UiT The Arctic University of Norway

Faculty of Science and Technology

Department of Technology and Safety

Methods for enhanced learning using wearable technologies

A study of the maritime sector

Hui Xue

A dissertation for the degree of Philosophiae Doctor

March 2023

METHODS FOR ENHANCED LEARNING USING WEARABLE TECHNOLOGIES

(A STUDY OF THE MARITIME SECTOR)

Hui Xue

A dissertation submitted to the Faculty of Science and Technology at 'UiT The Arctic University of Norway' in partial fulfilment of the requirements for the degree of Philosophiae Doctor in the Department of Technology and Safety.

Tromsø, Norway

2023

Supervised by:

Bjørn-Morten Batalden

Dilip K. Prasad

Ove T. Gudmestad

Foreword

I have prepared this dissertation in partial fulfilment of the requirements for the degree of Philosophiae Doctor at the Faculty of Natural Sciences and Technology (NT-fak), UiT The Arctic University of Norway. The research presented in this thesis has been conducted in the Department of Technology and Safety (ITS).

After completing MSc degree in Computer Science at UiT The Arctic University of Norway, I received a job offer as an IT consultant and two PhD offers from two different Universities. I found it was easy to accept the PhD offer at UiT because the university provided the opportunity to pursue research in natural science, which is my area of expertise, and acquiring a doctorate was my long-sought goal since childhood. However, my PhD journey was fraught with unexpected challenges.

Despite being aware of the inherent challenges and uncertainties commonly known about pursuing a PhD degree, new experiences in this journey continued to surprise and challenge me. The project I had initially planned ended much earlier than expected, leaving insufficient data for further research. As a result, I had to redirect my research. In addition, the COVID-19 pandemic forced the maritime training simulation to shut down for 10 months, making it impossible to conduct experiments and collect data. Furthermore, upon returning from parental leave, I faced a significant challenge in collecting data due to the ongoing impact of the COVID-19 pandemic on my workplace. Despite these obstacles, I persevered and invested a considerable amount of extra time into learning about maritime operations, with the aim of becoming a true professional in this field.

The journey from the first step to the very end of my PhD has been a unique experience in my life. However, such difficult circumstances could never make me doubt my resolve and commitment to continue. The support I received from the department and my new supervision team helped me overcome these obstacles, and the entire experience taught me some valuable lessons that I will carry with me throughout my career. I have developed self-adjustment abilities that have strengthened my potential for future challenges. As I approach the end of my PhD journey, I retrospect my time as a PhD candidate in the Department of Technology and Safety with gratitude and pride.

Tromsø, Norway

March 2023

Hui Xue (Helene)

Abstract

Maritime safety is a critical concern, as any navigation error or unsafe act can result in serious consequences or accidents for the crew, passengers, the environment and assets. Officers' competence plays a crucial role in avoiding these consequences or accidents, emphasising the need of proper training essential. However, the complex and rapidly evolving maritime industry challenges traditional training methods, such as classroom instruction, on-board training and simulation-based exercises. In response to these challenges, research interest has been growing in innovative training methods in maritime education and training, such as the use of virtual reality (VR) technology; augmented reality (AR) glasses; multi-sensor fusion training and mobile, collaborative and blended learning. However, further research is needed to assess the impact of these methods and establish best practices for designing and implementing future training programmes. Of particular concern is the seafarers' stress level experienced during maritime operations, with the potential for failure. To ensure that maritime training programmes remain relevant in the face of technological advancements and changing operational requirements, it is vital to continuously evaluate and develop new methods that effectively prepare maritime professionals for the potential challenges.

In this study, we explored the use of wearable sensors with biosignal data collection as a means of improving training performance in the maritime sector. With the deepening of the study, three experiments were conducted progressively with different purposes. The first experiment focused on situation awareness (SA) and the relationship between experience levels and biosignal data. The second experiment aimed to determine the effects of different training methods on cognitive workload, stress levels and decision-making skills in a complex scenario. For this, a complex towing operation was used in this experiment for the study content. The third experiment focused on the classification of scenario complexity and the stress levels of seafarers under different scenarios and the impact of stress on training performance. During the experiments, data were collected in three categories: (1) questionnaire data on the stress levels, workload and user satisfaction of auxiliary training equipment; (2) performance evaluation data on SA, decision-making, navigation and ship-handling abilities and (3) biosignal data, including electrodermal activity (EDA), body temperature, blood volume pulse (BVP), inter-beat interval (IBI) and heart rate (HR). In addition, data collected during the EU-funded Wearable Experience for Knowledge Intensive Training (WEKIT) project were analysed to investigate the potential of wearable technology, such as AR technology, in hands-on training and evaluate user satisfaction. Several statistical methods and machine-learning algorithms were used in the data analysis.

The present dissertation contributes to the advancement of the field of maritime education and training by exploring methods for enhancing learning in complex situations. Through a series of empirical studies, this research advances our understanding of the application of cutting-edge technologies, such as AR and wearable sensors, in maritime training. The utilisation of wearable sensors to collect biosignal data provides insights into the interplay between stress levels and training outcomes in the maritime industry. The results of this research emphasise the significance of incorporating biosignal data into maritime training programmes, because biosignal data have been shown to effectively gauge stress levels, classify the complexity of training scenarios and determine the seafarers' level of experience, which in turn has a significant influence on both safety and performance outcomes. To facilitate future research in this area, this dissertation proposes a conceptual training model that underscores the relationship between stress and safety factors and offers a framework for the development and evaluation of advanced biosignal data-based training systems.

Acknowledgements

"天行健，君子以自强不息" (tiān xíng jiàn, jūn zǐ yǐ zì qiáng bù xī)

– This saying, originating from a passage in the Book of Changes (I Ching), conveys the message that just as the heavens remain resolute and active, a noble person should also strive for continuous self-improvement. This quote emphasises the importance of perseverance and consistent effort in achieving success. As I reflect on my own journey towards obtaining a doctorate degree, I am reminded of the truth in this saying. However, I must acknowledge that without the unwavering support of my supervisors, friends, family and UiT, all of my hard work and dedication would have been in vain. Earning a doctoral degree holds immense significance for me, as it marks the culmination of a difficult 5-year journey full of challenges and obstacles. I extend my heartfelt and deepest gratitude to each and every one of you who has supported me, believed in me and stood by my side throughout this arduous journey.

While I had been working towards my goal for some time, it wasn't until my new supervision team was established that my true journey began, and I was able to make significant progress towards my objectives. It was a difficult time, as the COVID-19 pandemic was in full swing and the university was struggling to find ways to get everything back to normal. I had just returned from my parental leave when I received the news that my main supervisor and one of my co-supervisors were quitting the team. With only two published papers and no clear storyline in my research, I was at a loss, and the pandemic had made it nearly impossible to collect useful data for analysis. During this challenging time, Bjørn-Morten, who had been my previous co-supervisor, stepped up as my main supervisor. He, along with the administrative staff Gunn-Helene and Trine from our Department of Technology and Safety (ITS) provided me with invaluable support. Bjørn-Morten brought in Professor Gudmestad, an experienced professor, to join the new supervision team, and I am grateful for his guidance and mentorship. Despite the setbacks, my co-supervisor, Dilip, remained steadfast in his belief in me and chose to stay on the supervision team. I am grateful for his unwavering support and encouragement, which gave me the motivation and strength to keep pushing forward.

Before the pandemic, I gradually shifted my research focus to maritime training, which was the strongest area of Bjørn-Morten's research field. He helped me discover my interests, conducted a pilot study with me and gathered the first samples of data. I was amazed by his effective working style. Despite being the head of the department and being very busy, he always made time for me. I could easily catch him in person or reach out to him via call or message. Bjørn-Morten introduced me to the nautical science research group, which felt like home to me at UiT. I received a lot of assistance from the members of this group. I am grateful to Bjørn-Morten for being my main supervisor, leading such an excellent department and ensuring that my PhD journey becomes a success.

The encounter with Dilip occurred when I was taking a PhD-level course in the Computer Science Department on 'Computational Intelligence and Its Applications', where he was the lecturer. At that time, I was struggling to analyse the data I had obtained from the WEKIT project at the time. After the first lecture, I approached him with my research problem. Dilip's vast knowledge of machine learning, compassionate heart and selflessness impressed me. Immediately, I had an inner knowing that he was

the supervisor on whom I could place my trust and rely. I requested him to be my co-supervisor, and I am grateful that he agreed to work with me and support me in this project. The discussions between the two of us have been nothing short of excellent, and he has always been there for me. His prompt response to emails, even on weekends or late at night, has been invaluable. I thank Dilip for always being there for me and for being such a great co-supervisor.

As I began working with Professor Gudmestad, I quickly noticed that he preferred to address people by their last names. This led me to assume that he would prefer to be called by his last name as well. It was a special time when he agreed to be my co-supervisor, and I will always remember that afternoon when he came into my office. He was a nice, genial wise man who sat down, opened his notebook, held a pen and was ready to record our talking. I knew right away that this was him, Professor Gudmestad. He asked me about my research status, my problems and even some personal questions. When I told him that my current published research papers were not obviously connected, he encouraged me and convinced me that we can incorporate all the articles into my thesis and connect them through a storyline. The very next evening, I received an email from him with a hand-drawn sketch that turned out to be the flow diagram for my PhD dissertation. Despite his many years of experience and having supervised countless students, his unwavering passion for his work shines through in everything he does. I am grateful to Professor Gudmestad for challenging me and allowing me the space to grow. He has taught me that the devil is in the details; however, we must never forget the bigger picture.

Before I joined the nautical science research group, I felt isolated and alone in my work. As my previous supervisor was hired in a different department, it left me no choice but to navigate my research journey almost entirely by myself. However, joining the nautical science research group was a turning point for me. Though I only became a member towards the end of my journey, it made such a positive impact on both my research and quality of life. The amazing discussions, fantastic social events, endless brainstorming and cosy coffee time with Anders, Gudmund, Johan-Fredrik, Kåre, Magne-Petter, Prasad, Tae Eun and Øyvind were the reason that the small coffee table on the third floor became my favourite spot. The incredibly vast knowledge related to maritime operations, kind hearts and altruism of instructors Anders, Gudmund, Johan-Fredrik and Øyvind were instrumental in helping me conduct two successful experiments and obtain valuable data. I would also like to thank the students from the nautical science programme. Without their willingness to volunteer as a participant in my experiments, I could not have finished the work. Specifically, I want to express my deepest gratitude to Johan-Fredrik, who has been the most amazing co-author I could ever ask for. His enthusiasm for our project was infectious, and his constant flow of creative ideas never failed to inspire me. In addition, his kindness and selflessness in his constant availability to help others, including myself, was truly remarkable. Despite his gentle nature, he never hesitated to challenge my ideas and opinions, keeping me grounded and ensuring that our work was always at its best. I cannot thank Johan-Fredrik enough for his invaluable contributions the success of my PhD.

In the midst of the hustle and bustle of the technology building, a few special people always made time for me by lending an ear and a shoulder to lean on during the incredible moments we shared together. Alyona and Johana E., my former officemates, and Kaori, whom we always met at the lunch table, will never know how much their support meant to me. Their encouragement, valuable experience and sincere suggestions always lifted me up when I needed them the most.

I also want to extend special gratitude to my PhD and postdoc friends at the ITS: Afshin, Behrooz, Bright, Hadi, Hao, Hosna, Jibola, Khan, Leikny, Lise, Minh Tuan, Nikolai, Sushmit and Yufei. From

watching the Eurovision Song Contest and football nights to karaoke sessions and endless moments of fun and excitement, they brought joy to my days in inexpressible ways. They made my PhD journey colourful and unforgettable. I thank them all from the bottom of my heart.

During my time in Tromsø, I had the privilege of meeting many incredible people who enriched my life in so many ways, some for longer periods and some for shorter. I will always be grateful to TODOS, such a wonderful organization for Doctoral Students and Postdocs at UiT, for giving me the opportunity to cross paths with such amazing individuals. Verena and Pedro from the math department, Jenifer from the pharmacy department, Magnar from the chemistry department, and Marit Dagny from the machine-learning group – I thank them all for being a part of my journey in Tromsø. I shared countless unforgettable moments with them – from the aurora exploration night, the whale watching trip, the cabin trip, the ski trip, the Halloween party, the bar night, dinners, deep conversations and experiencing true friendship. I will always cherish the memories of the good food and wine I shared with them and the laughter that echoed through the halls. The guidance on buying an apartment, the insights into Norwegian culture and the experience shared about baby stuff that Magnar and Marit Dagny offered were invaluable. There is one person who deserves a special mention, Verena, to whom I owe a debt of gratitude that can never be repaid. She travelled all the way from Germany to Norway, enduring an 8-hour door to door journey to show up at my wedding. I thank her for being the Godmother of Lukas, extending her love and care for him and hosting my little family for 20 days in her apartment during the summer holiday. Those memories will stay with us for a lifetime, and I feel so grateful to have shared them with her.

Looking back on my journey to becoming a PhD candidate, I am overwhelmed with gratitude for the countless individuals who have helped shape me into the scientist and person I am today. Most importantly, I would like to acknowledge Puneet and Jarle, who hired me and gave me the chance to pursue a PhD. I would also like to express my appreciation for my friends and loved ones back in China, who have cheered me on from afar, providing me with invaluable advice and encouragement along the way. Their unwavering support has been a constant source of inspiration for me, and I feel lucky to have such incredible people in my life.

After almost three pages of gratitude, in which I surely forgot people to thank, comes the part where I could spend at least three more pages – my family. My parents, the backbone of my life, have always been my unwavering support system. They flew across the globe to the Arctic Circle to help me in the last 9 months of my PhD, took care of me, cooked for me, cleaned up after me and took care of my son. Their selflessness and love have been a constant source of inspiration. My in-laws also hold a special place in my heart, accepting me as their own daughter and showering on me unconditional kindness through their souls. I am incredibly grateful for their presence in my life. My son Lukas is the greatest gift to my life. His smile melts me, and his presence fills my life with joy. No matter how tired or stressed I was, his cheerful energy always lifted me up. Last but not least, my husband Lars Daniel, my soulmate and my best friend has been my constant companion through thick and thin. His unwavering belief in me, optimism and unyielding support have been the pillars that held me up during my PhD journey. He is the other half of my soul, and I am eternally grateful for his love.

Five and a half years have passed by in the blink of an eye. The PhD journey can be a very lonely place, even among amazing people; however all of the mentioned people made sure I never felt truly alone. As I embark on this new chapter of my journey, I am filled with a deep sense of gratitude and awe for all those who have helped me reach this point. I thank each of those who have played a part in my story from the bottom of my heart.

