Hill Book Company, 1978.

[29] A. Hamzah, A. G. Persada and A. F. Hidayatullah, "Towards a Framework of Mobile Learning User Interface Design," In *ICEEL '18*, 2018.

[30] K. Bongard-Blanchy and C. Bouchard, "Dimensions of User Experience - from the Product Design Perspective," *Journal d'Interaction Personne-Système,* 9 November 2015.

[31] Schmalstieg, A. Fuhrmann and G. Hesina, "Bridging multiple user interface dimensions with augmented reality," In *IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, Munich, 2000.

[32] A. Yeo, B. W. J. Kwok, A. Joshna, K. Chen and J. S. A. Lee, "Entering the Next Dimension: A Review of 3D User Interfaces for Virtual Reality," *Electronics,* p. 10.3390/electronics13030600, 3 13 2024.

[33] P. H. Goddard, S. McLeary and D. Gorney, Five Dimensions of User Experience Balancing business and user experience perspectives to create successful e-commerce sites, Fairfield: Human Factors International, 2008.

[34] M. A. Faudzi, Z. C. Cob, S. A. Sharudin, R. Omar and M. Ghazali, "The Effects of User Interface Design for Mobile Learning Application on Learner's Extraneous Cognitive Load: A Conceptual Framework," In *Asian CHI '23*, 2023.

[35] A. Marcus and E. W. Gould, "Cultural Dimensions and Global Web User-Interface Design: What? So What? Now, What?" In *6th Conference on Human Factors and the Web*, 2000.

[36] M. Sarsam and H. Al-Samarraie, "A First Look at the Effectiveness of Personality Dimensions in Promoting Users' Satisfaction With the System," *SAGE Open*, doi.org/10.1177/2158244018769125, 2018.

[37] Z. Wang, W. Liu and M. Yang, "Data-driven multi-objective affective product design integrating three-dimensional form and colour," in *Data-driven multi-objective affective product design integrating three-dimensional form and colour Zeng Wang, Weidong Liu, Minglang Yang Neural Computing and Applications (NCAA)*, pp. 15835–15861, 2022.

[38] J. J. Shen, K. Jin, A. Zhang, C. Breazeal and H. W. Park, "Affective Typography: The Effect of AI-Driven Font Design on Empathetic Story Reading," in *CHI EA'23*, pp. 1–7, doi.org/10.1145/3544549.35856252023.

[39] P. J. Thomas, The social and interactional dimensions of human-computer interfaces, New York: Cambridge University Press, 1995.

[40] V. G. Motti, "Wearable Technologies: a Roadmap to the Future," in *WebMedia'20: Proceedings of the Brazilian Symposium on Multimedia and the Web*, pp. 3-4, 2020, doi.org/10.1145/3428658.3431928.

[41] R. Sun, A. V. Wallop, G. Leslie and E. Y.-L. Do, "SoniSpace: Expressive Movement Interaction to Encourage Taking Up Space with the Body," in *DIS'23 Companion Publication*, 2023, pp. 279–283, doi.org/10.1145/3563703.359.

[42] M. O. Ellenberg, M. Satkowski, W. Luo and R. Dachselt, "Spatiality and Semantics - Towards Understanding Content Placement in Mixed Reality," in *CHI EA'23*2023, pp. 1–8, doi.org/10.1145/3544549.3585853.

Status quo of Digital Accessibility in Multinational Enterprises – an Exploratory Study

Andreas Deitmer
Instituto Universitário de Lisboa (ISCTE-IUL)
Business Research Unit (BRU-IUL)
Lisbon, Portugal
and
Technische Hochschule Mittelhessen (THM) -
University of Applied Sciences
BliZ - Study Centre for Blind and Visually Impaired Students (THM)
Gießen, Germany
e-mail: andreas_deitmer@iscte-iul.pt

Monika Maria Möhring
Technische Hochschule Mittelhessen (THM) -
University of Applied Sciences
BliZ - Study Centre for Blind and Visually Impaired Students (THM)
Gießen, Germany
e-mail: mmm@bliz.thm.de

J. M. Vilas-Boas da Silva
Instituto Universitário de Lisboa (ISCTE-IUL)
Business Research Unit (BRU-IUL)
Lisbon, Portugal
e-mail: jmvbs@iscte-iul.pt

Abstract— **Digital accessibility (DA) for disabled people is a hot topic, mandatory under UN and EU dispositions. Literature shows a research gap regarding relevant factors that influence the enforcement of DA in Multinationals (MNEs) (RQ1) in order to structure the problem. An exploratory study analysed the current status of DA in MNEs. Findings were consolidated by comparing them with experiences reported by FCEA conference of experts, confirming a notable deficiency in current literature regarding DA status-quo (RQ2). A normative approach, grounded in legal, legitimate, and ethical standards to steer corporate governance in matters of DA was proposed following Soft Systems Methodology.**

Keywords-Digital Accessibility in Multinational Enterprises; Exploratory study under Soft Systems Methodology (SSM); Normative approach to steer corporate governance.

I. INTRODUCTION

Digital accessibility (DA) refers to the extent to which digital products, resources, and services are available for people with disabilities [1][2][3]. Article 1 of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) states that ‘Persons with disabilities include those who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others’ [4]. Even though everyone benefits from barrier-free digital products, digital barriers mainly affect people with auditory, cognitive, physical, speech, and visual disabilities [5].

According to the World Wide Web Consortium (W3C), software, websites and mobile applications should be Perceivable, Operable, Understandable, and Robust (POUR) to be accessible for this target groups [6][7]. Technical criteria for accessible, digital content are provided within the Web Content Accessibility Guidelines (WCAG) [7]. After ratifying the UNCRPD [8], the European Union (EU) has performed a lot of activities regarding digital accessibility [9]. The Directive (EU) 2016/2102 obliges all member states of the EU, to incorporate the accessibility of the websites and mobile applications of their public sector bodies within their national legal systems. These new legislative changes have created a growing market for accessible digital products and services for public bodies, which may be an opportunity or a risk for market participants, depending on their ability to design their products and services accessible [10]. Due to the entry into force of Directive (EU) 2019/882 (‘European Accessibility Act’ (EAA)) and the resulting national legislative changes, digital products that are seen as important by the EU, have to be designed in a way that they are usable by people with disabilities [11]. Therefore, companies that manufacture such physical products and services, will also have to face increasing accessibility requirements, specifically the WCAG [7], within the next years.

However, not only companies that are obliged by law should act with accessibility in mind: the role of business enterprises has changed during the last decades and the concept of Corporate Social Responsibility (CSR) has become increasingly important [12]. While in the past, the goal of corporations has been solely profit maximization, nowadays companies are expected to have a positive impact on society and to consider social and environmental impacts in their business decisions [13]. Furthermore, by acting with CSR in mind, companies can benefit on many different areas [14]. As one aspect of CSR, DA can bring a lot of advantages, for example by driving innovation or by enhancing company's brand [15].

As described in [1] literature-based findings shows a research gap on how companies may draw on social, political & legal, organisational, and technical framework conditions in designing accessible digital products and services. Furthermore, there is also a gap in how well prepared multinational enterprises (MNEs) are, to fulfil the new DA requirements for their products and services. To close this gap, this paper presents an exploratory study about the status quo of the new digital accessibility requirements in MNEs and discusses possible solutions to increase the maturity of digital accessibility within these companies, to address the following research questions:

RQ1: What are the relevant factors that influence the enforcement of digital accessibility in MNEs?

RQ2: How well prepared are MNE to meet new digital accessibility requirements on their digital products and services?

The remaining of this paper is organized as follows. Section II describes the methodology and the exploratory case study. Section III contains the analysis of the findings. Section IV discusses these findings based on an expert conference about the German transposition of the EAA and concludes.

II. METHODOLOGY

Soft Systems Methodology (SSM) is a learning approach for tackling complex, real-world problems, positioned in organizational contexts where the problem itself is not clearly defined. The methodology emphasizes understanding the problem situation from various viewpoints without imposing a predefined structure, which is crucial during the initial stages of SSM (i.e., Stages 1 & 2). SSM is made up of seven stages, as shown in Fig. 1 [16][17].

A preliminary literature review was performed to perceive the unstructured problem situation and it showed a gap regarding the status of DA in MNEs [1]. To investigate this status, an exploratory case study within MNEs was carried out and presented in this paper. Semi-structured interviews with experts [18] were conducted between 22/11/2023 and 31/01/2024, previously harmonized by the Ethics Council at ISCTE-IUL. Appendix I includes the interview questionnaire, guided by the work of Qu and Dumay [19]. The aim was to gather as many perceptions of the research questions as possible from a diverse group of people. This helped in capturing a wide range of insights and understanding the complexity of the situation.

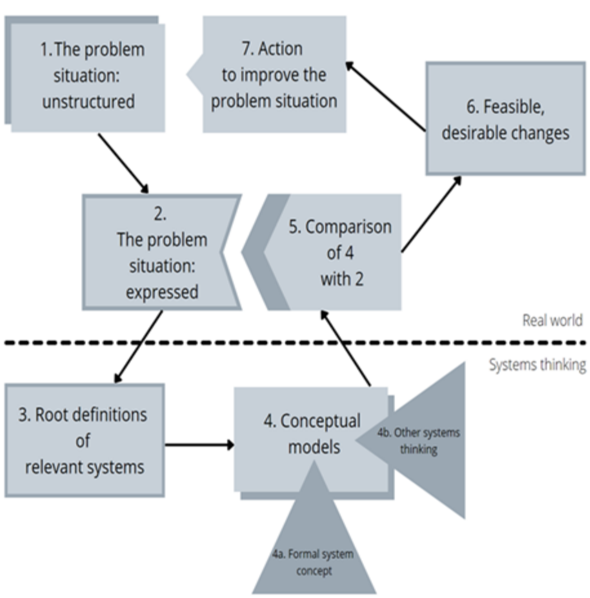

Figure 1. The seven stages model of SSM. Adapted from [16].

To address *RQ2*, the participants were asked to estimate the relevance of digital accessibility within their company, based on a Likert scale [20], which contains grades from ‘very irrelevant’ to ‘very relevant’ and the maturity of digital accessibility within their company by using the scale according to Herget (Appendix I). Maturity levels offer a good basis for systematically identifying strategies to improve the current situation and provide a basis for discussion and reflection within a company, as the attested maturity levels always correspond to the personal interpretations of the stakeholders interviewed [21]. Subsequently, a conference about the German transposition of the EAA [22], was used to discuss the findings of the exploratory study enabling to address *RQ1* pursuing a bottom up approach.

After the interviews, the data analysis was performed. In the first step, all transcriptions of the interview data and conference contributions were loaded into MAXQDA Analytics Pro [23] and text quotations, which might contain relevant data for this research were identified. After this initial analysis, relevant text quotations were identified, referenced either by in-vivo codification (codes consists of the words used by the informants) or marked up as ‘new Code’, using the expressions of the interviewees whenever it was possible [24][25]. In a next step, codes that overlap with other codes were highlighted and merged.

To address RQ1 and to attempt to structure the problem situation for Stage 2 of SSM, a sketch was created out of the findings provided by the discussion of the interview data based on the conference contributions. According to [26], Rich Pictures are preferable for the expression of relationships to the linear writing and support the holistic thinking about a situation [27]. The sketch was drawn by using Insight Maker, a free and open-source simulation and modelling application [28].

Concise statements that capture the essence of systems that appear relevant to the problem situation start emerging from the discussion of the findings. These are the root definitions (SSM Stage 3) that serve as the foundation for developing conceptual models, which are not intended to be direct representations of reality but rather will serve as intellectual tools to facilitate discussion and debate about the problem situation and potential improvements, in future. The process is iterative, with learning and adaptation occurring as new insights emerge [26][29].