Tromsø, Norway

March 2023

Helene

薛慧

Table of Contents

1 Introduction....1
1.1 Background....1
1.2 Motivation for the PhD project....4
1.3 Problem statement....4
1.4 Objectives of the PhD project....5
1.5 Employed technique list....5
1.6 Outline of the thesis....6
2 Theory and related work....7
2.1 Theoretical framework....7
2.2 Maritime education and training (MET)....10
2.2.1 Situation awareness....10
2.2.2 Decision-making....12
2.3 Wearable technology (AR, VR, wristband, etc.)....13
2.4 Biosignal data and stress applied in maritime training....14
2.5 Machine learning....14
2.6 Summary....15
3 Research design and methodology....17
3.1 Tools and measures utilized in the study....17
3.1.1 Assessment of AR user satisfaction....17
3.1.2 Development of Situation Awareness Global Assessment Technique (SAGAT) queries 18
3.1.3 Stress and workload assessments in maritime training scenarios....20
3.2 Design of the experiment and data collection....20
3.2.1 Experiment I....21
3.2.2 Experiment II....21
3.2.3 Experiment III....23
3.3 Analytical methods....25
3.3.1 Statistical method....25
3.3.2 Biosignal data analysis in ML....28
3.4 Summary....33
4 Results and discussion....34
4.1 List of papers....34
4.2 Research findings....35

4.3 Discussion ... 43
4.3.1 Relationship between the articles ... 43
4.3.2 Strengths of the work ... 44
4.3.3 Limitations of the work ... 44
4.3.4 Implications of the work ... 45
4.3.5 Contributions to the research field ... 47
5 Conclusion and future work ... 49
5.1 Concluding remarks ... 49
5.2 Future work ... 50
References ... 52
Paper I ... 61
Paper II ... 87
Paper III ... 97
Paper IV ... 115
Paper V ... 135

List of Tables

Table 1: Summary of techniques used..6
Table 2: Contribution of the PhD candidate to the papers in the PhD thesis 35

List of Figures

Figure 1: Flow diagram of the PhD work. 5
Figure 2: Stress response curve [40] according to Nixon P: Practitioner 1979 [45] and Yerkes–Dodson Law (Yerkes RM, Dodson JD) [43]. 9
Figure 3: Flow diagram of the theoretical framework. Note: developed by the author. 9
Figure 4: Flow diagram of the literature review. Note: developed by author. 10
Figure 5: Flowchart showing how the employed data collection methods are structured and techniques are applied to solve the research questions 17
Figure 6: Hierarchical task analysis of radar tasks in the navigation process. Noting that AIS stands for automatic identification system, the true vector is related to the North and the relative vector is related to the motion of your own ship. 19
Figure 7: The procedure for developing the SAGAT queries. 19
Figure 8: Chart layout with the route and horizontal lines indicating the four SAGAT stops. 21
Figure 9: View from the simulator at UiT, The Arctic University of Norway. The location where the critical situation took place is shown on the map and its corresponding 3D views in two different directions of vision. 22
Figure 10: The planned route Sandnessundet consists of five straight legs. Waypoint 5 is under the middle point of the bridge. 23
Figure 11: Illustration photo of the sailing route of one of the participants and geographical points of the traffic situations. 24
Figure 12: Mixed-methods approach for stress level analysis in maritime training. Note that ML stands for machine-learning. 24
Figure 13: Sample respiratory rate (RR) intervals with activity windows recorded by a participant. Note that the gaps between windows are the time between activities. 24
Figure 14: Illustration of a simple CNN for data classification in this work. Note that in deep stacking, layers can be repeated several or many times. 32
Figure 15: A map of an example of a risky decision and its corresponding 3D view. One of the participants decided to cross the narrow passage between the tugboat and the disabled ship, an action that requires considerable ship-handling skills. 39
Figure 16: Conceptual model of a stress-based maritime training programme. 41
Figure 17: Comparison of machine-learning results from five different algorithms in different ways of pre-processing data. 42
Figure 18: Deviation from the planned route for the experimental scenario (orange) and the control scenario (blue) in the comparison graph. (Experts suggest that the deviation of less than 100 m is negligible.) 42
Figure 19: Participants sailed routes from the control scenario (left) to the experimental scenario (right). The magenta lines represent the planned route. 43
Figure 20: An example of maritime education and training hierarchical work. Note that IMO stands for International Maritime Organization, NMA stands for Norwegian Maritime Authority and STCW stands for International Convention on Standards of Training, Certification and Watchkeeping for Seafarers. 46
Figure 21 : Future work for a reliable and safe maritime training system. 50

List of Abbreviations and Definitions

AIS	Automatic identification system
ANOVA	Analysis of variance
AR	Augmented reality
ARPA	Automatic radar plotting aid
ATV	Automated transfer vehicle
BVP	Blood volume pulse
CNN	Convolutional neural network
COA	Course of action
COLREG	Convention on the International Regulations for Preventing Collisions at Sea
CPA	Closest point of approach
CTB	Cargo transfer bag
DWT	Discrete wavelet transforms
ECG	Electrocardiogram
EDA	Electrodermal activity
EI	Expected improvement
EEG	Electroencephalography
EOG	Electrooculogram
FV	Features vector
GP	Gaussian process
GPS	Global positioning system
HOC	Higher-order crossings
HR	Heart rate
HRV	Heart rate variability
HS	A small bulk carrier, Hagland Saga
HTA	Hierarchical task analysis
IBI	Inter-beat interval
IMO	International Maritime Organization
ISS	International Space Station
KNN	K-nearest neighbours
KSAs	Knowledge, skills and abilities
LDA	Linear discriminant analysis
MARS	Mission awareness rating scale
MET	Maritime education and training
MKO	More knowledgeable other
ML	Machine learning
NASA-TXL	NASA Task Load Index
NAVAID	Navigational aid
NDM	Naturalistic decision-making
NM	Nautical miles
PPG	Photoplethysmogram/Photoplethysmography
QUIS	Questionnaire for user interaction satisfaction
RPD	Recognition-primed decision-making
RR	Respiratory rate
SA	Situation awareness

SAGAT	Situation awareness global assessment technique
SARS	Situational awareness rating scales
SART	Situational awareness rating technique
SA-MA	SA for maritime navigation and collision avoidance
SD	Standard deviation
SFV	Statistical-base feature vector
SGUS	Smart glasses user satisfaction
STAI	State-trait anxiety inventory
STCW	International Convention on Standards of Training, Certification and Watchkeeping for Seafarers
SVM	Support vector machine
TSR	Temporary stowage rack
UHF	Ultra-high frequency
UI	User interface
UiT	UiT The Arctic University of Norway
VAS	Psychometric evaluation of a visual analogue scale
VHF	Very high frequency
VR	Virtual reality
WEKIT	Wearable experience for knowledge intensive training
WFV	Wavelet-based feature vector
ZPD	Zone of proximal development
2D	Two-dimensional Euclidean space
3D	Three-dimensional Euclidean space

List of publications

Paper I.
H. Xue, P. Sharma, and F. Wild, “User satisfaction in augmented reality-based training using Microsoft HoloLens,” *Computers*, vol. 8, no.1, p. 9, 2019. https://doi.org/10.3390/computers8010009

Paper II.
H. Xue, B.-M. Batalden, and J.-F. Røds, “Development of a SAGAT query and simulator experiment to measure situation awareness in maritime navigation,” in: N. Stanton, Eds. *Advances in Human Aspects of Transportation*, AHFE 2020, *Advances in Intelligent Systems and Computing*, vol. 1212. Cham: Springer. https://doi.org/10.1007/978-3-030-50943-9_59

Paper III.
H. Xue, B.-M. Batalden, P. Sharma, J. A. Johansen, and D. K. Prasad, “Biosignals based driving skill classification using machine learning: A case study of maritime navigation,” *Applied Sciences*, vol. 11, no. 20, p. 9765, 2021. https://doi.org/10.3390/app11209765

Paper IV.
H. Xue, J. F. Røds, Ø. Haugseggen, A. J. Christensen, B.-M. Batalden, and O.T. Gudmestad, “A study on the effects of rapid training method on ship handling, navigation and decision-making skills under stressful situations”.
(Manuscript is submitted to *The Journal of Navigation*.)

Paper V.
H. Xue, Ø. Haugseggen, J.-F. Røds, B.-M. Batalden, and D. K. Prasad, “Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing performance.”
(Manuscript is submitted to *Transportation Research Interdisciplinary Perspectives.)*

Other publications

H. Xue, P. Sharma, F. Wild, R. Klemke, I. Koren, W. Guest, A. Vovk, B. Limbu, J. Schneider, and D. Di Mitri, “Requirement analysis and sensor specifications – Final version” (WP 3 | D3.4). Wearable Experience for Knowledge Intensive Training (WEKIT Grant agreement ID: 687669), 2017. Accessed: Jan. 3, 2023. [Online]. Available: https://cordis.europa.eu/project/id/687669/results

H. Xue, J.-F. Røds, and B.-M. Batalden, “The impact of safety factors on decision-making in maritime navigation,” 2023. Accepted by AHFE (Applied Human Factors and Ergonomics) 2023 open-access proceedings.

Chapter 1

1 Introduction

1.1 Background

The maritime sector is defined as consisting of shipping, ports, marine and maritime business services industries, each of which comprises a diverse array of activities [1]. It plays a crucial role in global economic growth, as over 90% of the world's trade relies on sea transportation [2]. Ensuring maritime safety is a critical concern, as any unsafe act - such as collisions, groundings, fires, or pollution - can have serious consequences for the crew, passengers, assets and the environment. Human errors, such as mistakes in ship handling and inadequate decision-making, account for a significant proportion of maritime accidents, particularly those related to navigation [3]. Studies reveal that officers' competence plays a significant role in shipping accidents [4], [5], and proper training is beneficial in equipping maritime professionals with the necessary knowledge and skills to limit navigation errors, handle complex situations and avoid unsafe acts. To achieve this, the unique operating environment of the shipping industry and the need for effective training methods are vital.

Maritime training typically includes both theoretical and practical components, and the specific type of training programme, as well as the country in which it is offered, can impact the content, duration, and focus of the training [6]. The theoretical components cover subjects such as maritime law, navigation, ship stability, meteorology and cargo handling and other related topics, depending on the specific training program and its objectives [7]. Practical components include hands-on experience on board ships, training in ship-handling simulators and other relevant training exercises that simulate the scenario to train the students with traffic clearance, towing and docking operations [8]. Maritime training curricula are typically organised according to the standards set by the International Maritime Organization (IMO) [9]. For maritime education and training, the IMO has developed the convention of standards, known as the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW), which sets out minimum training and certification standards for masters, officers and other personnel on-board ships [10]. The STCW divides maritime training into several levels, including basic training, advanced training and specialised training [11]. Basic training covers the essential knowledge and skills required for the safe operation and maintenance of ships, while advanced and specialised training is tailored to specific job roles and duties on board ships [12]. STCW also requires seafarers to complete periodic refresher training to maintain their competency and stay up-to-date with the latest developments in the industry [13].

In recent times, maritime education and training (MET) has transformed from the informal apprenticeship model to a more formal, structured and internationally unified education with defined learning outcomes for certification and promotion [6]. This system involves a blend of theoretical classroom instruction and practical hands-on training, and simulators are increasingly utilized in training programs [14]. Simulation-based training exercises provide trainees with a controlled environment to practice handling scenarios, including emergency situations, in a safe and controlled manner. However, the real-world environment is often more complex than what can be tested out in practice, and individuals need to be prepared to handle such complex situations confidently. Simulators offer an avenue to bypass the safety implications and associated costs of on-the-job training while providing the benefits of repetitive learning in a realistic, safe, and controlled environment [15].

In the maritime industry, a growing number of companies operate with a limited workforce, requiring employees to possess diverse technical skills and qualifications. Personnel who are well-trained and highly qualified are a valuable asset to any organization. However, there are inconsistencies within the education and training system that prevent the efficiency of the training. Specifically, the assessment system has shifted the focus from acquiring the necessary knowledge and skills for onboard ship operations to passing competency examinations [16]. As maritime education and training can be costly [17], it is essential to implement effective training methods and achieve the training objectives efficiently.

Studies have shown that the competence of instructors and the teaching methods used are critical to improving the learning effectiveness of MET programmes [18]. Such programmes should prioritize practical training opportunities and emphasize the development of the necessary knowledge and skills for onboard ship operations, rather than solely focusing on passing competency examinations. In addition, MET have challenges to adapt to new technology and innovations to keep up with changing market requirements, meeting new international standards, addressing language and cultural barriers, ensuring safety and providing practical training opportunities [19], [20]. Hence, it is crucial for MET to address these challenges and provide high-quality training programmes to produce competent and well-trained seafarers.

To overcome these challenges, a need for more effective training methods has drawn research attention that can enhance individuals' ability to handle complex situations in the maritime sector. The implementation of innovative training methods that can address the limitations of current training methods can be beneficial for improving safety in the maritime industry. Human factors have been identified as contributors to safety incidents in the maritime industry [21]. While operators can contribute to the safety of complex systems by adapting to new and unforeseen events, it is important to address how to limit the negative effects of human behaviours [22]. Therefore, maritime training programmes should be designed to mitigate the negative impact of human factors on safety, and address how to limit the negative effects of human behaviours.

In recent years, research interest has been growing in exploring innovative training methods in the maritime sector to enhance the learning experience, improve information retention and prepare maritime professionals to handle complex situations. To achieve these goals, a variety of innovative approaches have been introduced and applied in MET. These approaches include the use of virtual reality (VR) technology, augmented reality (AR) glasses, multi-sensor fusion training, mobile learning, collaborative learning, and blended learning.

VR technology provides a fully immersive virtual environment for training, making the experience more engaging and enjoyable [23]. The use of VR technology in maritime safety education allows trainees to experience simulations of hazardous situations and practice responding to them in a safe and controlled environment [24]. This technology can help overcome some of the logistic and safety challenges associated with traditional maritime training methods. In addition, due to their portability and ease of use, VR technologies allow students to train at their convenience, increasing opportunities for training and enhancing their knowledge and allowing trainees to practice handling complex situations in a simulated environment that closely resembles the real world [25].

AR glasses offer a semi-immersive experience, allowing students to learn and practice related knowledge through digital information overlay in the real world. Students can learn related knowledge and practice through the application set up in the AR glasses, reducing the repetitive work of the instructor [3]. Using AR technology to overlay digital information onto the real world provides a more interactive and engaging training experience. This allows trainees to practice handling complex situations in a simulated environment that closely resembles the real world, providing a valuable learning experience that is not possible through traditional training methods.

A multi-sensor fusion training method is another innovative approach that has gained research focus in recent years. A pilot study in maritime training employed a multi-sensor fusion framework, using the training method of briefing/debriefing in the simulator; collecting audio, video, eye-tracking data, etc., and visualising operational procedures, thereby achieving the goal of improving the maritime operation skills of seafarers [26].

The use of mobile devices and tablets in training has gained popularity in recent years, which can also be employed in maritime training programmes due to their ability to provide flexible access to training materials and simulations. This allows trainees to access and engage with training materials at their own pace and at a convenient location, providing a valuable and flexible learning experience.

Some other innovative approaches are collaborative learning and blended learning. Collaborative learning involves working with others to solve problems and complete tasks; it provides an opportunity for individuals to learn from each other and improve their skills. Blended learning combines traditional training methods with modern technology to create a more comprehensive, efficient, and effective training experience, as trainees receive training that is tailored to their individual needs and learning styles.

The use of innovative training methods in maritime education and training has gained popularity in recent years; however, more innovative training methods have yet to be researched and developed. This is a crucial area of study, as the maritime industry faces unique and challenging operating environments. Effective training is essential for avoiding unsafe acts, such as collisions and groundings, which is a key factor in enhancing maritime safety. Considering this, a need arises for further research to identify best practices and assess the impact of innovative training methods on the acquisition and retention of the necessary knowledge and skills for safe and efficient maritime operations. This can be achieved through a systematic and comparative analysis of both innovative and traditional training methods, which can provide valuable insights for the design, implementation and evaluation of future training programmes in the maritime sector. The development of new training methods can also play an important role in meeting the potential evolving needs of the maritime industry. With advances in technology and changing operational requirements, maritime training programmes must keep pace and remain relevant.

This will require the continued development and evaluation of new training methods that can effectively prepare maritime professionals for the challenges they are likely to face in their work.

1.2 Motivation for the PhD project

This PhD work was initially part of the EU-funded Wearable Experience for Knowledge Intensive Training (WEKIT) project and was carried out at the Department of Technology and Safety, UiT The Arctic University of Norway. The WEKIT project is multi-disciplinary research that provides an innovative learning experience enabled by tailoring existing wearable smart devices and sensors [27]. However, the WEKIT project ended in 2019 in the middle of my PhD work. However, due to insufficient resources after the closure of the project, my PhD was dedicated to focusing on maritime sector training, as we could utilise our maritime simulator training centre in Tromsø and develop novel simulator-based training and assessment methods. To enhance learning opportunities and continue the previous research direction of WEKIT, wearable sensors are deployed as technologically advanced training aids in maritime training. Analysing the data collected from wearable sensors is one of the essential parts of this thesis, and my background in computer science is just right for this job.

The motivation for this PhD project is, '*To study enhanced learning to handle complex situations in the maritime sector and better understand maritime education and training from theoretical and methodological perspectives*'.

1.3 Problem statement

Based on the defined problems and purposes of this research, this PhD work was conducted with attention to approach the following research questions (RQ):

RQ1: Can wearable technology, for example AR technology, be used in hands-on training satisfactorily?

RQ2: How does experience affect maritime SA and stress levels?

- RQ2.1: Which methods can be used to measure trainees' performance in maritime navigational tasks, specifically SA during maritime navigation, and how does experience impact the navigator's SA?
- RQ2.2: Does experience affect the stress levels of navigators in maritime navigation tasks?

RQ3: How can efficient training progress be made during stressful maritime tasks?

RQ4: Can we build a system to analyse the objective stress levels of navigators based on biosignal data?

According to the research questions, this PhD work is structured as follows:

The main focus of the research questions in this PhD project is illustrated in Figure 1, which highlights the domain, experimental methods and data analysis methods used. The use of AR glasses as an innovative training tool has garnered research interest across a range of sectors due to their potential for hands-on training. RQ1 is dedicated to evaluating user satisfaction with AR glasses and their application to gain a better understanding of the feasibility of implementing AR in maritime training. As depicted in Figure 1, the experimental approach is centred on utilising wearable sensors and simulator-based training, which addresses the latter four research questions and aims to investigate the efficacy of innovative training methods in the maritime sector. The focus of the data analysis in RQ2.2 and RQ4 is to gain a deeper understanding of the impact of trainees' stress levels on experience levels and their training performance through biosignal data analysis. RQ3 emphasises the development of decision-making skills through training in complex situations. To provide a comprehensive understanding of the study and arrive at compelling results, various techniques, including descriptive statistics and machine-learning algorithms, were employed in the data analysis.