III. DIGITAL ACCESSIBILITY IN MULTINATIONAL ENTERPRISES - AN EXPLORATORY CASE STUDY

The upcoming section is structured to align with the initial research questions posed in the study. It will include (a) Interview Questions: A detailed list of the questions asked to the stakeholders; (b) Stakeholder Responses: A comprehensive summary of the answers provided by the surveyed stakeholders; and (c) Preliminary Analysis: An initial interpretation and examination of the responses in relation to the research questions.

A. Selection of experts to interview in multinational enterprises

The research involved engaging with stakeholders from various enterprises listed among the 500 largest MNEs in the world, as ranked by [30]. Three individuals, including CEOs, managers, and inclusion officers from three different MNEs, agreed to participate in the interviews. The study aimed to understand whether the current state of digital accessibility (DA) in MNEs differs across sectors.

To assess the impact of new legal requirements on DA, employees from two distinct sectors were interviewed. On one hand, Technology Industry [31] is frequently impacted by new DA criteria due to the nature of its operations. The study explored the assumption that technology-related companies might have an easier time incorporating technological changes, such as DA, into their processes. On the other hand, Pharmaceutical Industry [31] is only partially affected by new DA regulations, but this sector is significantly influenced by other legal frameworks [32]. This could offer a contrasting viewpoint on how DA is implemented and perceived.

The selection of managers and inclusion officers as interviewees was intentional to gain insights from different levels of the corporate hierarchy. This approach also sought to understand the corporate structures and contexts better, particularly focusing on technological and knowledge exchanges within companies [33]. The findings from these interviews are expected to contribute to a more nuanced understanding of DA's integration into MNEs' operations.

B. Findings from interviews carried out in the exploratory case study

The first part of the interviews concerned gathering information about RQ2, i.e., How well prepared are MNE to meet new DA requirements on their digital products and services? The topics addressed were grouped into four categories, as follows.

1) Relevance and maturity of digital accessibility (DA)

The participants were asked to rate the relevance of digital accessibility within their company [20][21].

All participants within the pharmaceutical industry emphasized the relevance, i.e., the degree to which DA is applicable, and assessed it as 'neither irrelevant nor relevant'. The level of maturity of DA [34] is found to range between 'not existent', i.e., no consciousness present and 'initial', i.e., first engagement with the way of a certain behavior and activity has been made. However, DA is found as being important, i.e., of significant value. Interview partner 2 (IP02) describes it like this:

'*Nothing has happened in our company yet, but it is important to us. But we haven't dealt with it enough.*' (IP02).

In contrast, participants from the technology industry rated DA as 'relevant' and its level of maturity as close before 'managed', i.e., specifications, guidelines, sanction systems, process descriptions are almost in place, measures are close to be taken according to predefined patterns, and the company's own good practices are tending to be in place, providing the benchmark for action. IP03 indicates the situation in his company as follows:

'*@@ MNE gamma ## is currently making a huge move, i.e., a big shift towards digitalization and accessibility and I think digitalization is very, very important.*' (IP03).

The disparity in the perceived applicability (all interviewees) and value (IP02) of DA within the pharmaceutical industry, as indicated by the interviewees, does indeed raise concerns. It suggests a need for more in-depth research to understand the barriers and opportunities for DA in this sector. While IT providers are seen as more advanced in recognizing the relevance and achieving maturity in DA, there seems to be some ambiguity about whether their focus is on DA specifically or on digitalization as a whole. Clarifying this distinction is crucial for developing targeted strategies that enhance DA without conflating it with broader digital transformation efforts (IP03).

2) Knowledge of DA regulations

Since MNEs are affected by new legal requirements regarding digital accessibility, e.g., [11], the participants were asked about their knowledge of legal requirements for DA that apply to their company, as well as about corresponding internal guidelines within their company. All interviewees from the pharmaceutical industry stated, that they were not aware of any legal requirements about DA and that no DA policies had been established in their companies. The preconditions apparently differ in the technology industry. In particular, IP03 stated that he was actively involved in the development of international standards (IP03):

'Yes, [..]. I know the EU guidelines, I know pretty much everything about the American market. In part, I am familiar with the Japanese market, [..]. The interesting thing is [..] the central standard for us is EN 301549, which I have been working on since 2012 and which is now entering the next round.' (IP03).

IP03 also stresses the importance of internal policies and monitoring DA activities by a central unit:

'Since 2018 we have also anchored this in the so-called inclusion agreement stating that all IT services for @@ MNE gamma ## should be accessible. No, have to [be digital accessible]. So, this requirement is directly included in the inclusion agreement. It is also one of our main drivers that a large part of the internal software solutions come across the table here in the Competence centre. This means that internal IT programmes are first checked for accessibility and only then they are released.' (IP03).

A suggestion to include stakeholders from civil society in the discussion on DA appears to be insightful. It would provide a more comprehensive understanding of how legal requirements are perceived by those directly affected versus their potential to create business opportunities for IT providers.

Additionally, exploring the internal policies that diverge from legal requirements, which have been scarcely mentioned, could reveal valuable insights into the internal drivers and barriers to DA implementation. This broader approach to stakeholder engagement could help balance the perspectives between regulatory compliance and the practical, lived experience of DA among end-users.

3) Possible degree of DA within different departments

IP03 also stressed out the need for his MNE to decide, which degree of digital accessibility in which working environment made sense or even would be possible:

'Internally, we have the problem that we run a lot of digital solutions that cannot actually fulfil within the scope of accessibility criteria. So, we have 360,000 employees, of which well over half work in the production area. We maintain a lot of IT in the area of production control and when I'm in the rolling mill and in production control, then I no longer have to worry so much about accessibility. This means that one of our main problems at the moment is: when and to what extent does software have to be accessible? Where is it 'worth it' and where should you keep your hands off?' (IP03).

This is about developing requirements to express DA business policy (IP03) regarding when, to what extent, if it worths it and for whom, i.e., which disabilities [5].

4) Action taken to increase DA, implementation and effectiveness

Afterwards, the participants were asked about the existing actions to increase digital accessibility in their companies. As one example in the pharmaceutical industry was given by interview partner 1 (IP01) is the accessibility of the intranet according to the WCAG [7] in their company. The participant from the technology industry (IP03) reported on a wide range of actions to anchor digital accessibility within his company:

'I always say our strategic approach is Accessibility by Design. Construction would be building-related to quote an example, we have about 15,000 document templates in the company. We started with this and said these 15,000 document templates, let it be PowerPoint, let it be oh... all kinds of stuff. They are at first built so that they are accessible a priori. This means that if people use this, then they can first use a basic level of accessibility. So, the same applies [..].

We use so-called user interface libraries and style guides for how our user interfaces look. It is very important that we simply work together to ensure that these things work as they should.' (IP03).

The feedback from the interviewees suggests that the current actions implemented for DA in the pharmaceutics are not fully effective. For instance, IP01's observation indicates the need for physical checks for intranet and application accessibility. The mention of 'Accessibility by Design' in document templates is also noted as being too ambiguous. Furthermore, while the development of user interface libraries is a positive step (IP03), it lacks specific guidelines for different disabilities. Finally, the absence of a clear measure of effectiveness in these initiatives is a critical oversight.

To address these shortcomings, it is recommended that regular follow-ups are executed, detailed guidelines for each type of disability are established, clear effectiveness metrics of DA actions are introduced. The second part of the interviews concerned gathering information about RQ1, i.e., *What are the relevant factors that influence the enforcement of digital accessibility in MNEs?* The topics addressed were grouped into two categories (a) & (b), as follows.

a) Status quo of digital accessibility

Participants were asked to describe the status-quo of DA within their respective MNE, guided by specific keywords. The findings are depicted and summarised in Fig. 3, a sketch about the *status quo* of DA in all analysed MNEs. For instance, IP03 describes the stakeholder involvement in his company like this:

'The topic of accessibility has actually been promoted at @@ MNE gamma ## by the representative office for disabled employees [..] So, in management, I'm more familiar with diversity and inclusion, and you have to make sure that you're working under this umbrella. The topic of accessibility has been strongly addressed in the IT department; [..] it has not yet been addressed in the area of procurement and it has only been addressed to a limited extent in the area of tender development.' (IP03).

Overall, while there is some level of awareness and targeted implementation of DA, specially in the IT department, a holistic strategy is essential for achieving full maturity in DA practices, e.g., stakeholder involvement. IP03 points out the disparity in DA efforts, underscoring the potential for enhanced and more evenly distributed initiatives throughout the organization.

b) Factors influencing the maturity of digital accessibility

Finally, the interviewees were asked to name further important factors from their perspective that would be capable of influencing the maturity of digital accessibility [34] within their companies. The following factors were mentioned. Interview partner 1 (IP01) summarizes this for MNE alpha as follows:

'*It [the pharmaceutical industry] is very strongly regulated. And I think, that this strong regulations in other business areas sometimes prevent the implementation of digital accessibility or even digitalisation.'* (IP01).

The maturity of DA [34] is indeed shaped by multiple factors, with the regulatory environment being a significant one. As per the insights from Interview Partner 1 (IP01), the pharmaceutical industry's dense legal framework can sometimes serve as a barrier, fostering a risk-averse mindset that emphasizes compliance over innovation. This cautious approach prioritizes the identification and prevention of potential DA issues. However, IP01 also implies that while compliance is necessary, it should not stifle progress.

For DA to truly advance, it must be supported by flexible and proactive strategies that allow for innovation within the bounds of regulation.

IV. Discussion and conclusions

The exploratory study on DA in MNEs acknowledges the increasing significance of DA for both public sector organizations and private businesses. As a result, MNEs are now confronted with evolving legal and social mandates related to DA. The study reveals that there is a notable deficiency in current literature regarding the actual state of DA in MNEs. To bridge this gap, the study was designed to examine how MNEs currently align with some DA requirements and to structure the identified challenges associated with DA implementation. This research is pivotal in shaping a comprehensive understanding of DA's integration into MNE operations and guiding future improvements.

A. Cross-checking the results of the exploratory study

The German transposition of the European Accessibility Act (EAA) provides a valuable reference point for cross-checking the results of the exploratory study on DA with practical insights from companies actively working towards implementing DA in their processes.

The discussion at the conference, as reported by [22], aimed to consolidate the findings of the exploratory study by comparing them with the experiences of companies that are navigating the requirements of the EAA, as follows:

1) Digital accessibility as an overall-process

All presenters confirmed the message that DA was an interdepartmental task. This is also regulated in international process norms regarding digital accessibility, as conference participant 1 (CP01) stressed out:

'The implementation of the processes described in 'DIN EN 17161 Design for all' must take place at all organisational levels in the company in order to be anchored sustainable and independently of individuals and organisational units.' (CP01).

CP02 mentions the importance of a systemic approach to implementing digital accessibility into company-wide processes:

'What is also another challenge is the structuring, to bring digital accessibility in the software development process. So, everyone who works in large corporations knows how complex the processes are and there is not just one person who decides 'now we'll make it barrier-free', but rather we need a system to get it down to the bottom to pass it on to the operational level and that is a huge challenge.' (CP02).

Conference participant 3 (CP03) highlights, that to increase the positive customer experience, digital accessibility must be implemented along the complete supply chain:

'We're not just talking about making these channels barrier-free, but actually the entire process from searching and buying through to delivery, [..]. So it's about [making all the processes and the entire chain] here accessible.' (CP03).

The review emphasizes a systemic and holistic approach to DA in MNEs (CP01), highlighting the importance of integrating DA into both intra-firm (CP02) and inter-firm processes (CP03). It identifies Organisational Structuring, Supply Chain Management, and Information Technology as key areas for DA readiness as Root Definitions (RDs).

The review also points out that Customer Service and Stakeholder Theories are crucial for understanding the customer's role in DA, aligning with the trend of customer engagement as a business priority as RDs. Overall, the review suggests that MNEs should adopt a strategic, inclusive, and customer-centric approach to DA, which is supported by contemporary business practices and theories [35].