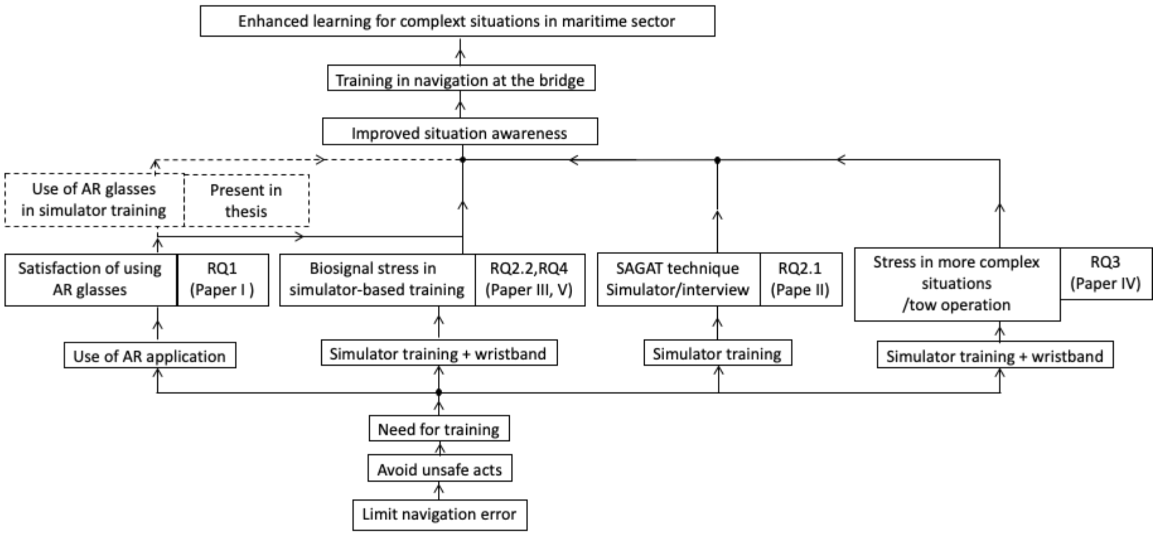


Figure 1: Flow diagram of the PhD work.

1.4 Objectives of the PhD project

This study aims to address the need for effective training methods that can help individuals in the maritime sector limit navigation errors, avoid unsafe acts and improve overall maritime safety. The primary objective of the thesis is to investigate the effectiveness of existing tools utilized in maritime training to achieve the training objectives in an innovative way. Specifically, the study focuses on exploring innovative and effective training methods that can enhance individuals' ability to handle complex situations in the maritime sector. The assessment of cognitive states and their impact on decision-making and performance are important components of maritime training and will be examined in this study. To enhance individuals' abilities to handle complex situations in the maritime sector, this study explores the concept of maritime complex situations through various simulator scenarios, including towing operations, emergency situations, complex traffic situations, and tasks with low visibility. The goal is to identify effective training methods that can help individuals improve their performance in handling such situations.

1.5 Employed technique list

Research in maritime training education is usually mainly concerned with human factor studies; thus, it does not involve highly sophisticated physical and data sciences. However, this study is different from the traditional method, in that, as a multi-disciplinary study, it involves a variety of methods employed to address the aforementioned research questions, which are included but not limited to Table 1.

Table 1: Summary of techniques used

No.	Title	Main techniques employed
Paper I	User satisfaction in augmented reality-based training using Microsoft HoloLens	Descriptive statistics, exploratory data analysis, inferential statistics and parametric estimation
Paper II	Development of a SAGAT query and simulator experiment to measure situation awareness in maritime navigation	Inferential statistics, development of SAGAT query and semi-interview
Paper III	Biosignal-based driving skill classification using machine learning: A case study of maritime navigation	Data pre-processing, feature selection, deep learning, convolutional neural network (CNN) and inferential statistics
Paper IV	A study on the effects of rapid training method on ship handling, navigation and decision-making skills under stressful situations	Descriptive statistics, exploratory data analysis and inferential statistics
Paper V	Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing performance	Data pre-processing, descriptive statistics, inferential statistics, feature selection, exploratory data analysis and representative machine-learning algorithms

1.6 Outline of the thesis

This thesis consists of five distinct sections. The first section provides an overview of the pressing concerns surrounding safety in the maritime industry as well as the training requirements in maritime education. The second section comprises a comprehensive review of the relevant literature. The third section outlines the research methodology employed in this PhD study. The fourth section presents the results and contributions of this PhD project, including the associated scientific publications and applications as well as a discussion of these findings. The fifth and final section offers concluding remarks and recommendations for future work. A comprehensive bibliography is also provided at the end.

Chapter 2

2 Theory and related work

This chapter summarises the important literature in the related area. This chapter, thus, serves five main purposes. Firstly, it presents the theoretical framework for optimizing learning and practice for performance. Secondly, it presents a brief introduction to the current situation of maritime education and training (MET) in general with special attention paid to covering situation awareness (SA) and decision-making. Thirdly, it gives a brief description of wearable technology applied in assisting maritime training. Fourthly, it highlights the state-of-the-art biosignal data and stress applied in maritime training. Finally, it introduces machine learning (ML) as a method of biosignal data analysis.

2.1 Theoretical framework

In the maritime industry, a particular set of competencies is required to succeed, which encompasses knowledge, skills, and abilities (KSAs) [28]. These KSAs can be developed through both training and experience [29]. For instance, successful maritime navigation tasks demand a deep understanding of the relevant equipment, policies, rules, strategies, and procedures. To illustrate, navigators must possess knowledge of international regulations, such as Convention on the International Regulations for Preventing Collisions at Sea 1972 (COLREGs), be adept at using nautical charts, comprehend bridge resource management principles, and have a strong grasp of the fundamentals of radar and automatic radar plotting aids (ARPA) [10].

Skills are technical or manual proficiencies which are usually learned or acquired through training or hands-on experience and should be observable and measurable [30], [31]. In maritime operations, some of the skills are essential for ensuring safety at sea, including problem-solving, cognitive skills, situational awareness, decision-making, route planning and following, as well as social and communication skills [10]. These proficiencies are necessary to navigate and operate vessels effectively in various maritime contexts.

Abilities refer to an individual's present, demonstrable capacity to apply various knowledge and skills simultaneously in order to complete a task or perform an observable behavior [32]. Abilities are considered traits because they exhibit a degree of stability over relatively long periods of time [33]. However, it is recognized that abilities may develop over time and with exposure to multiple situations [34]. In the context of maritime operations, navigators are required to possess a specific set of abilities. After acquiring the necessary knowledge and gaining proficiency in relevant skills, navigators should be able to demonstrate the ability to perform tasks effectively. These include operating equipment and correctly applying information, executing planned decisions, maintaining situational awareness, communicating effectively, demonstrating leadership, and taking appropriate actions following a collision or grounding [10], [35]. These abilities are essential for ensuring the safety and efficiency of maritime operations. Navigators who possess these abilities are better equipped to handle the demands of their job and can perform their duties with greater confidence and effectiveness. As such, the development and assessment of abilities is an important consideration for employers and organizations seeking to ensure the success of their maritime operations.

In the field of maritime training, there exists a discrepancy between the actual proficiency levels and the desired proficiency levels of KSAs required for effective maritime navigation [36]. One way to bridge this gap is by improving the training methods [37]. The Zone of Proximal Development (ZPD) is a concept in educational psychology from Vygotsky (1978) [38] that can serve as a useful framework for enhancing the training methods in maritime education and training [39].

The ZPD is the gap that exists between what a learner can do independently and what they can achieve with guidance and support from a more knowledgeable other (MKO) [40], [41]. In the context of maritime navigation training, the MKO could be an experienced navigator, a qualified instructor, or a training program that provides guidance and support to learners. The training program can identify a learner's ZPD and tailor the instruction to meet their specific needs. By providing guidance and support within a learner's ZPD, the training program can facilitate the learner's growth and development in the KSAs required [42] for effective maritime navigation. This approach enables learners to acquire new knowledge, skills, and abilities that are just beyond their current level of competence, while still within their reach with guidance and support from the MKO [43].

To achieve optimal learning outcomes, it is essential to find the appropriate challenge point for trainees, which enables them to reach their ZPD [42]. The concept of the challenge point, developed by Guadagnoli and Lee, emphasizes the importance of identifying the ideal level of difficulty for a learner to maximize their learning and skill development [44]. As noted by Marteniuk (1976), providing too much information or difficulty can overwhelm the learner, hindering their ability to learn and reducing the potential benefits of the learning process [45]. It is therefore important to ensure that the learning process is structured in a repeatable and stress-resistant manner, allowing trainees to build on their existing knowledge and skills gradually [46].

Moreover, studies have indicated that psychophysiological states such as cognitive workload and stress levels significantly affect performance [47]–[49], [50]. The Yerkes–Dodson law is an empirical relationship between pressure and performance that was developed by psychologists Robert M. Yerkes and John Dillingham Dodson in 1908 [51], explains that performance increases with physiological or mental arousal, but only up to a point. When levels of arousal become too high, performance decreases [52]. Additionally, Nixon published the Stress Response Curve in 1979, illustrating how performance is improved by a certain amount of stress, but then rapidly decays if the stress level (arousal) crosses the fatigue threshold [53]. Research also shows that stress is not always dysfunctional, and when it is positive, stress can prove to be one of the most critical factors in increasing organizational productivity. If not positive, stress can contribute to a variety of physical and mental disabilities in a person and can lead to depression, haste, and poor performance [54]. This relationship between stress and performance is illustrated graphically in Figure 2, based on the aforementioned theories. In the maritime industry, stressors, such as mental and work stress, can lead to fatigue and reduced performance in seafarers, thereby increasing the risk of incidents and accidents [55]. Therefore, understanding the impact of stress and arousal on performance is critical in the development of effective maritime training programs.

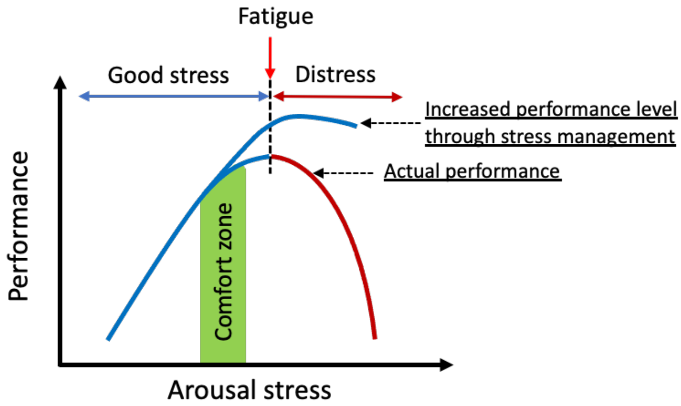


Figure 2: Stress response curve [40] according to Nixon P: Practitioner 1979 [45] and Yerkes–Dodson Law (Yerkes RM, Dodson JD) [43].

Based on the above theory studied, the theoretical framework is developed and presented in Figure 3. The framework illustrates the connection between the challenge point and the ZPD, and how stress management is involved in identifying the appropriate challenge point to improve the training performance of the learner. By identifying the optimal level of difficulty of the training program, the learning process can be structured to provide enough stimulation for learners to maximize their learning and skill development without overwhelming them, thereby optimizing learning outcomes. Meanwhile, to ensure that the learning process is structured effectively, trainers can consider the Yerkes-Dodson law and the Stress Response Curve, which demonstrate the relationship between stress and performance. By understanding and managing stress levels appropriately, trainers can tailor their instruction to meet the specific needs of each trainee, thus promoting optimal learning outcomes and efficient acquisition of competencies such as knowledge, skills, and abilities.

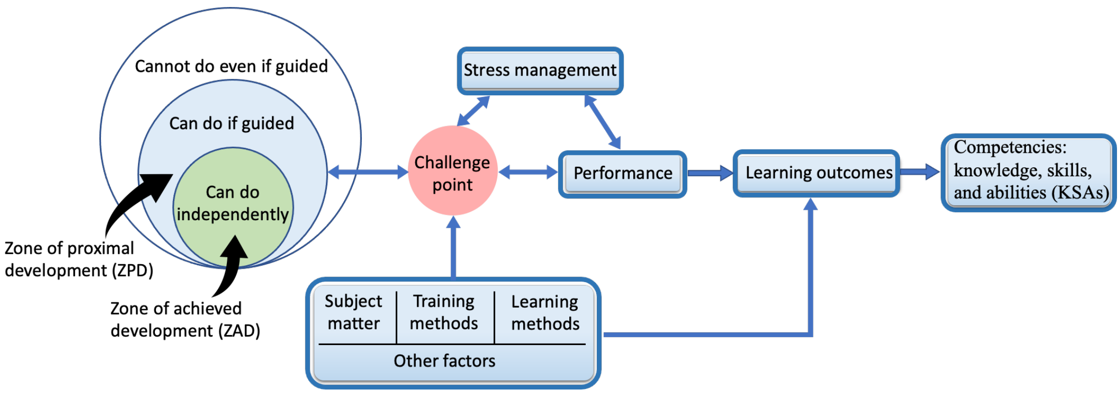


Figure 3: Flow diagram of the theoretical framework. Note: developed by the author.

Drawing on the theories discussed above and the present project, a comprehensive literature review was conducted. Figure 4 illustrates the framework of the literature review undertaken in this study. The review primarily focused on improving training methods for situation awareness and decision-making skills in MET, as well as exploring the application of wearable technology, stress monitoring in maritime training, and biosignal data analysis. The review examined the latest research in these areas, with a particular emphasis on integrating machine learning methods for the analysis of biosignal data. By integrating these innovative approaches into maritime training programmes, it can. enhance training effectiveness and promote the development of learning outcomes and essential competencies.

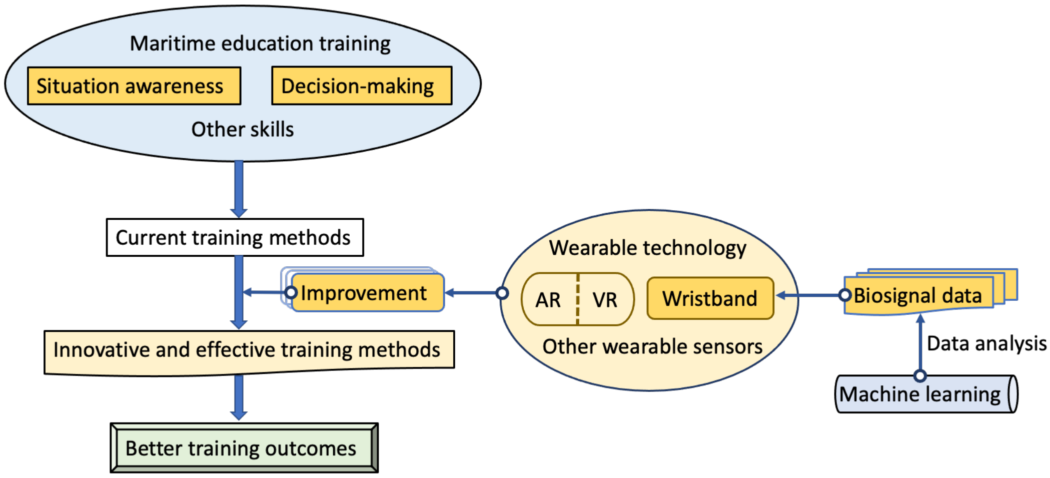


Figure 4: Flow diagram of the literature review. Note: developed by author.

2.2 Maritime education and training (MET)

Maritime education and training (MET) play a crucial role in enhancing safety in the maritime industry, given the complexity and dynamism of the sector. To effectively manage risks and handle emergencies at sea, seafarers must possess a solid foundation of knowledge and skills. As such, maritime training programmes aim to equip seafarers with the necessary knowledge and skills and provide them with opportunities to apply and practice these skills in simulated and real-life situations [6].

However, designing and delivering high-quality maritime training programmes is a complex and challenging task that takes into consideration various factors, such as student levels, exercise design and assessment of learning outcomes [56]. To ensure that these programmes meet the specific needs of the learners, the training programme design should include a variety of exercises and assessments that test the learners' knowledge, skills and abilities to apply these skills in real-life situations.

One significant aspect of maritime education training is the consideration of stress and its impact on learning and performance. A relationship exists between stress and training outcomes: when stress levels become excessive, they can have adverse effects on both health and performance, thereby compromising compliance and participation in safety performance. Based on the maximal adaptability model, stress to a certain degree can also benefit training outcomes [55]. Therefore, it is critical to have a comprehensive understanding of the relationship between stress and training outcomes and develop stress-based training systems that reduce the impact of stress on seafarers.

In maritime training, a wide range of skills is typically taught and trained, including navigation, ship handling, communication and SA and decision-making [15]. These skills are fundamental to ensure safety at sea, making training in SA and decision-making skills particularly crucial in maritime education [59].

2.2.1 Situation awareness

Situation awareness (SA) is a critical component of performance in complex domains, such as the maritime industry, where the operational environment is constantly changing and requires the ability to perceive, understand and anticipate potential threats [60]. SA refers to the cognitive process of perceiving and understanding elements of the environment, comprehending their meaning, and projecting their status in the near future [61]. This process enables individuals to make informed decisions, especially in high-stakes scenarios such as maritime navigation and operations, based on their perception, understanding, and predictions of the current situation [62].

However, acquiring and maintaining SA can be challenging in the complex maritime sector [63]. Seafarers require training in various aspects of SA, including environmental awareness, navigation and communication skills, risk assessment and emergency response procedures [64]. Only conducting SA-related exercises can only improve trainees' capacity to analyze a complex situation and may not have a clear impact on manoeuvre performance [65]. Effective training programmes must be flexible and adaptable to meet the ever-evolving demands of the maritime environment. These programmes can be developed through a combination of theoretical training and practical exercises, such as simulation-based training and real-life scenario training [66], [67]. MET programmes play a crucial role in developing SA in seafarers and provide them with the necessary knowledge, skills and attitudes for safe and efficient ship operations.