2) Organisational anchoring of DA in multinational enterprises

It is also noteworthy that all of the presenting companies maintain a central office to monitor and control digital accessibility activities (CP01):

'The strategic management level with the company management, inclusion officers and ideally a Chief Accessibility Officer ICT, the so-called CAO, is significantly responsible for assigning roles, responsibilities and authorisations, in addition to the development and communication of visions, policies and goals.' (CP01).

CP04 *explains* their corporate concept for accessibility in their branch as follows:

'We at @@ MNE eta ## are organised into six divisions [..] and all [..]report on their activities to a central accessibility and human-centred design team [..].' (CP04).

CP04 also identifies four key areas for the company-wide cultural integration of digital accessibility:

'We are concentrating on four areas here, i.e., standardisation and regulation. Then there is external relations and internal promotion, which is also very important. And then, in principle, the whole thing is included in the sustainability area, which also includes inclusion, accessibility, diversity and so on.' (CP04).

Fig. 2 illustrates these four key domains of digital accessibility from the perspective of the MNE CP04.

Figure 2. Four key focus areas of accessibility of a MNE in the industrial sector. Adapted from [22].

The study identifies the role of the CAO (CP01) in strategic management, sustainability practices, and Society 5.0 principles as potential RDs for research framework. Society 5.0 is a concept that envisions integrating technology to solve societal issues and promote development [36]. Additionally, the design of IT systems that leverage external knowledge is crucial, highlighting the importance of Absorptive Capacity - the ability of an organization to utilize external knowledge. These components are suggested as RDs to evaluate DA maturity in MNEs and to inform strategy development.

3) International and harmonised standards as a key for the enforcement of DA

Several speakers (e.g., CP03) highlighted the need for internationally harmonized criteria for digital accessibility (CP03):

'For us, it is important that there are standardised requirements in the EU. It makes no sense that there are different markets within the EU, that there are different criteria for different countries, including for me as a person with a disability. My disability doesn't change when I move from one country to another.' (CP03).

If a company or product is affected by several rules, they must be compatible, as explained by CP04:

'*On the one hand, of course, we have the regulation, which in Germany actually primarily comprises the Interstate Broadcasting Treaty. Then, there is the content side, we have the linear broadcast providers, there are the network service providers, then to name also the platforms, the producers and the users. And that is not quite so easy to implement, in particular, if some groups are excluded from the regulatory side, while the device side then gets the regulation, because without content, this television cannot display anything'* (CP04).

Also the understanding and interpretation of the legal regulations (CP03) is seen as a requirement. Harmonizing and standardizing regulatory frameworks across the EU (CP03 & CP04) is complex due to the involvement of many stakeholders from various countries. To address this, developing organizational capabilities in absorptive capacity is essential. These capabilities are key to ensuring consistent interpretation and application of knowledge, both internally and from external sources. Stakeholders within the organization, civil society, and mechanisms of standardization and regulation are crucial. They serve as potential RDs and play a significant role in the maturity of DA.

4) Actions taken to increase digital accessibility

Actions taken by the MNEs of the conference participants, to increase DA were introduced. As already mentioned within the interviews, E-learning programs seem to be an appropriated action to increase knowledge of DA, as confirmed by CP04:

We have e-learning programmes that were already introduced in 2018 in Japan for all colleagues. These [programmes] are also being improved every year [..].' (CP04).

CP04 also highlighted the importance of close networking between internal and external stakeholders to alight DA from as many perspectives as possible:

'Then we set up an Accessibility Champions Network, where of course, individual colleagues are not obliged to participate, but we now have a large group who are connected to this network either because they are personally affected or because of product development issues and so on. We exchange ideas and then we also work together with the disability organisations for product tests and much more.' (CP04).

The conference proceedings underscore the gaps in DA practices revealed by the case study, suggesting a more in-depth physical examination is necessary. They stress the value of creating networks that include all stakeholders to address these issues effectively (CP04). Findings reinforce the necessity for precise DA requirements and acknowledge the significance of Absorptive Capacity, e.g., in product development. This supports the premise that DA considerations should be integral to the design process from the outset.

i) Preliminary graphical sketch of the DA *status quo*

The interviews have yielded a comprehensive view of DA in MNEs, identifying key factors that affect DA enforcement. A preliminary sketch, Fig. 3, visualizes these factors and their interrelations. Next steps involve refining this sketch into a 'rich picture' for clarity and synthesizing the insights to present a detailed and nuanced understanding of DA's current implementation, pinpointing both strengths and enhancement areas.

B. Outcomes of the study

1) Research question RQ2

To address Research Question 2 (RQ2), which examines the readiness of MNEs to meet new digital accessibility standards, it was observed that companies in technology sectors and those offering digital products and services are more advanced in integrating digital accessibility into their processes. This observation is consistent across all participants and presenters, who acknowledged the significance of digital accessibility from their perspectives. However, there is often ambiguity in responses regarding whether the relevance and maturity discussed pertain to Digital Accessibility (DA) or to digitalization in general, which underscores the concept of 'DA by design'.

It is essential for both external and internal stakeholders to be systematically engaged to establish comprehensive social requirements that supplement legal mandates. The scarcity of robust examples of implemented actions suggests a lack of effective guidance, potentially due to inadequate involvement of associations, insufficient absorption of external knowledge, and limited internal engagement. Furthermore, business policies need to articulate specific disabilities to be addressed, which will focus research and development efforts on enhancing digital accessibility.

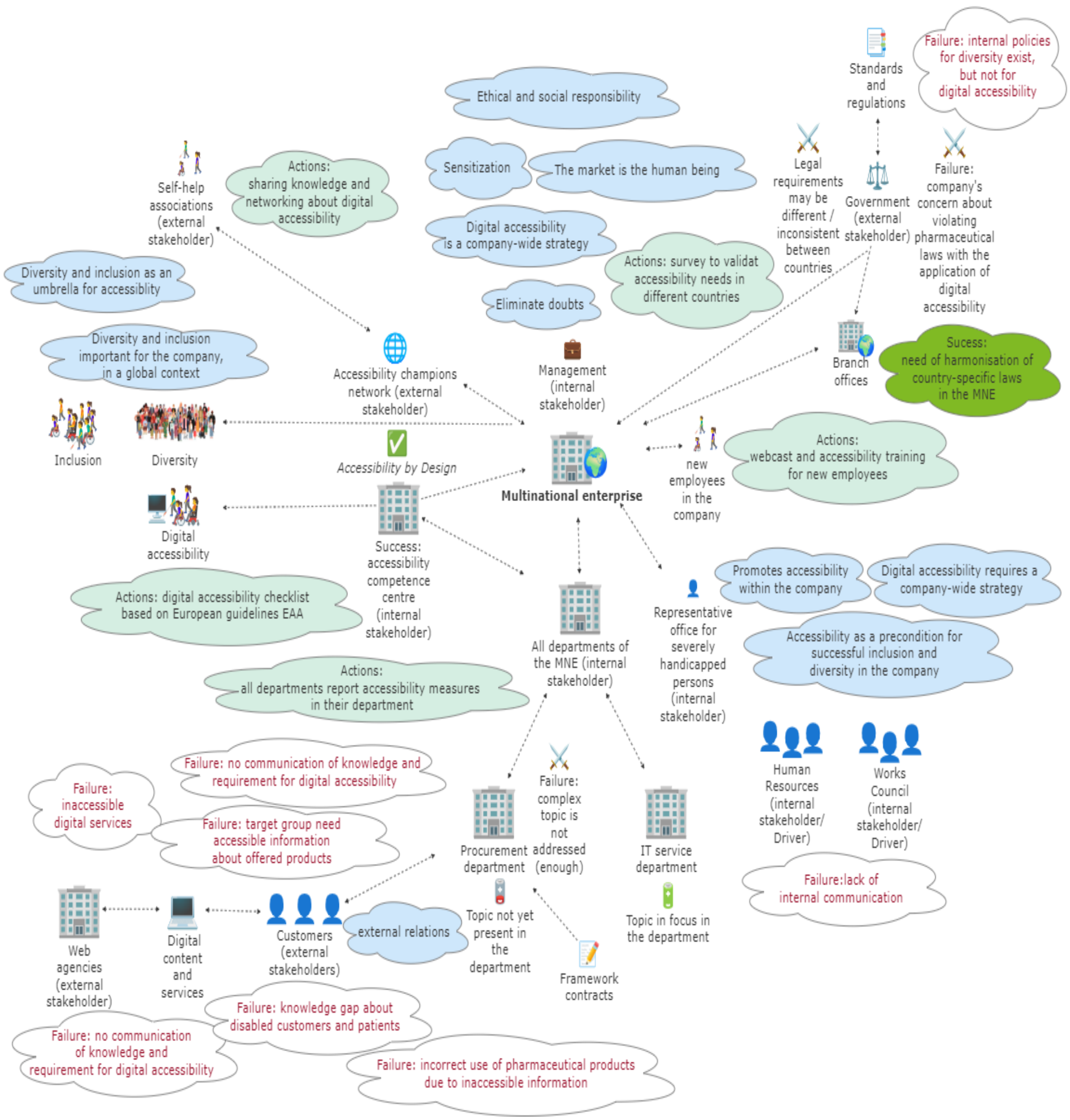

Figure 3. The first comprehensive ongoing holistic sketch of the status quo of DA for MNEs, synthesized from interviews and perspectives presented at the EAA conference [22].

2) Research question RQ1

To address the first research question on the factors affecting the implementation of Digital Accessibility (DA) in Multinational Enterprises (MNEs), we identified a notable gap. This gap, evident from Fig. 3, the discussion exercise, and the analysis of findings related to RQ2, points to the absence of an initial guiding framework to structure the problem-situation. We propose a normative approach, grounded in established legal, legitimate, and ethical standards [37][38], to steer corporate governance in matters of DA. Such an approach would gain legitimacy through compliance with legal and regulatory mandates that foster diversity and inclusion, echoing the contemporary concerns of Society 5.0 and the United Nations SDGs. Therefore, it would be subject to the principles of ethics and social responsibility, ensuring that corporate decisions align with these broader societal values. To establish a consequent systematic and holistic strategic management process that is embedded within the company's structure, the following principles should be adhered to:

- Respect *ethics and corporate social responsibility*; ensuring that all company activities align with ethical standards and contribute positively to society.
- Integrate relevant *externalities from stakeholders*, including (a) civil society knowledge from associations, networks of individuals, experts, etc., (b) technical knowledge encompassing technology and IT advancements, (c) legal regulatory compliance, e.g., EAA, (d) governmental policies that may impact the company's operations, (e) advocacy and lobbying efforts that align with the company's interests, (f) standardization and harmonization to ensure consistency and quality.
- Promote the *transference, assimilation, and application of knowledge* within the company's processes through absorptive capacity.
- Develop *organizational capabilities that structure the organization* by (a) encouraging customer engagement and fostering a customer-centric approach and, (b) focusing on DA by design, ensuring external/internal requirements alignment.

C. Further work

As we advanced in structuring the research problem, it became evident that various knowledge domains needed to be incorporated into the formulation of the root definitions for a conceptual framework for MNEs to implement digital accessibility within their company-wide processes, which is the third stage of Soft Systems Methodology (SSM). It is suggested that an intermediate literature review is pursued to help a more robust structuring of questions to support a more credible exploratory study to definitely establish the root definitions. In qualitative research, the number of interviewees is not fixed and can indeed vary widely. The key is to reach a point of data saturation, where additional interviews do not yield new insights or themes. Some researchers suggest that around 20 interviews may be a good starting point for a robust study, but this number is flexible and depends on the research objectives and richness of data collected to capture a diverse range of perspectives and to provide a deep understanding of the phenomenon.

D. Expected future contributions

Firstly, theoretical development of a validated conceptual model that is well-aligned with the principles of Soft Systems Methodology (SSM). A solid foundation for stages 4a and 4b of SSM will result from incorporating propositions, relationships, dimensions, and measures, and representing these graphically. This model will serve as a critical research tool for assessing the current state of digital accessibility within MNEs. Therefore, the research also offers practical benefits to MNEs striving to improve their DA practices.