The assessment of SA is crucial for measuring the effectiveness of training programmes and ensuring safe and efficient maritime operations [68]. SA assessment can be carried out through either subjective or objective measures [69]. Subjective assessment of SA relies on self-reporting and the subjective perceptions of the trainee, instructor, or evaluator. Questionnaires are typical subject measures. During a training exercise or simulation, trainees may be asked to complete a questionnaire or provide verbal feedback, on their level of SA. Such questionnaires can measure an individual's SA level based on his or her perception of the current situation, understanding of the situation and prediction of future events [70]. Techniques such as Situational Awareness Rating Technique (SART) [70], Mission Awareness Rating Scale (MARS) and Situational Awareness Rating Scales (SARS) [71] are applied in subjective measurements. While subjective assessments can provide valuable insights into trainees' perceptions of their own SA, they can also be influenced by biases and may not accurately reflect the trainee's actual level of SA. Therefore, objective measures of SA are also necessary to provide a more accurate assessment.

Objective measurement involves the use of standardised tests or performance metrics to evaluate an operator's SA skills. One such technique is the Situation Awareness Global Assessment Technique (SAGAT), which uses the freeze online probe method to objectively measure SA [72]. Another example is performance measures. Performance metrics provide objective measurements of an individual's SA skills, including measures of effectiveness and measures of performance. Measures of effectiveness typically assess reaction time, accuracy, and response time, while measures of performance evaluate an individual's performance on specific tasks or subtasks, such as tracking time or reaction time to a specific event [73]. Performance-based measurement of SA is indeed based on the subject's observable behaviors and actions. When combined with other measures of SA, performance-based measures can provide a more comprehensive evaluation of an individual's SA skills [74]. This multi-dimensional approach can help identify broader concerns about the actions taken by operators and improve the effectiveness of training programs.

In addition, several other tools (simulator, computer, telephone, video and audio recording equipment, eye-tracking device, pen and paper, etc.) also aid in SA measurements [75]. Simulation-based training places operators in realistic virtual environments where they can practice and demonstrate their SA skills. This provides a controlled and repeatable platform for evaluating SA skills. Eye-tracking [76] and physiological monitoring devices, such as wearable devices, can track an individual's eye movements and physiological responses in real-life situations and provide insights into their level of SA. Video-based analysis records and analyses an individual's behaviour and actions to assess SA skills in real-life situations [77].

Moreover, it is essential to consider factors such as workload, fatigue, stress and distractions, as they can impact SA and decision-making abilities of individuals [73]. The selection of the appropriate assessment tool or method depends on the specific requirements and goals of the evaluation as well as the resources and expertise available. A combination of tools and methods can provide a more comprehensive and accurate assessment of a seafarer's SA skills [74].

Overall, the assessment of SA skills is a critical aspect of improving decision-making skills and preventing errors and incidents [78]. Particularly, objective SA measures are necessary if simulators are to be used to evaluate the skills and training [79]. While there are some objective SA measures available, more research is needed to develop and refine objective measures in the maritime domain. Objective SA measurements are not well developed in the maritime domain, which makes it challenging to evaluate and improve seafarers' SA skills. Utilising appropriate tools and methods, in conjunction with considering relevant factors, can enhance safety and ensure the continued success of maritime operations. It is important to invest in the development of reliable and valid measures of SA to assess the effectiveness of training programs and improve the overall safety of maritime operations.

2.2.2 Decision-making

The importance of effective decision-making in the maritime industry cannot be overstated, as the safety and well-being of the crew, passengers and vessels are contingent upon it [80], [81]. Effective decision-making involves SA and the ability to process information, weigh options and choose the best course of action. Seafarers must be able to make informed decisions quickly and effectively, even under high-stress situations and adverse conditions.

The effective decision-making process is a crucial aspect of maritime operations, where rapid and accurate decision-making can distinguish between success and failure [82]. To assist in the decision-making process, various decision-making models have been developed and applied in the maritime industry. These models provide a structured and systematic approach to decision-making and can be used to help seafarers identify, analyse and resolve complex problems in a timely and efficient manner. One such model is the 'observe, orient, decide, act' (OODA) loop that emphasises the importance of speed and agility in decision-making [83]. Another model is the 'strengths, weaknesses, opportunities and threats' (SWOT) analysis that helps to identify and evaluate internal and external factors that may impact a decision [84], [85]. Naturalistic decision-making (NDM) is a cognitive approach that emphasises the importance of experience, expertise and context in the decision-making process [86]. NDM views decision-making as an iterative process and plays a crucial role in ensuring safe and efficient operations in the maritime industry.

To support NDM in maritime operations, seafarers must be trained to develop and refine their decision-making skills. This training can take the form of simulation-based exercises, scenario-based training and practical experience in real-life situations. NDM also involves continuous monitoring and feedback, allowing for the refinement of decision-making skills over time.

Other decision-making models commonly used in various domains, including maritime operations, include rational, bounded rationality [87], intuitive decision-making, satisficing and administrative models [88]. These models offer different approaches to decision-making and can be used in different situations, depending on the constraints and available information.

The choice of a decision-making model to be used in maritime operations depends on various factors, including the nature of the problem, the available resources and the goals and objectives of the operation. However, all of these models provide a systematic and structured approach to decision-making that can help improve the accuracy and efficiency of maritime operations.

MET play a critical role in the development of decision-making skills among seafarers [65]. Training in effective decision-making skills must comprise a combination of theoretical training and practical exercises [65]. Theoretical training includes lectures, case studies and group discussions, while practical exercises include simulation-based training, real-life scenarios, gaming, video-based training and behaviour-based training.

In high-stakes scenarios, such as navigation and emergency response, decision-making skills become particularly crucial, requiring the employment of a multi-faceted approach that encompasses both theoretical and practical training as well as continuous monitoring and feedback. Opportunities for feedback and reflection through self-assessment, peer assessment or instructor feedback can provide seafarers with valuable insights into their areas of strengths and those that need improvement [89].

Moreover, having a high level of SA enables seafarers to make informed decisions and respond promptly and appropriately to changes in their environments. In high-stress situations, decision-making can become impaired, leading to errors and potentially hazardous outcomes [90], [91]. Hence, improving decision-making skills is critical for ensuring safety in the maritime industry and can be achieved through effective training programmes that help seafarers acquire the knowledge and skills required to manage risk, make informed decisions and respond appropriately in emergencies.

In conclusion, the utilisation of a combination of theoretical training and practical exercises, along with opportunities for feedback and reflection, can help enhance decision-making skills and ensure the continued success of maritime operations.

2.3 Wearable technology (AR, VR, wristband, etc.)

Wearable technology refers to electronic devices that are designed to be worn by the user and provide a variety of features, such as AR, VR, and the health and fitness tracking field [25]. In the maritime sector, the integration of wearable technology, such as AR, VR, and wristbands, has the potential to improve training outcomes, enhance navigational decision-making and promote overall safety in marine transportation [92], [93]. For example, AR and VR technology can provide a more immersive and interactive learning experience for seafarers by simulating real-life situations and allowing them to practice and apply their skills in a safe environment. In addition, wristbands equipped with sensors can be used to monitor the physiological responses of seafarers, such as heart rate (HR), body temperature, blood pressure volume, and skin conductivity (GSR), which can provide valuable insights into their level of stress and performance during training and operations [94]. These data can provide insight into navigators' stress and experience levels and can be used to tailor training programmes to meet the unique needs of each individual. By collecting and analysing data from wearable devices, it is possible to create personalised training programmes that address the specific strengths and weaknesses of each navigator, ultimately leading to improved learning outcomes and increased safety.

Overall, wearable technology has the potential to revolutionise the way maritime training is conducted and enhance the safety of the maritime industry by improving the effectiveness of training programmes and increasing navigators' ability to handle complex situations.

2.4 Biosignal data and stress applied in maritime training

Biosignal data refer to various physiological signals generated by the body, which can be measured using various instruments, such as electroencephalograms (EEGs), electrocardiograms (ECGs) and electromyograms (EMGs) [95]. These signals provide valuable information about the body's physiological responses to various internal and external stimuli, such as stress, as well as external stimuli like long working hours, irregular shift patterns, and physically demanding tasks [96], [97].

Stress is a common issue in many industries, including the maritime industry, and it has a significant impact on the health and performance of employees [49], [98]. In the maritime sector, stress can lead to navigation errors, decreased attention and memory and decreased decision-making ability, all of which can lead to compromise in the safety of life at sea [99].

Objectively quantifying seafarers' stress levels using biosignal data is vital because self-reported stress levels are often unreliable. For example, the level of cortisol, a stress hormone, can be measured in saliva or blood, providing an objective measure of stress levels of an individual [100]. By analysing biosignal data, researchers and practitioners can gain a deeper understanding of the physiological responses to stress and develop more effective strategies for managing stress in the workplace.

In the maritime sector, the use of biosignal data to analyse stress levels can be especially valuable in training programmes, as stress levels can impact the learning outcomes and performance of navigators. By monitoring stress levels and providing targeted feedback, training programmes can help minimise the negative impact of stress on trainees' performance and decision-making abilities.

2.5 Machine learning

Machine learning (ML) is a rapidly growing subfield of artificial intelligence concerned with developing algorithms and models that can improve performance over time through experience. It has the potential to significantly impact various industries, including the maritime sector, where it can be used to enhance decision-making, situation awareness (SA) and overall safety.

In the context of maritime safety, ML can be applied to large amounts of data, such as automatic identification system (AIS) data, to predict vessel behaviour, traffic patterns and shipping routes [101]–[104]. In addition, ML can also be used in combination with biosignal data analysis to determine the level of stress experienced by seafarers and its impact on their performance. Biosignal data, such as HR, electroencephalogram (EEG) signals and skin conductance, can provide valuable information on the physiological and emotional states of individuals, and ML algorithms can be used to identify patterns and relationships in data that are not easily noticeable by the human eye [96].

By combining ML and biosignal data analysis, it is possible to gain a deeper understanding of the relationship between stress levels and decision-making in maritime workers and develop new approaches and technologies to improve safety and training outcomes. The ability to predict stress levels in real time can help identify factors that contribute to stress and inform the development of effective training programmes.

Overall, the application of ML and biosignal data analysis in the maritime sector has the potential to revolutionise the way maritime safety is studied and improved, leading to better decision-making, improved training outcomes and a safer working environment for seafarers.

2.6 Summary

In summary, the theoretical framework proposed in this study has the potential to improve MET. By identifying the appropriate challenge point for learners, trainers can optimize learning outcomes and promote skill development. The Yerkes-Dodson law provides guidance on the level of stimulation required for learners to achieve their full potential without becoming overwhelmed, thereby reducing performance. Incorporating the ZPD, challenge point, and stress response curve concepts can significantly enhance stress-resistant learning and bridge the gap between actual and desired levels of proficiency in the KSAs required for effective maritime operations.

Furthermore, the results of the literature review highlight the use of wearable technology and biosignal data analysis to help in enhancing training at sea. This information can assist in the development of effective training programs for maritime professionals.

In conclusion, identifying the ZPD and "challenge point" of learners and providing appropriate guidance and support can facilitate the acquisition of necessary KSAs for effective maritime navigation. By incorporating the theoretical framework proposed in this study, training programs can be optimized for better outcomes, thereby promoting skill development and enhancing the safety and efficiency of maritime operations.

‘The goal is to turn data into information, and information into insight’.

Carly Fiorina, former CEO of Hewlett-Packard

Chapter 3

3 Research design and methodology

This section summarises the main techniques applied in this thesis. Each of these techniques, experiment design, data collection and data analysis will be discussed in detail in the relevant subsections. Figure 5 shows a flowchart of the way experiments and data collection methods are structured and data analysis techniques are applied to solve the research questions.

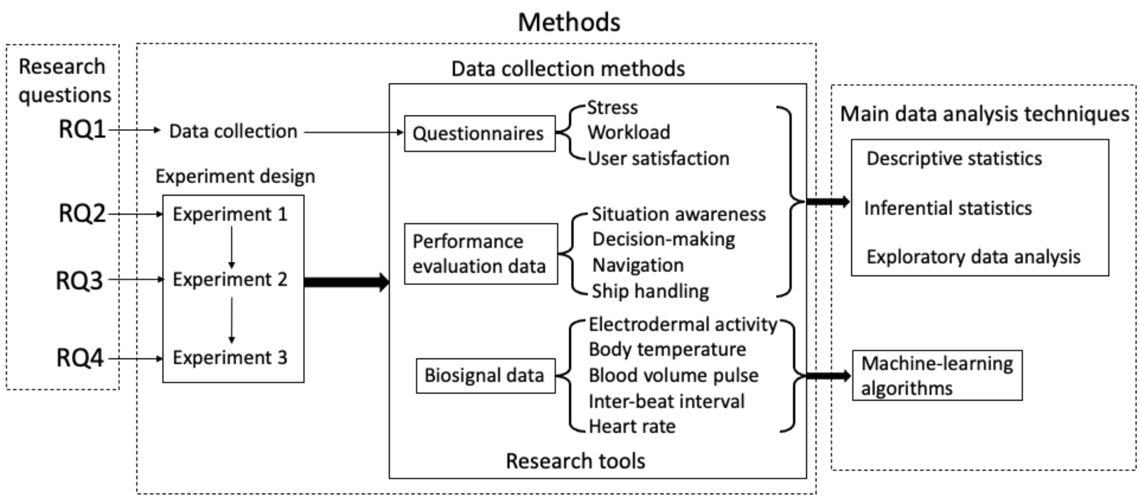


Figure 5: Flowchart showing how the employed data collection methods are structured and techniques are applied to solve the research questions.

3.1 Tools and measures utilized in the study

The purpose of this section is to describe the tools and measures used in this study.

3.1.1 Assessment of AR user satisfaction

The assessment of AR user satisfaction is a key element of technology acceptance of AR and wearable technologies [105]. AR user satisfaction is dependent on both the design of the user interface (UI) and the choice of the AR hardware. There are several concepts and subjective measures for evaluating the user experience of AR services. With regards to the user, satisfaction questionnaires are common tools used to evaluate a user's experience. One such tool — the Questionnaire for User Interaction Satisfaction (QUIS) — is designed to assess users' subjective satisfaction with specific aspects of the human–

computer interface [106]. The results of QUIS facilitate new developments by addressing reliability and validity problems found using its satisfaction measurements. Therefore, the measure is highly reliable across many types of interfaces. The other tool is The Smart Glasses User Satisfaction (SGUS) questionnaire was created for the WEKIT trials. It is a tool designed to assess users' subjective satisfaction with smart glasses. The general objective of the questionnaire is to understand the potential end users' central expectations of AR services with smart glasses, especially from an experiential point of view. The following is a brief description of these two questionnaires.

- Questionnaire for user interface satisfaction (QUIS)

The Questionnaire for User Interaction Satisfaction (QUIS) measures subjective satisfaction with specific aspects of the interface and interaction between the user and the AR application. QUIS consists of a demographic questionnaire, a six-scale measure of overall system satisfaction, and hierarchically organized measures. The measures include the following specific interface factors [107]: screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation. Each area is measured by a seven-point scale according to the user's overall satisfaction with the interface and the above factors.

- Smart glasses user satisfaction (SGUS)

SGUS is a method and measure to scrutinize aspects, such as an enhanced perception of the environment, interaction with the augmented environment, implications of location and object awareness, user-created AR content, and the new AR features that users typically use [108]. In this study, the smart glasses used for the different use cases were Microsoft HoloLens. SGUS measures subjective satisfaction on the basis of different features associated with user satisfaction, such as the content and interaction with the content. SGUS is based on evaluation criteria for web-based learning [109] and statements evaluating the user experience of mobile augmented reality services [108]. Some of the items from the table "Evaluation criteria for web-based learning—a framework" [109] and table "Examples of formative subjective statements with regard to the value and overall goodness of the service in terms of the UX category in question" [108] were picked and modified. SGUS consists of 11 items (statements) on a seven-point Likert scale (1–7) [18]. The 11 statements include three categories of evaluation criteria, which are general interface usability criteria, AR interaction-specific criteria for an educational AR app, and learner-centered effective learning [109].

3.1.2 Development of Situation Awareness Global Assessment Technique (SAGAT) queries

In this study, the Situation Awareness Global Assessment Technique (SAGAT) for SA for maritime navigation and collision avoidance (SA-MA) is developed and evaluated using interviews with experts and a simulator experiment. The interviewees, who have extensive experience as navigators in the merchant fleet and the navy, were consulted to develop a Hierarchical Task Analysis (HTA) of the navigation and collision avoidance tasks. Each task was defined with a goal to achieve, and the HTA was adapted to different situations and needs [110]. The HTA was developed based on a literature study and discussions with subject matter experts. Figure 6 presents the HTA of radar tasks in the navigation process, which was used to list the navigation and collision avoidance tasks as input to the SA queries.