NOTE

The first version of this article was published in the EurOMA 2024 conference proceedings and will only be made available to conference participants. The proceedings will not be indexed, so there are no conflicts with the publication in this IARIA Journal [39].

ACKNOWLEDGMENT

We acknowledgment all the companies and interview partners who participated voluntarily and contributed to this research on inclusiveness. An esteemed thanks to both universities, Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL) and Technische Hochschule Mittelhessen (THM) - University of Applied Sciences, which are an important basis for this international, inclusive research. A special acknowledgment to the entire IARIA committee for the invitation to publish this article in the journal, followed by the conference contribution to the special track about DA at the IARIA conference in 2023.

APPENDIX I

A. Exploratory study: Digital Accessibility in MNEs

- From your personal perspective, how would you rate the relevance of digital accessibility in your company?
  - Very relevant
  - Rather relevant
  - Neither relevant or irrelevant
  - Rather irrelevant
  - Very irrelevant
- How would you rate the maturity level of digital accessibility within your company?
  - Level 0: Non-existent - i.e., no consciousness present
  - Level 1: Initial - i.e., first engagement with the way of a certain behavior and activity has been made. Analyses, process descriptions, policies are evaluated for their suitability and initial steps are taken to implement them.

    - Level 2: Managed - i.e., Specifications, guidelines, sanction systems, process descriptions for handling business transactions are in place, measures are taken according to predefined patterns, and the company's own good practices are in place, providing the benchmark for action.
    - Level 3: Defined - i.e., in addition to clearly regulated specifications and business processes, responsibilities and exceptions are fixed, benchmarks and targets have been developed, and continuous improvement in task execution and collaboration is targeted.
    - Level 4: Quantitatively Managed - i.e., the business processes are systematically evaluated and compared with the targets, deviation analyses are carried out, and optimizing measures are taken and checked for their effect. Reporting on deviations and target achievement has been introduced and forms the starting point for continuous optimization.
    - Level 5: Optimizing - i.e., here is a permanent orientation towards best practices within and outside the company, and all business processes are permanently evaluated and optimized.
- If you are aware of legal requirements for digital accessibility, like EU Directive 2019/882 (European Accessibility Act) [11], that affect your company, please name them.
- If you are aware of guidelines (policy: mandatory guidance of action) for digital accessibility within your company, please name them.
- If measures are currently being carried out in your company to increase digital accessibility in your company: How are these implemented in day-to-day business and how is their effectiveness monitored?
- In your opinion, which stakeholders, (environmental) factors, like economic systems, social norms and values, interests/goals or framework conditions, like policies (policy: mandatory guidance of action), influence the implementation of digital accessibility in your company? Key words for this question: drivers, stakeholders, responsibilities, interests/goals, policies, and legal framework. Note to this question: This is an accompanying process to develop a rich picture to better reflect the question in your organization. For demonstration purposes and shows an example of a Rich Picture, with the question of what aspects influence the implementation of guidelines in a hospital, based on [40]. The picture is a simplification of the real world. For this research, the illustration is created by the interviewer using Insight Maker, a free and open-source simulation and modelling application.
- Do you have any other questions or suggestions for this research?

References

[1] A. Deitmer, M. M. Möhring, and J. Vilas-Boas da Silva, ‘Digital accessibility in multinational enterprises: A meta study’. In M. M. Möhring (Ed.), SMART ACCESSIBILITY 2023 : The Eighth International Conference on Universal Accessibility in the Internet of Things and Smart Environments IARIA, 2023, pp. 1-5, ISBN: 978-1-68558-084-1, Available from: http://hdl.handle.net/10071/28927 2024.05.12

[2] J. E. Hellbusch and K. Probiesch, Barrierefreiheit verstehen und umsetzen: Webstandards für ein zugängliches und nutzbares Internet. Translation in English: Understanding and implementing accessibility: Web Standards for an Accessible and Usable Internet, 1st ed., dpunkt.verlag, Heidelberg, 2011.

[3] M. Kulkarni, ‘Digital accessibility: Challenges and opportunities’, IIMB Management Review, vol. 31, no. 1, pp. 91-98, Mar. 2019, doi:10.1016/j.iimb.2018.05.009.

[4] United Nations (UN). *Convention on the Rights of Persons with Disabilities: CRPD*. [Online]. Available from: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html 2024.05.12

[5] W3C. *Diverse Abilities and Barriers*. [Online]. Available from: https://www.w3.org/WAI/people-use-web/abilities-barriers 2024.05.12

[6] Information Resources Management Association (Ed.), Accessibility and Diversity in Education: Breakthroughs in research and practice, IGI Global, Hershey, Pennsylvania, 2020.

[7] W3C. *Web Content Accessibility Guidelines (WCAG) 2.2*. [Online]. Available from: https://www.w3.org/TR/WCAG22 2024.05.10

[8] United Nations (UN). *Convention on the Rights of Persons with Disabilities: CRPD*. [Online]. Available from: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html 2024.05.10

[9] D. Ferri and S. Favalli, ‘Web Accessibility for People with Disabilities in the European Union: Paving the Road to Social Inclusion‘, Societies, vol. 8, no. 2, p. 40, 2018, doi:10.3390/soc8020040.

[10] The European Parliament and the Council of the European Union. *Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies (Text with EEA relevance)*. [Online]. Available from: https://eur-lex.europa.eu/eli/dir/2016/2102/oj 2024.05.10

[11] The European Parliament and the Council of the European Union. *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (Text with EEA relevance)*. [Online]. Available from: https://eur-lex.europa.eu/eli/dir/2019/882/oj 2024.05.10

[12] P. Katerattanakul, S. Hong, H.-M. Lee, and, H.-J. Kam, ‘The effects of web accessibility certification on the perception of companies’ corporate social responsibility‘, Univ Access Inf Soc, vol. 17, issue 1, pp. 161-173, 2018, doi:10.1007/s10209-017-0532-1.