The SAGAT was initially developed to assess pilots' SA across the three SA levels [60]. In this study, the SAGAT procedure comprises ten steps: defining tasks, developing SA queries, selecting participants, briefing participants, pilot runs, task performance, freezing the simulation, administering SA queries, evaluating query answers (query answer by a subject matter expert), and calculating SAGAT scores [111]. The procedure of developing SAGAT Queries is shown in Figure 7.

The SA queries used in the simulator experiment were developed based on the results of the SA requirements analysis [112]. The queries encompass all three levels of SA for a global measure and were administered at four stops along a fixed course line. These stops provided a consistent and controlled environment for the administration of the queries.

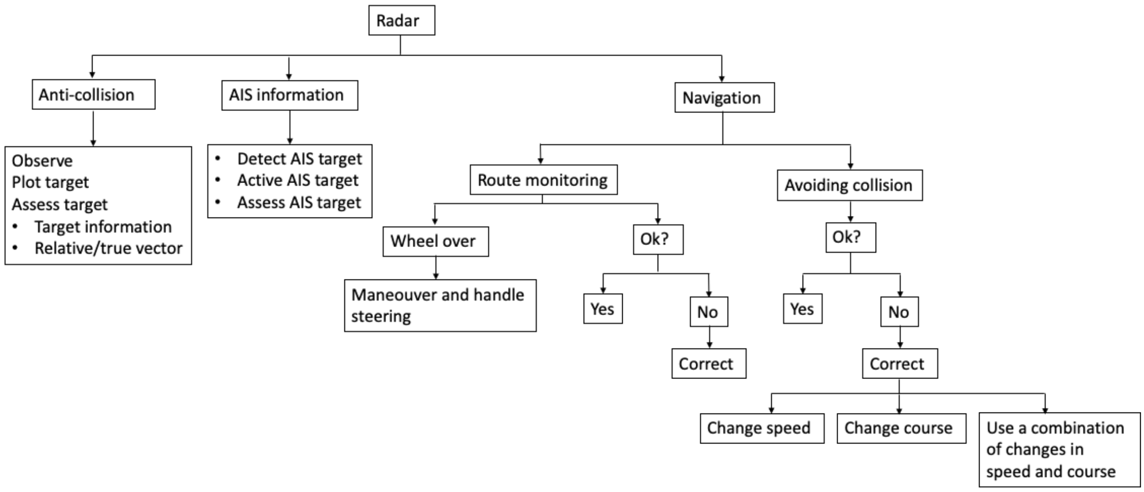


Figure 6: Hierarchical task analysis of radar tasks in the navigation process. Noting that AIS stands for automatic identification system, the true vector is related to the North and the relative vector is related to the motion of your own ship.

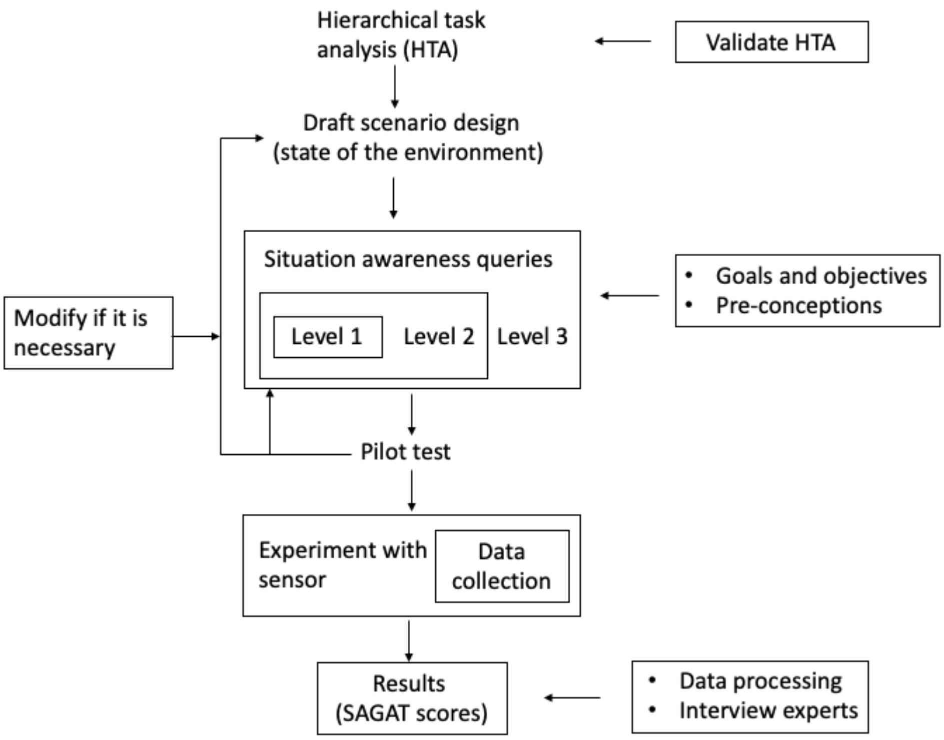


Figure 7: The procedure for developing the SAGAT queries.

3.1.3 Stress and workload assessments in maritime training scenarios

This study provides a comprehensive examination of stress and workload assessments in maritime training scenarios through the use of questionnaires and biosignal data analysis. The objective stress levels were analyzed using biosignal data collected through wristbands worn by participants during maritime tasks. Machine learning algorithms were applied to the biosignal data for analysis. In addition, subjective stress levels and workload were assessed using validated questionnaires. Statistical tests were performed to determine the significance of subjective stress level and workload differences among various complex scenarios. To access the subjective workload, NASA Task Load Index (NASA-TLX) was utilized, while the subjective stress was assessed by the state-trait anxiety inventory (STAI) form Y-1. The following is a brief description of these two questionnaires.

- NASA Task Load Index (NASA-TLX)

NASA-TLX is a widely recognized assessment tool that is used to evaluate the perceived workload of participants in a given task [113], [114]. It consists of six categories that are rated by participants following the completion of the sailing scenario. These categories include mental demand, physical demand, temporal demand, performance, effort, and frustration level [115]. The ratings are then converted to a ten-point scale score, with 0 representing low levels and 10 representing high levels. The scores from each subscale are combined to give an overall workload score.

- State-trait anxiety inventory (STAI) form Y-1

STAI Y-1 form is a widely used self-assessment tool for evaluating state and trait anxiety in individuals [116]. The questionnaire consists of 20 items that measure state anxiety, which refers to a temporary and situational emotional state [117]. Participants are required to rate the intensity of their anxiety symptoms. The scores obtained from the STAI Y-1 form are commonly classified into three categories: "no or low anxiety" (20-37), "moderate anxiety" (38-44), and "high anxiety" (45-80). These ranges are used as a benchmark to classify the level of anxiety experienced by the participants.

3.2 Design of the experiment and data collection

The purpose of this section is to detail the data collection procedures employed in this study. Specifically, data were collected through three experiments designed for this research including three categories: (1) questionnaire data on the stress levels, workload and user satisfaction of auxiliary training equipment; (2) performance evaluation data on SA, decision-making, navigation and ship-handling abilities and (3) biosignal data, including electrodermal activity (EDA), body temperature, blood volume pulse (BVP), inter-beat interval (IBI) and heart rate (HR).

In this study, user satisfaction data were collected through an experiment conducted by the WEKIT project. The remaining data were collected through the following experiments designed specifically for this study:

- Experiment I: An experiment examining SA in maritime navigation among both experts and students
- Experiment II: An experiment measuring stress levels and classifying the complexity of sailing scenarios
- Experiment III: An experiment examining stress in a complex situation, specifically comparing the learning outcomes of different teaching methods in towing operations

3.2.1 Experiment I

In this experiment, a maritime navigation task was designed to investigate the relationship between two factors: navigating experience and stress. The task was performed on a 240° view simulator equipped with Kongsberg Digital's K-sim Navigation software using a vessel model called BULKC11 (which has an overall length of 90 m and a moulded beam of 14 m). The task consisted of the following two parts: sailing and completing Situation Awareness Global Assessment Technique (SAGAT) queries on a frozen simulator screen. Each participant sailed a 40-min. voyage with four stops, with each section of the sailing lasting approximately 8–12 min. During the sailing section, participants completed the SAGAT queries in approximately 15 min. (4 stops with an average of 4 min. to answer the SAGAT queries). The entire experiment lasted approximately 55 min. For the four stops, each stop was conducted within a fixed range on the course line (see Figure 8). An expert completed the same SA queries with the correct answers on the simulator, and the participants' answers were compared with the expert's results. During the experiment, each participant wore a wearable device to collect biosignal data.

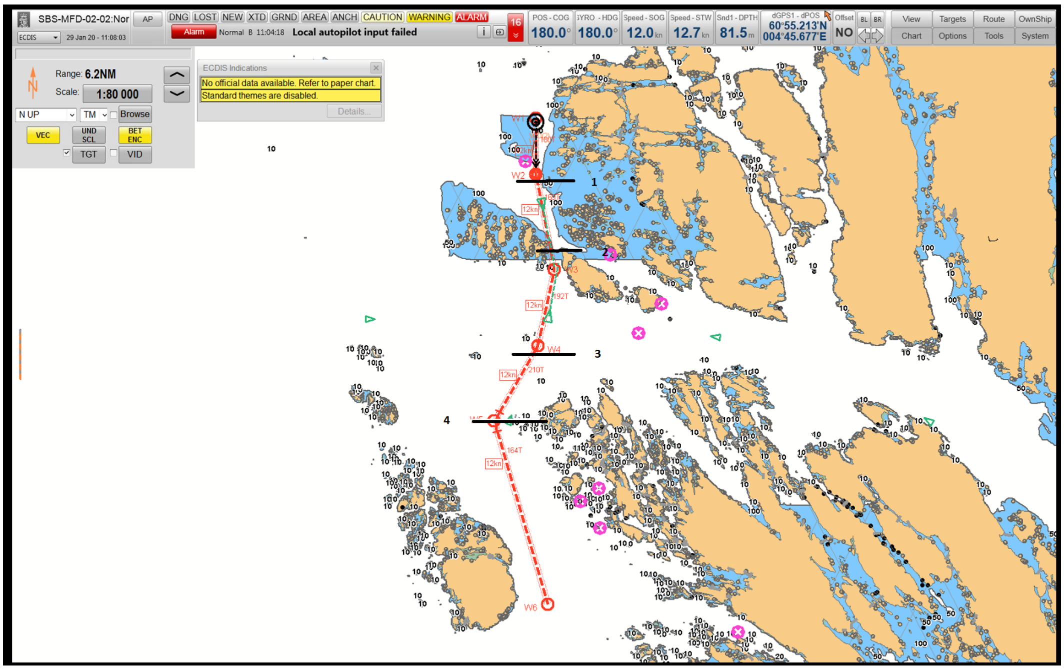


Figure 8: Chart layout with the route and horizontal lines indicating the four SAGAT stops.

Hierarchical task analysis (as shown in Figure 6) was used to map the navigation and collision avoidance tasks. The results of the simulator experiments and inputs from subject matter experts were used to determine the SAGAT score. During the experiment, questionnaire data, performance data and biosignal data were collected for analysis.

3.2.2 Experiment II

This experimental study aimed to investigate the effects of different training methods on cognitive workload, stress levels and decision-making during a towing operation scenario. The experiment used a within-subject factorial design where second-grade students were assigned to the control group and first-grade students were assigned to the experimental group, resulting in a quasi-experiment design due to the lack of random assignment of participants to groups [118]. The control group consisted of second-grade students who had completed 32 simulator exercises or 96 h, including two externally evaluated examinations, while the experimental group consisted of first-grade students who had completed half of the required simulator exercises (16 exercises or 48 h, including two externally evaluated examinations). In addition, the experimental group received a project-specific rapid training course specifically designed for the towing operation task, including a 20-min. video lecture and 1 h of hands-on training and practice on the simulator. Within each group, pairs of participants were assigned as teams to complete the towing operation under the designed scenario. The participants were randomly assigned to two tugboats, and they could communicate with each other via maritime VHF radiocommunication and with the instructor station using UHF radiocommunication.

In this experimental study, the scenario took place in good weather and involved the failure of both engines at a specific location during the towing operation. The location of the engine failure was decided to be near Ryøya Island, which is located south of Kvaløya and southwest of Tromsø, and the geographical locations were held constant throughout the experiment. Three-dimensional views of the sailing route in two different directions of vision can be found in Figure 9.

Data collection for this study involved the measurement of several dependent variables, including training methods, cognitive workload, stress levels and decision-making. These variables were assessed using a range of methods, including the Empatica E4 Wristband for biosignal data collection, questionnaires about workload and stress levels and a customer decision-quality rating scale to measure decision-making performance.

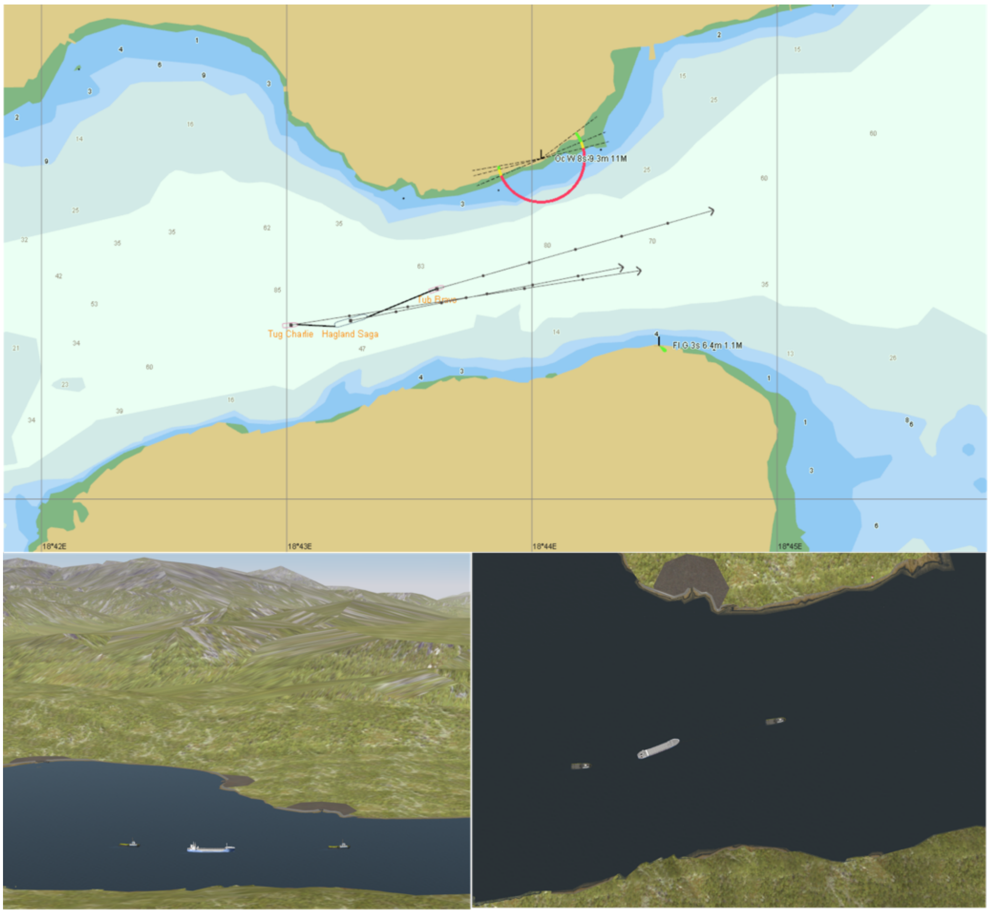


Figure 9: View from the simulator at UiT, The Arctic University of Norway. The location where the critical situation took place is shown on the map and its corresponding 3D views in two different directions of vision.

3.2.3 Experiment III

In this experiment, two complex scenarios were designed for maritime navigation tasks to examine and classify different scenarios in terms of complexity and the stress levels of seafarers. The control scenario was conducted under fair weather conditions with six events, while the experimental scenario was conducted under snowy weather conditions. In addition, the experimental scenario included four more events than the control scenario to create different levels of difficulty. Other simulated variables, such as location and traffic situation, were held constant across both scenarios. The experiment was conducted using three simulator bridges equipped with Kongsberg Digital's K-sim Navigation software, each with either a 240° or 360° view, and three independent instructor stations. The vessel model used was the BULKC11 Hagland Saga, a small bulk carrier with a length between perpendiculars of 85 m. The sailing route chosen was the Sandnessundet strait in Tromsø, Norway, which is frequently used for navigational training for nautical students at UiT. The route consisted of six waypoints (as shown in Figure 10), with a significant turn at the third waypoint, and the participants encountered two fishing vessels and a tug (as shown in Figure 11).

Data collection for this experiment involved the use of a medical-grade wearable device, the Empatica E4 Wristband, to collect biosignal data from the participants (as shown in Figure 12). These data included sample respiratory rate (RR) intervals with activity windows (as shown in Figure 13). In addition to biosignal data, questionnaires about workload and subjective stress levels were collected after each section of the navigational tasks (as shown in Figure 12). To assess learning objectives, performance was measured using the following criteria:

- Number of position fixes on the chart
- Deviation from the planned route and score based on the distance from the planned course using the assessment tool (K-Sim Instructor) in the simulator
- Score based on the closest point of approach (CPA)

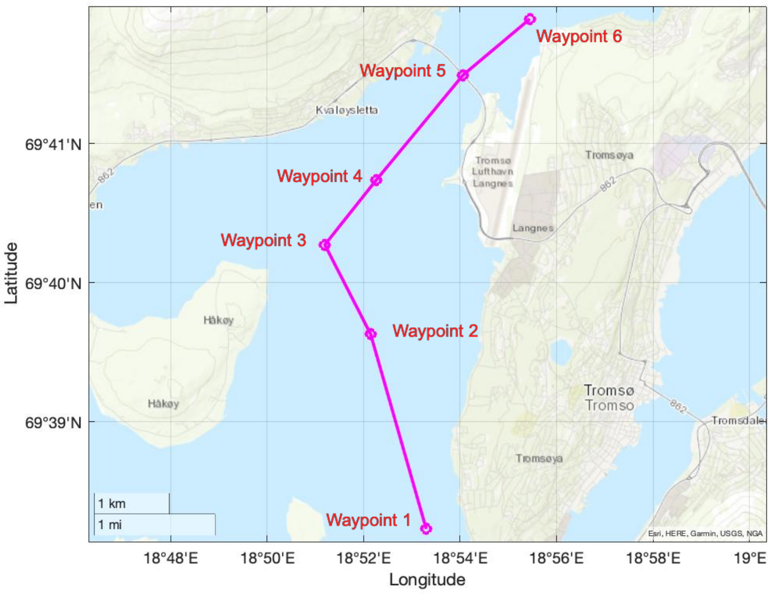


Figure 10: The planned route Sandnessundet consists of five straight legs. Waypoint 5 is under the middle point of the bridge.

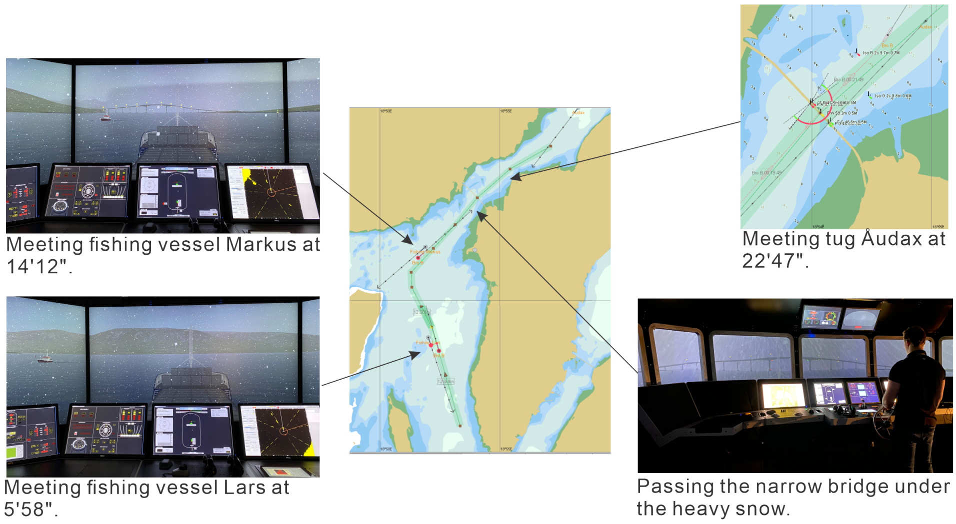


Figure 11: Illustration photo of the sailing route of one of the participants and geographical points of the traffic situations.

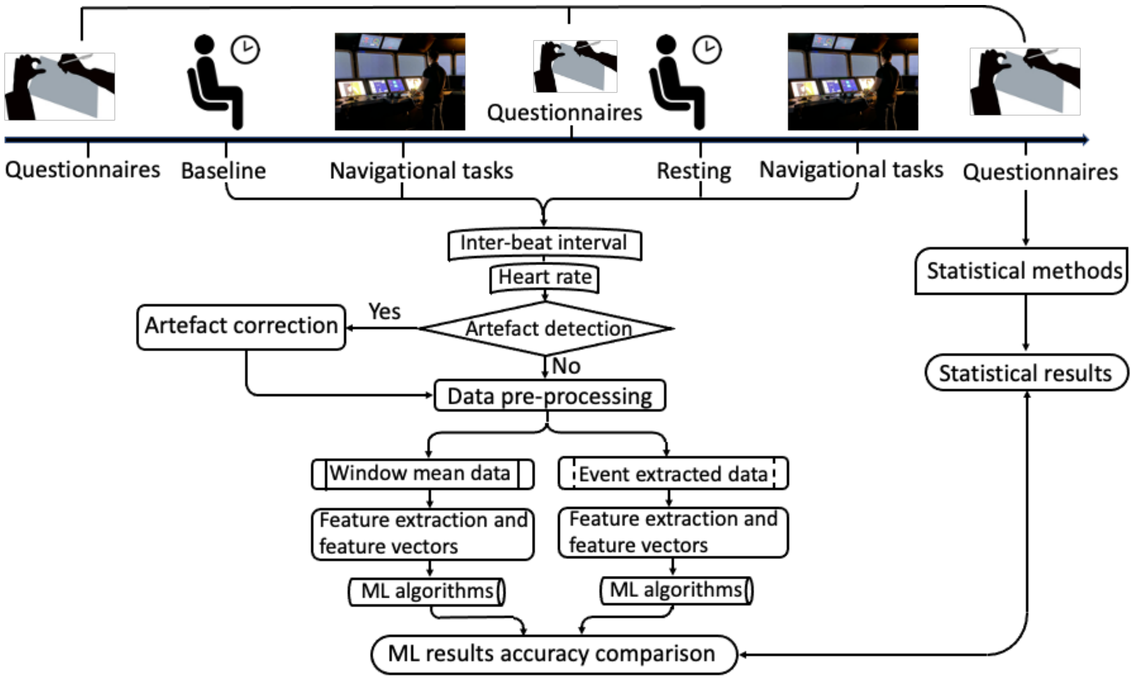


Figure 12: Mixed-methods approach for stress level analysis in maritime training. Note that ML stands for machine-learning.

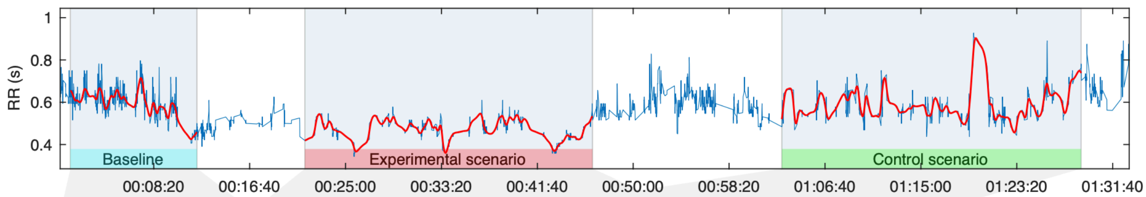


Figure 13: Sample respiratory rate (RR) intervals with activity windows recorded by a participant. Note that the gaps between windows are the time between activities.

3.3 Analytical methods

To answer the research questions, several analytical procedures were adopted, which are described below.

3.3.1 Statistical method

The research tools employed in this work consisted of questionnaires that included measures of user satisfaction, perceived stress and workload as well as performance evaluation data. The collected data were analysed using a combination of statistical methods, including the analysis of variance (ANOVA), the Kruskal–Wallis H test, the Spearman rank correlation coefficient and the Welch Two-Sample t-test to conduct a robust and comprehensive analysis of the data and identify any significant relationships and patterns that may exist between the specific variables of interest, which were the data collected in the study. The results of this analysis were then used to inform the development of a stress-based training model for maritime navigation, with the goal of improving the quality and effectiveness of such programmes and ultimately enhancing safety at sea.

In this PhD work, it is noteworthy that the statistical methods were not calculated manually; instead, the analyses were performed using the programming language R.

The following are the statistical methods used:

(1) Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is a statistical method used for determining information about the means between two or more groups. The analysis used variances between and within samples. In this PhD work, one-way and two-way ANOVA have mainly been used in the analysis of user satisfaction.

Depending on the type of ANOVA being used and the assumptions met by the data, different formulae are used to calculate ANOVA.

For one-way ANOVA, the formulae for the F-statistic are Equations (1), (2), and (3) [119]:

$$F = \frac{MS_{\mathrm{b}}}{MS_{\mathrm{w}}} \tag{1}$$

$$MS_{\mathrm{b}} = \frac{SS_{\mathrm{b}}}{df_{\mathrm{b}}} = \frac{\sum_{j=1}^{k}(\bar{X}_j - \bar{X})^2}{n-k} \tag{2}$$

$$MS_{\mathrm{w}} = \frac{SS_{\mathrm{w}}}{df_{\mathrm{w}}} = \frac{\sum_{j=1}^{k}(\bar{X}_j - \bar{X})^2}{k-1} \tag{3}$$

where

- MS_{b} is the mean square between groups, which is calculated as the sum of squares between groups divided by the degrees of freedom between groups.

- MS_{w} is the mean square within groups, which is calculated as the sum of squares within groups divided by the degrees of freedom within groups.
- SS_{b} is the sum of squares between groups.
- SS_{w} is the sum of squares within groups.
- df_{b} are degrees of freedom between groups.
- df_{w} are degrees of freedom within groups.
- n is the total of all the data sets combined together.
- k is the number of groups.
- $\bar{X}$ is the total mean.
- $\bar{X}_j$ is the mean in different groups with j donated to the j^{th} group.

The *p*-value was then calculated using the F-distribution with the appropriate degrees of freedom.

For a two-way ANOVA, the formula for the F-statistic is more complicated as it includes multiple sources of variation, such as the main effects of each independent variable and the interactions between and among them. The formula can be derived by dividing the total sum of squares into component sums of squares, each of which represents a different source of variation.

In this work, a special case of one-way ANOVA, a t-test, was also used. In the presence of only two groups, the F-statistic from ANOVA will be equivalent to the t-statistic from the t-test; thus, the calculated p-value will be the same. However, ANOVA is more powerful when more than two groups are present, and it allows the test of the overall effect of the independent variable and not just compare two groups.

(2) Kruskal–Wallis H test

The Kruskal–Wallis H test (also known as 'one-way ANOVA on ranks') is a rank-based non-parametric statistical test that can be used to determine whether there are statistically significant differences between two or more independent groups on a continuous or ordinal dependent variable [120]. This test was applied to the data obtained from the questionnaires used to assess the participants' stress levels and workloads.

The test statistic was calculated using Equation (4) [121] and the degrees of freedom were determined using Equation (5). The corresponding p-value was calculated using the chi-square distribution with corresponding degrees of freedom.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i (\bar{R}_i - \bar{R})^2 \tag{4}$$

$$df = k - 1 \tag{5}$$

where

- H is the test statistic.
- N is the total data sample size (e.g. three groups rated for two different-level tasks) for each questionnaire.
- k is the number of groups compared.

- n_i is the sample size for group i.
- $\bar{R}_i$ is the average of the ranks in a group i.
- R is the average of all the ranks among all samples.
- df is the degree of freedom.

(3) Spearman rank correlation coefficient

The Spearman rank correlation coefficient, also known as Spearman's rho (ρ), is a statistical measure used to determine the strength and direction of the association between two variables that are measured on an ordinal or continuous scale. Unlike Pearson's correlation coefficient, which is used to measure the linear association between two continuous variables, Spearman's rank correlation coefficient is based on the ranks of the data rather than the actual values.

The formula for the Spearman rank correlation coefficient is given in Equation (6) [119]:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n} \tag{6}$$

where

- ρ is the Spearman rank correlation coefficient.
- $\sum_{i=1}^{n} d_i^2$ is the sum of the squared differences between the ranks of the two variables.
- n is the number of observations.
- $d_i = r_i - s_i$, where r_i is the rank of variable x and s_i is the rank of variable y

It is important to note that the value of ρ ranges from −1 to 1, where

- A value of 1 indicates a perfect positive correlation, meaning as one variable increases, the other variable also increases, and as one variable decreases, the other variable also decreases.
- A value of –1 indicates a perfect negative correlation, meaning as one variable increases, the other variable decreases, and as one variable decreases, the other variable increases.
- A value of 0 indicates no correlation, meaning there is no relationship between the two variables.

Spearman's rank correlation coefficient is commonly used when the data. do not meet the assumptions of normality and linearity required for Pearson's correlation coefficient. It is often used in non-parametric statistical analyses and can be used to examine the correlation between ordinal or continuous variables.

(4) Welch two-sample t-test

The Welch Two-Sample t-test is a statistical method used to compare the means of two independent groups when the variances of the two groups are not necessarily equal [122], [123].

The basic steps for conducting the Welch Two-Sample t-test are as follows:

- State the null and alternative hypotheses.
- Determine the level of significance.

- Collect and organise the data.
- Check the assumptions of normality; this test makes no assumptions about the variances between the two groups.
- Calculate the test statistic and the p-value.
- Make a decision and interpret the results.

Welch's t-test defines the statistic t by the following formula of Equation (7) [124]:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{7}$$

where

- $\bar{X}_1$ and $\bar{X}_2$ are the means of the two groups.
- s_1 and s_1 are the sample variances of the two groups.
- n_1 and n_2 are the sample size of the two groups.

3.3.2 Biosignal data analysis in ML

Biosignal data are often high-dimensional, time-series data that can be challenging to analyse using traditional methods. The purpose of this section is to detail the process of using ML techniques to analyse and extract information from biosignal data.

A. Data pre-process

Using ML for biosignal data analysis can be challenging, as the data are often highly variable and noisy. In addition, the quality and quantity of the data may be limited, and the data may be collected under different conditions and from different individuals, leading to variability in the data. Therefore, it is important to pre-process and clean the data, such as by removing noise or artefacts, before applying ML techniques. In this study, following methods were applied for pre-processing the data: (1) downsample the data to the same frequency; (2) normalise the data; (3) detect and remove the artefacts; (4) collect the window mean data and (5) select the event extracted data. The details can be found in the research papers that are attached as an appendix at the end of the dissertation.

B. Feature selection

ML techniques can be used to extract features from biosignal data. In this study, the following features were extracted and used alone or in combination to improve the performance of a ML model.

(1) Statistic-based features

In this study, statistic-based features were extracted from the biosignal data set using a range of statistical analysis techniques. These features included measures of central tendency, such as the mean and median, and measures of dispersion, such as the range and standard deviation. In addition, the study extracted measures of skewness and kurtosis, which describe the symmetry and peakedness of the data distribution, respectively. These features were collected as feature vectors including mean vectors, standard deviation vectors, variance vectors, median skewness vectors and kurtosis vectors that have been extracted from the data set (shown in Equation (8)). These vectors were a numerical representation of each feature, containing the values of the feature for each data set sample. Using these feature vectors in ML models was beneficial in identifying and classifying the biosignal effectively, providing useful insights and improving the system's overall performance.

$$F_s = [\mu_{X_i}, \sigma_{X_i}, V_{X_i}, Sk_{X_i}, K_{X_i}], \qquad (i = 1,2, \dots l) \tag{8}$$

where:

- F_s are the statistic-based feature vectors.
- μ_{X_i} is the mean of the data set.
- σ_{X_i} is the standard deviation of the data set.
- V_{X_i} is the variance of the data set.
- Sk_{X_i} is the skewness of the data set.
- K_{X_i} are the kurtosis vectors.
- X_i is the data set of each participant in each task, where l is the length of X_i

(2) Wavelet-based features

In this study, wavelet-based features were extracted from biosignal data using a technique called discrete wavelet transform (DWT). In particular, the study used Daubechies wavelets (the number of vanishing moments was 4), which are a commonly used set of wavelets known for their mathematical properties and suitability for signal processing [125].

To extract the wavelet-based features, the study computed the coefficients of the Daubechies wavelets at different scales. In this case, the scales used were 2, 4 and 8 to obtain three levels of resolution for the wavelet coefficients. The result of this process was a matrix with three rows representing the different scales and columns equal to the length of the data for each participant in each task.

Then, two different types of wavelet-based features were computed from this matrix, as seen in Equation (9). The first type was the sum of the square of the wavelet coefficients, as shown in Equation (10), and the other type was the sum of the product of the square of the wavelet coefficient and the natural logarithm of the square of the wavelet coefficient, as shown in Equation (11). These features capture and describe different aspects of the signal, providing additional information that may be useful for the identification and classification of biosignals.

$$F_W = [F_{W_1}, F_{W_2}] \tag{9}$$

$$F_{W_1} = \left[\sum_{i=1}^{l} Y_{L_1}^2, \sum_{i=1}^{l} Y_{L_2}^2, \sum_{i=1}^{l} Y_{L_3}^2\right] \tag{10}$$

$$F_{W_2} = \left[\sum_{i=1}^{l} (Y_{L_1}^2 \cdot \ln(Y_{L_1}^2)), \sum_{i=1}^{l} Y_{L_2}^2 \cdot \ln(Y_{L_2}^2), \sum_{i=1}^{l} Y_{L_3}^2 \cdot \ln(Y_{L_3}^2)\right] \tag{11}$$

where

- F_W is the wavelet-based feature.
- F_{W_1} and F_{W_2} are the two different ways of computing.
- Y is the Daubechies wavelet coefficient in three levels L_1, L_2 and L_3 .
- l is the length of the data sample.