[13] T. Stobierski. *Types of Corporate Social Responsibility to Be Aware Of*. [Online]. Available from: https://online.hbs.edu/blog/post/types-of-corporate-social-responsibility 2024.05.10

[14] *European Commission (EC)*. Corporate social responsibility & Responsible business conduct. [Online]. Available from: https://ec.europa.eu/growth/industry/sustainability/corporate-social-responsibility-responsible-business-conduct_en 2024.05.10

[15] W3C, *The Business Case for Digital Accessibility*. [Online]. Available from: https://www.w3.org/WAI/business-case 2024.05.10
[16] P. Checkland, Systems thinking, systems practice, John Wiley & Sons Ltd, p. 164, 2008.
[17] J. Manjate, J. and J. Vilas-Boas, Aplicação da Soft Systems Methodology em Investigação Qualitativa - Teoria e Prática, Translation in English: Applying Soft Systems Methodology in Qualitative Research - Theory and Practice. Lisboa: Edições Silabo, p. 5, 2021.
[18] M. Ebrahimi, Applications of soft systems methodology for organizational change. (Premier reference source). Hershey, PA, USA: IGI Global, Business Science Reference, p. 136, 2020.
[19] S. Q. Qu and J. Dumay, The qualitative research interview. Qualitative Research in Accounting & Management, vol. 8, pp. 238-264, 2011. doi:10.1108/11766091111162070.
[20] J. F. Hair, M. Wolfinbarger, A. G. Money, P. Samouel, and, M. I. Page, Essentials of Business Research Methods. Routledge. p. 245, 2019. ISBN: 9780429511950.
[21] J. Herget, Unternehmenskultur im Reifegrad-Modell, English translation: Corporate culture in the maturity model. Berlin, Heidelberg: Springer Gabler, pp. 83-97, 2020.
[22] Federal Centre of Expertise on Accessibility (FCEA), Bitkom and DIN. *Review of the conference on the Act of the Directive (EU) 2019/882 of the European Parliament and of the Council on the accessibility requirements for products and services", Conference* '*Barrierefreiheitsstärkungsgesetz: Hintergründe, Umsetzung und Praxisbeispiele*', 2022, Berlin. Available at: https://bit.ly/conference-accessibility_berlin-2022 2024.05.10
[23] U. Kuckartz and S. Rädiker, *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. *Springer eBooks Social Sciences*. Springer, 2019. doi: 10.1007/978-3-030-15671-8
[24] U. Flick, Triangulation: Eine Einführung, English translation: Triangulation: An introduction, 3rd ed., Wiesbaden: VS Verlag für Sozialwissenschaften | Springer Fachmedien Wiesbaden GmbH, p. 105, 2011.
[25] D. A. Gioia, K. G. Corley and, A. L. Hamilton, Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology (Organizational Research Methods), pp. 15-31, January 2013.
[26] P. Checkland, Systems thinking, systems practice, John Wiley & Sons Ltd, 2008.
[27] M. Ebrahimi, Applications of soft systems methodology for organizational change. Hershey, PA, USA: IGI Global, Business Science Reference, pp. 11-12, 2020.
[28] S. Fortmann-Roe, Insight Maker: A general-purpose tool for web-based modeling & simulation. Simulation Modelling Practice and Theory, vol. 47, pp. 28-45, 2014. doi: 10.1016/j.simpat.2014.03.013.
[29] J. Manjate and J. Vilas-Boas, Aplicação da Soft Systems Methodology em Investigação Qualitativa - Teoria e Prática. Lisboa: Edições Silabo, 2021.
[30] Forbes. *The Global 2000*. [Online]. Available from: https://www.forbes.com/lists/global2000 2024.06.01
[31] United Nations, International Standard Industrial Classification of All Economic Activities, Revision 4. New York: UNITED NATIONS PUBLICATION, 2008, ISBN: 978-92-1-161518-0.
[32] I. Heikkinen, S. Eskola, V. Acha, A. Morrison, C. Walker, C. Weil, A. Bril, M. Wegner, T. Metcalfe, S.-D. Chibout, M. Chlebus, Role of innovation in pharmaceutical regulation: A proposal for principles to evaluate EU General Pharmaceutical Legislation from the innovator perspective, Drug Discovery Today, vol. 28, issue 5, 2023, 103526, ISSN 1359-6446, doi: 10.1016/j.drudis.2023.103526.
[33] M. M. Möhring, Innovation in a High Technology B2B Context: Exploring Supply Networks, Processes and Management, ed. 2014. Springer Fachmedien Wiesbaden, 2014. doi: 10.1007/978-3-658-05721-3.
[34] WAI (Web Accessibility Initiative). *Accessibility Maturity Model*. Available at: https://www.w3.org/TR/maturity-model 2025.05.10
[35] S. Vivek, S. Beatty, and, R. Morgan, Customer Engagement: Exploring Customer Relationships Beyond Purchase. The Journal of Marketing Theory and Practice. Vol. 20. pp. 127-145. 2012, doi: 10.2307/23243811.
[36] M. Fukuyama, 'Society 5.0: Aiming for a New Human-Centered Society', Japan Spotlight, pp. 47-50, 2018.
[37] C. Hinings, and P. Tolbert, *Organizational Institutionalism and Sociology: A Reflection. The Sage Handbook of Organizational Institutionalism*. pp. 473-490, 2008. doi: 10.4135/9781849200387.n20.
[38] W. R. Scott, 'W. Richard Scott, Institutions and Organizations. Ideas, Interests and Identities', M@n@gement, vol. 17, no. 2, p. 136, 1995.
[39] EurOMA 2024. *Call for papers*. [Online]. Available from: https://euroma2024.org/call-for-papers 2024.06.01
[40] H. Augustsson, K. Churruca, and, J. Braithwaite, Re-energising the way we manage change in healthcare: the case for soft systems methodology and its application to evidence-based practice", BMC Health Services Research, vol. 19, p. 4, 2019. doi: 10.1186/s12913-019-4508-0.