(3) Higher-order crossing (HOC)-based features

Higher-order crossings (HOC)-based features, also known as zero-crossing-based features, are a set of features extracted from the analysis of the patterns of zero-crossings in a signal. Zero-crossing, a commonly used concept in signal processing, refers to the point at which the signal changes from positive to negative or vice versa [126]–[128][123]–[125]. In this study, the HOC features were extracted in the following steps:

Computation of the difference between adjacent elements was done in the data series in different orders. The k^{th} order difference is shown in Equation

(12). From $\nabla^{k-1}Z_t$ in the Equation (13), a binary process $X_t^{(k)}$ is defined in Equation

(13). The count of the symbol changes from $X_t^{(k)}$, D_k, was calculated in Equation

(14).

$$\nabla^{k-1}Z_t = \sum_{i=1}^{k} C_{i-1}^{k-1}\,(-1)^{i-1}Z_{t+1-i} \quad with \quad C_{i-1}^{k-1} = \frac{(k-1)!}{(i-1)!\,(k-1)!} \tag{12}$$

$$X_t^{(k)} = \begin{cases} 1, & \nabla^{k-1}Z_t \geq 0 \\ 0, & \nabla^{k-1}Z_t < 0 \end{cases} \tag{13}$$

$$D_k = \sum_{t=2}^{N} [X_t^{(k)} - X_{t-1}^{(k)}]^2 \tag{14}$$

where

- $k = 1,2, \ldots$
- ∇^0 is the zero-mean data series.
- D_k is the count of symbol changes in k^{th} order.

In this study, the extraction of HOC-based features from the biosignal data was represented by a vector consisting of the number of axis crossings in a zero-mean data series outlined in Equation (15). The resulting HOC-based features were found to be beneficial in improving the performance of ML models used in this study, providing useful insights and better accuracy in identifying and classifying biosignals.

$$F_{\text{HOC}} = [D_1, D_2, \ldots, D_L], \quad (1 < L < J) \tag{15}$$

where

- F_{HOC} is the HOC-based feature.
- J donates the maximum order of the estimated HOC.

- L is the HOC order used in this study.
- D_1 denotes the number of axis crossings in the zero-mean data series, D_2 denotes the number of axis crossings in the first difference of the series, D_3 denotes the number of axis crossings in the second series and so on.

C. ML models

In this PhD research, a comprehensive analysis of various ML models was undertaken to make predictions or make decisions based on biosignal data. A diverse set of models was employed, including support vector machine (SVM), K-nearest neighbours (KNN), Naive Bayes, linear discriminant analysis (LDA), logistic regression and convolutional neural network (CNN). These models were selected based on their suitability and effectiveness in dealing with biosignal data.

In particular, CNN is a better approach for the time-series classification tasks of biosignal data compared to other algorithms. A CNN is a type of neural network that is particularly well suited for image and video data. Using a CNN for analysing biosignal data treats the biosignal time-series data as a 2D image, where x-axis denotes time and y-axis denotes amplitude values. Then, a CNN can be applied to the 2D image to extract features in the data. The extracted features can then be used to classify the biosignal data into different classes or conditions. CNNs are based on the idea of a 'convolution', which is a mathematical operation that combines input data (such as an image) with a set of learnable filters (also known as kernels or weights) to produce a set of output features. These features are then processed by additional layers in the network to extract increasingly complex representations of the input data.

(1) Architecture of CNN

The architecture of a CNN typically consists of several layers (see Figure 14), including the following:

Convolutional layers that perform convolution operations on the input data to extract features [129].

Pooling layers that reduce the spatial dimensions of the features to reduce the computational cost and introduce some form of translation invariance [129].

- Fully connected layers that are usually added to the end of CNNs to generate global semantic information and perform classification based on the features extracted by previous layers [130]. In simple words, it can be interpreted that fully connected layers take the outputs of the previous layers and use them to make predictions or decisions.

It is noteworthy that the convolution layer and the polling layer can be repeated several or many times. The output of the pooling layer becomes the input of the next convolution layer. Similar to fully connected layers, they can also be stacked.

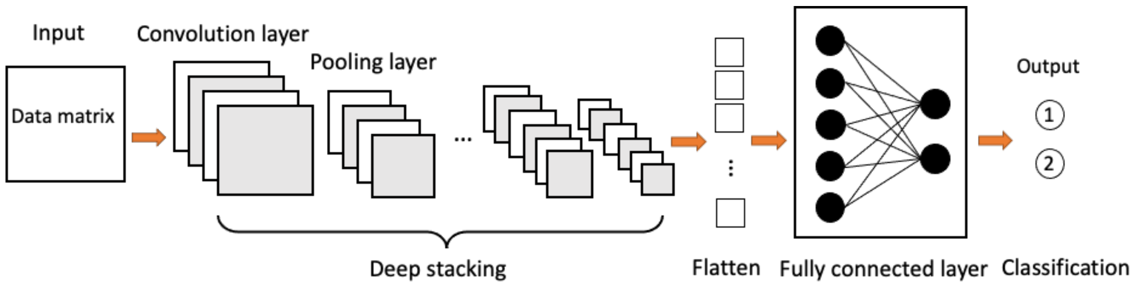


Figure 14: Illustration of a simple CNN for data classification in this work. Note that in deep stacking, layers can be repeated several or many times.

(2) Bayesian optimisation in CNN

In this study, Bayesian optimisation was employed to optimise the hyper-parameters of a CNN to improve its performance on a specific task. Hyper-parameter optimisation is a crucial step in ML, as it seeks to find the optimal set of hyper-parameters for a model that maximises its performance. The study focused on optimising the following hyper-parameters: number of layers, number of filters in each layer, learning rate, batch size and number of training epochs. These hyper-parameters have a significant impact on the performance of the CNN, and finding the optimal values can be challenging and computationally expensive.

Bayesian optimisation is a powerful method for automating the process of hyper-parameter optimisation by efficiently exploring the hyper-parameter space and selecting the best set of hyper-parameters. It is an effective and computationally efficient method for fine-tuning ML algorithms compared to manual grid searches. In this process, the next parameter settings depend on the performance of the previous configurations. The configurations are inferred and decided by the relationship between the hyper-parameter settings and the model performance. Bayesian optimisation for hyper-parameter automatic tuning is a frequently used method for automating the tuning of hyper-parameters, and it was applied in this study for finding an optimal set of hyper-parameters [131], [132].

The process of Bayesian optimisation for CNNs in this study is summarised as follows:

- Initialising the Gaussian process (GP) model with a set of random hyper-parameters and evaluating CNN performance on the validation set
- Using the GP model to predict the performance of the CNN for different hyper-parameter settings
- Using an acquisition function such as Expected Improvement (EI) to decide the next set of hyper-parameters to evaluate
- Evaluating the CNN performance on the validation set for the chosen hyper-parameters and updating the GP model with the new observations
- Repeating the process of predicting, choosing and evaluating until a stopping criterion is met

(3) Implementation

The ML models were implemented using various libraries and frameworks, specifically TensorFlow and Scikit-learn, in the Python programming language. In addition, some aspects of the study also utilised MATLAB to assist in building, training, testing and deploying the ML models. The use of these tools allowed for the efficient and accurate handling of the data, facilitating the analysis and interpretation of the results.

3.4 Summary

This chapter presented the research methodology adopted in this study. The focus of the methodology was on experimental design, data collection and data analysis methods. The objectives, procedures and types of data collected for the three experiments established to answer the research questions were discussed in detail.

The research methodology employed both statistical and ML methods for data analysis. Statistical methods including analysis of variance (ANOVA), Kruskal–Wallis H test, Spearman rank correlation coefficient and Welch Two-Sample t-test were applied. For the analysis of biosignal data, ML methods were primarily used, with a focus on application of convolutional neural networks (CNN) and Bayesian optimisation as innovative methods.

The combination of statistical and ML methods was aimed at providing a comprehensive understanding of the research questions and ensuring the validity of the findings. Thus, this chapter focused on providing a clear understanding of the research methodology, including the data collection procedures, data analysis methods and the techniques used to analyse the data collected.

Chapter 4

4 Results and discussion

This section presents a summary of the research results via a list of academic papers, as follows from the main findings and discussion.

4.1 List of papers

This PhD thesis includes five scientific articles that are completed or published through international scientific journals or various international conferences on related topics. This thesis includes pre-prints of the following five scientific papers and the contributions of the papers are listed in Table 2:

Paper I.
H. Xue, P. Sharma, and F. Wild, "User satisfaction in augmented reality-based training using Microsoft HoloLens," *Computers*, vol. 8, no.1, p. 9, 2019. https://doi.org/10.3390/computers8010009

Paper II.
H. Xue, B.-M. Batalden, and J.-F. Røds, "Development of a SAGAT query and simulator experiment to measure situation awareness in maritime navigation," in: N. Stanton, Eds. *Advances in Human Aspects of Transportation*, AHFE 2020, *Advances in Intelligent Systems and Computing*, vol. 1212. Cham: Springer. https://doi.org/10.1007/978-3-030-50943-9_59

Paper III.
H. Xue, B.-M. Batalden, P. Sharma, J. A. Johansen, and D. K. Prasad, "Biosignals based driving skill classification using machine learning: A case study of maritime navigation," *Applied Sciences*, vol. 11, no. 20, p. 9765, 2021. https://doi.org/10.3390/app11209765

Paper IV.
H. Xue, J. F. Røds, Ø. Haugseggen, A. J. Christensen, B.-M. Batalden, and O.T. Gudmestad, "A study on the effects of rapid training method on ship handling, navigation and decision-making skills under stressful situations".
(Manuscript is submitted to *The Journal of Navigation*.)

Paper V.
H. Xue, Ø. Haugseggen, J.-F. Røds, B.-M. Batalden, and D. K. Prasad, "Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing performance."
(Manuscript is submitted to *Transportation Research Interdisciplinary Perspectives.)*

Table 2: Contribution of the PhD candidate to the papers in the PhD thesis

Contributor role	Paper I	Paper II	Paper III	Paper IV	Paper V
Conceptualisation	✓	✓	✓	✓	✓
Experiment design			✓	✓	✓
Data collection		✓	✓	✓	✓
Formal analysis	✓	✓	✓	✓	✓
Investigation		✓	✓	✓	✓
Methodology	✓	✓	✓	✓	✓
Software	✓	✓	✓	✓	✓
Visualisation	✓	✓	✓	✓	✓
Validation	✓	✓	✓	✓	✓
Writing the manuscript	✓	✓	✓	✓	✓
Submitting and revising	✓	✓	✓	✓	✓

4.2 Research findings

This section summarises the research findings of the five included papers, which can be found in the appendix. The papers are presented together with the research questions that they answer. Paper I studied user satisfaction with wearable technology for hands-on training, while Paper II to Paper V focused on enhanced learning with wearable technology in maritime sector training.

RQ1: Can wearable technology, for example AR technology, be used in hands-on training satisfactorily?

Paper I – User satisfaction in augmented reality-based training using Microsoft HoloLens

When a new technology is introduced into the traditional training field, user satisfaction should be one of the top priorities for research and testing. This research has been addressed in Paper I. This paper investigated user satisfaction in AR technologies applied to training in three different industrial sectors, namely, aeronautics, medicine and astronautics.

AR technologies provide a different user experience than that of, for example mobile phone apps. The user interacts with the surrounding real world, combining inputs from the environment with digital augmentations. Therefore, in the past decade, AR has been increasingly employed for a number of training applications, such as medical education [125], rehabilitation engineering [133], automotive safety [134], task assistance [135] and manufacturing [136]. However, for the successful adoption of AR-based training across different domains, one of the key factors, user satisfaction, has never been well studied.

User satisfaction with AR can be broken down into two components: satisfaction with the interaction and that with the delivery device. In this study, Microsoft HoloLens AR glasses were used to experience an AR application designed for training and learning purposes. This application consisted of the following two modes: *recorder* and *player*. The *recorder* was designed to capture an expert's workplace experience and combine it with the technical documentation associated with a given scenario. The *player* was used to re-enact the scenario to verify the recordings and was usually employed to train a novice for the scenario.

This study had the following specific goals: (1) Test and observe user satisfaction with AR applications and AR glasses. (2) Find if experts and students are satisfied with the prototype application. (3) See if the application can increase interest in learning new skills. (4) Evaluate whether users find the application easy to use.

For these purposes, a questionnaire was implemented as a tool to evaluate and assess users' subjective satisfaction, and later interviews were conducted to better understand the results of the questionnaires. The measures of the questionnaires included the following specific interface factors: screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing and software installation. The results were analysed and reported using descriptive statistics.

From the results of analyses of the questionnaires, it was observed that most participants were satisfied with the AR glasses and applications. It was also found that the system and content helped the participants to accomplish the task quite well, and their attention was captivated in a positive way. In other words, the results show that the user interface was well-designed. When using AR glasses, the user is able to see 'useful information' displayed in close proximity to each relevant part or object in the real world. The main factors of age, gender, education level, roles of the participants and organisations did not have significant effects on satisfaction with using smart glasses and AR applications. However, computer/internet knowledge level did influence user satisfaction. Participants who had better computer/internet knowledge were more satisfied with smart glasses and AR applications. There was no significant interaction between these factors. Since most participants had a moderate or better than a moderate level of knowledge of using computers and the internet, it can be predicted that most educated people can easily accept smart glasses and AR applications.

Overall, the results of this study indicated that using augmented reality glasses for hands-on training is a feasible approach, and that both teachers and learners reported acceptable levels of satisfaction with the approach.

RQ2: How does experience affect maritime navigation tasks?

- **RQ2.1**: Which methods can be used to measure trainees' performance in maritime navigational tasks, specifically SA during maritime navigation, and how does experience impact the navigator's SA?

- **RQ2.2**: Does experience affect the stress levels of seafarers in maritime navigation tasks?

The research questions RQ2 including RQ2.1 and RQ2.2 are answered by Paper II and Paper III.

Paper II – Development of a SAGAT query and simulator experiment to measure situation awareness in maritime navigation

Paper III – Biosignal-based driving skill classification using machine learning: A case study of maritime navigation

This study is the first experiment that focused on research on maritime training. In maritime, many ship collisions and groundings occur due to navigators' erroneous SA. In particular, unsafe acts and pre-conditions for unsafe acts are important causes of ship collisions and groundings [137]. Many researchers have found that 71% of navigators' errors are SA-related problems [138]. Hence, SA is a key skill in maritime training. It is interesting to discover how experience affects maritime navigational tasks.

This experiment compared SA performance and stress levels between experienced navigators and nautical science students. RQ2 was broken down into two secondary research questions: RQ2.1 and RQ2.2, which were addressed in separate papers; Paper II answered research question RQ2.1 about performance assessment methods and the role of experience in SA performance, while Paper III addressed research question RQ2.2 regarding stress levels between experienced navigators and nautical science students based on biosignal data collected by wearable sensors during navigational tasks.

In Paper II, using SAGAT as a tool, a method for measuring SA–MA was developed. The procedure for developing the SAGAT queries can be found in Figure 7. HTA was used to map the navigation and collision avoidance tasks, and the results of simulator experiments and input from subject matter experts were used to determine the SAGAT score. The study found that it was difficult to measure SA–MA, especially for levels 2 and 3 of SA. The results also showed a difference in SA–MA between the experienced navigators and novices (students with less experience). Hence, experience was found to have an effect on a navigator's SA. Therefore, it can be concluded that SA performance can be improved through training and practice.

In Paper III, the focus was on detecting changes in stress levels between experienced navigators and novices. The study aimed to investigate whether differences existed in stress levels between experienced seafarers (experts) and novices (students) as they performed navigation tasks in a simulated maritime environment. The goal was to improve training methods for maritime students and potentially improve safety in the maritime industry by identifying and addressing any differences in stress levels between experts and students. To do this, biosignal data, including electrodermal activity (EDA), body temperature, BVP and HR, were collected from a wearable sensor as indicators of stress levels. These data were then analysed using a ML algorithm called the convolutional neural network (CNN). To validate the study, the results from the CNN analysis were compared to subjective measurements of workload using the NASA Task Load Index (NASA-TLX) tool.

The results of the CNN analysis showed that the biosignal data from the experts could be categorised differently from those of the novices, which was consistent with the results of the NASA-TLX ratings. This suggested that the proposed algorithm was successful in detecting differences in stress levels between experts and novices, as measured by both the biosignal data and NASA-TLX ratings. Here is a summary of the key findings of the study:

- Biosignal data from experienced seafarers (experts) and novices (students) can be classified using a ML algorithm with an accuracy of 75.5%.
- Subjective measurements of workload using the NASA-TLX tool showed differences between experts and students.
- The results of the NASA-TLX ratings showed that experts had a smaller workload than students.
- The study suggested that there may be a relationship between workload, stress and SA in maritime navigation, with experts experiencing less stress and having better SA scores than students.
- The sample size was small; hence, further research with a larger population is needed to confirm these findings.
- The study's results may contribute to the development of an automated assessment system for evaluating SA performance in maritime navigation.

RQ3: How can efficient training progress be made during stressful maritime tasks?

Paper IV – A study on the effects of rapid training method on ship handling, navigation and decision-making skills under stressful situations

This research question is answered mainly in Paper IV. In navigation tasks, decision-making skills are critical for safe sailing [139], especially in tasks such as collision avoidance, where the navigator must decide which means to use for determining the risk of collision and taking appropriate action [140]. In addition, environmental stress is one of the dominant factors that causes accidents at sea [81], [141], [142]. Working at sea is inherently stressful [143]–[145], especially when faced with rapidly changing situations and the need to make many decisions under pressure [146]. Decision-making situations can also increase the individual's stress level and affect decision-making ability under uncertainty by altering the underlying decision-making mechanism [147]. As a result, high stress levels can lead to flawed decision-making, which can be dangerous in a maritime setting.

This study had the following specific goals:

- Determine the learning objectives of the training programme. These should be specific and measurable and should reflect the skills and knowledge that the participants will need to demonstrate to complete the task(s) successfully.
- Investigate the effectiveness of simulator-based training in improving decision-making skills in the maritime industry.
- Investigate the stress levels of the participants during a particular task.
- Assess or evaluate participant learning outcomes.
- Understand how stress affects decision-making.

For these purposes, a project-aimed rapid training programme was developed to focus on the towing operation, and both the subjective and objective stress levels of the participants were measured during the tasks. The study included a simulated critical situation of an engine failure in the tugs at a critical location during which the participants were required to make decisions. A custom decision-quality rating scale was proposed to assess decision-making quality. Therefore, the interactions between workload, stress and decision quality can be determined. Finally, the study compared the outcomes of the rapid training programme with those of the routine training programme in terms of manoeuvring, navigational and decision-making skills.

The key findings of the study are as follows:

- Compared to the experimental group, participants in the control group, who received more routine training, had a higher average score on the decision-quality rating scale and a shorter average time for the back tugboat to cut the line in the emergency situation.
- Participants in the control group perceived higher levels of stress than those in the experiment group, and objective stress levels, as measured by HR, also increased significantly during the towing operation compared to the relaxing time in both groups.
- The NDM model was found to be suitable for analysing the participants' decision-making in this study. From the NDM framework, the main protocol used was the recognition-primed decision-making (RPD) mode. The results suggest that the control group used the RPD mode more frequently, while the experimental group had difficulty following the RPD mode because of their lower level of knowledge and experience.

In summary, this research found that (1) project-aimed rapid training can give participants enough knowledge to make efficient decisions in stressful and critical situations to some degree and (2) different training methods can affect the decision-making model applied by the participants. The findings suggest that participants who had received conventional teaching over a longer period demonstrated a deeper understanding of how to apply their knowledge and skills in unfamiliar and critical situations than those who received rapid project-aimed training. However, for the experimental group that received the rapid training, the decisions they made under time pressure in critical situations were often creative and unconventional, rather than based on recognition-primed decision-making. This suggests that the rapid training approach may encourage learners to explore different decision-making strategies beyond what they have learned, which could be beneficial in certain contexts but may also present challenges. This hypothesis is supported by high number of different solutions from this group. As these solutions are analysed, some that are not in line with either the crash course theory or other simulator-based exercises may be found. For example, a participant may have a risky decision to pass between the front tug and the object being towed (see Figure 15).

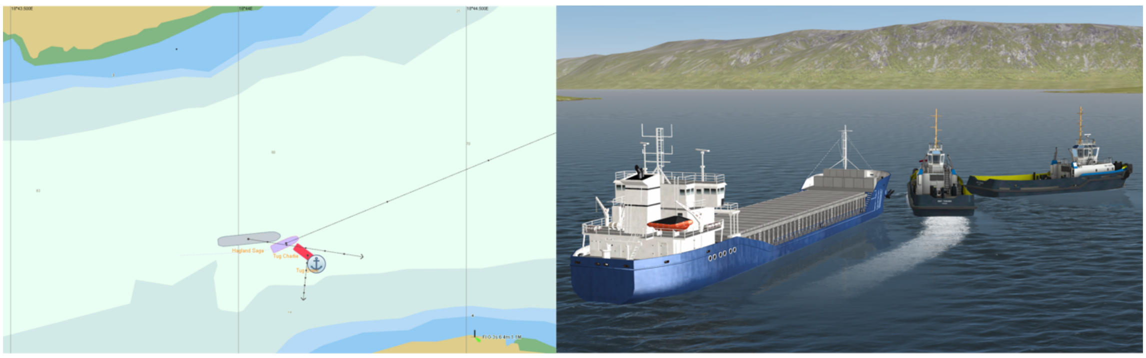


Figure 15: A map of an example of a risky decision and its corresponding 3D view. One of the participants decided to cross the narrow passage between the tugboat and the disabled ship, an action that requires considerable ship-handling skills.

This study made several contributions to the field of maritime education and training (MET):

- First, it provided evidence that project-aimed rapid training can be an effective way to train a certain skill in an efficient and cost-effective way in a safe environment. This is important because maritime training requires high investment and incurs high running costs [24], and the quality of MET is critical for ensuring the safety of life at sea.
- Second, the study used a decision-quality rating scale to evaluate the decision-making of participants, providing a more comprehensive understanding of the decision-making process in the maritime industry. This is valuable, as decision-making is a critical skill for navigators, and it is one of the main causes of maritime accidents.
- Third, the study used subjective and objective stress measures to assess the impact of training on stress levels. This allowed for a more complete understanding of the effect of training on stress in the maritime industry.

Overall, the findings of this study have the potential to inform the development of more effective and efficient training programmes in the maritime industry, which can ultimately improve navigational safety.

RQ4: Can we build a system to analyse the objective stress levels of navigators based on biosignal data so that it can improve training outcomes in maritime training programmes?

Paper V – Assessment of stress levels based on biosignal during simulator-based maritime navigation training and its impact on sailing performance

This research question was addressed in Paper V. The study described the critical impact of stress on safety and training outcomes in the maritime industry. It highlighted the challenges in assessing learning outcomes and performance and the need for objective stress analysis. The study aimed to build a system for analysing the objective stress levels of navigators based on biosignal data, with the goal of improving training outcomes in maritime training programmes. This study explored the relationship between stress and training outcomes in the maritime industry by comparing self-reported stress levels with objective stress levels measured from biosignal data. It also investigated the relationship between stress levels and safety factors and the impact of stress levels on training programmes by assessing learning outcomes and performance. A conceptual model was also proposed in Figure 16, demonstrating the relationship between safety factors and stress levels and highlighting the connection between stress and maritime training programmes.

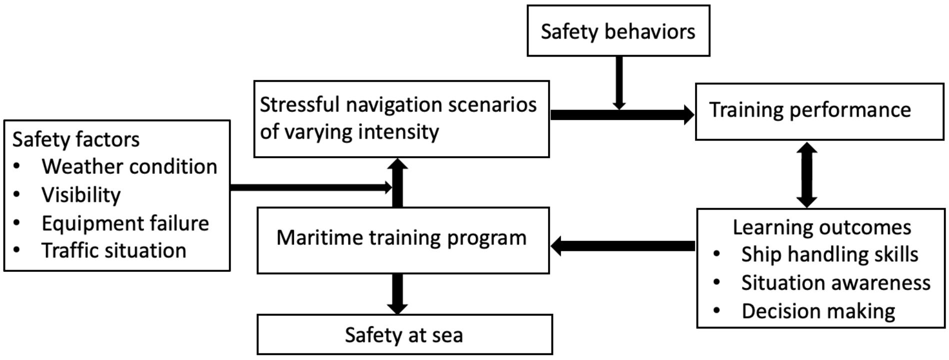


Figure 16: Conceptual model of a stress-based maritime training programme.

In this study, the subjective stress levels of the participants were self-reported using several questionnaires. Statistical methods were employed to analyse the results. Objective stress levels were assessed by analysing HR data obtained from IBI data collected via wearable sensors. Five different ML algorithms were selected to analyse the biosignal data, and their results were compared using three different methods of pre-processing the HR data: using the original data directly, using event-extracted data, and using window mean data. The results, as shown in Figure 17, indicated that when using the event extraction method, all five ML algorithms achieved high accuracy in classifying the stress levels of the participants from a complex training scenario and an easy training scenario. Both subjective stress levels and objective stress levels have good in-line results.

The study determined that the control and experimental scenarios resulted in different levels of stress for the participants, which affected their performance. The impact on training performance was evaluated through the proposed performance criteria and an examination of the deviation from the planned route. Results from the deviation measurements, shown in Figure 18, indicate that most instances of deviation in the control scenario remained within 300 m of the planned route, with the majority between 0 and 200 m away. In contrast, the deviation in the experimental scenario was primarily greater than 200 m. Moreover, participants in the experimental scenario were able to return to the planned route more swiftly, while participants in the control scenario took more time to do so (as shown in Figure 19). The results suggest that the deviation may have been caused by participants starting their turns too late or not turning back towards the planned route quickly enough. In addition, the time taken to return to the planned route after deviation was substantial, with some participants passing Waypoint 4 before returning to or nearing the planned route, which coincided with the point at which they encountered the second fishing vessel, resulting in a closer passing distance. The analysis of performance measures also revealed that the participants were not intense.

In conclusion, the results of the study contribute to a better understanding of the relationship between stress and training outcomes and have the potential to improve safety and optimise training programmes in the maritime industry.

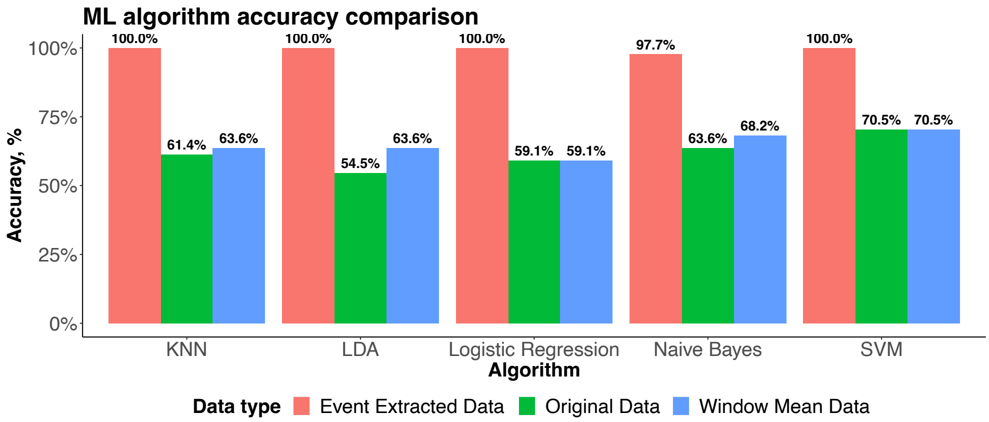


Figure 17: Comparison of machine-learning results from five different algorithms in different ways of pre-processing data.

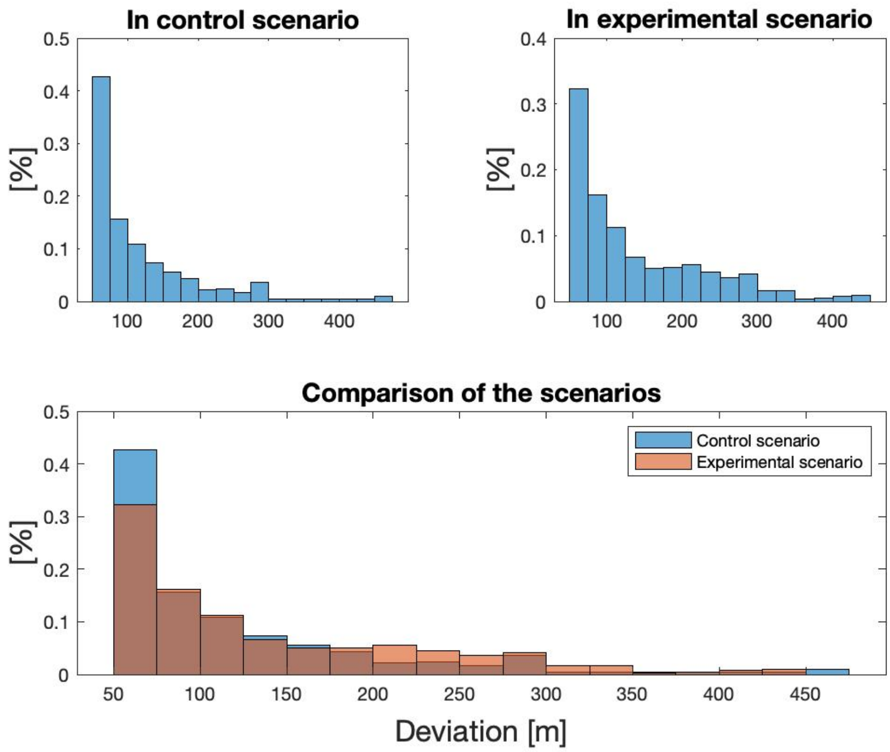


Figure 18: Deviation from the planned route for the experimental scenario (orange) and the control scenario (blue) in the comparison graph. (Experts suggest that the deviation of less than 100 m is negligible.)

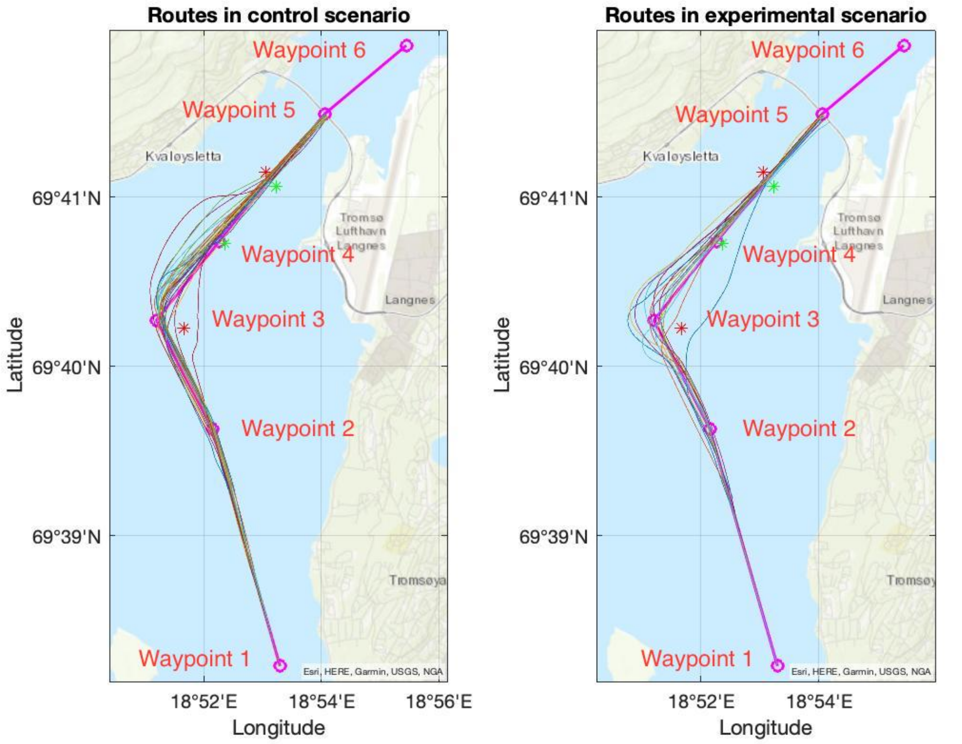


Figure 19: Participants sailed routes from the control scenario (left) to the experimental scenario (right). The magenta lines represent the planned route.

4.3 Discussion

This section discusses the research outcomes from various aspects. It covers the connections between articles, strengths of the work, limitations, and implications. Additionally, it evaluates the contribution of the research to the field. This in-depth discussion provides a deeper understanding of the research outcomes and advances knowledge in the field.

4.3.1 Relationship between the articles

For this research, we sought to limit navigation errors, avoid unsafe acts and improve training and learning methods in the maritime sector. All five papers were built together to form the foundation of this dissertation and answer the research questions proposed in the introduction section. These research questions are connected by their focus on exploring different aspects of training and performance in maritime navigation, with a particular emphasis on using technology and measuring the impact of experience and stress on performance. Together, these research questions aim to provide insight into the optimisation of training and performance within the context of maritime navigation tasks. Paper I addressed the research question whether wearable technology, specifically AR-based training using Microsoft HoloLens, can be used for hands-on training and the satisfaction of users with this technology. This question is important, as it lays the groundwork for the subsequent papers' focus on enhancing learning outcomes in the maritime sector. Papers II–V, supporting RQ2–RQ4, focused on how this type of technology can be used to enhance learning outcomes.

4.3.2 Strengths of the work

The work presented in this study has several strengths, primarily in the innovative and practical solutions proposed for improving safety and enhancing learning outcomes in the maritime industry. In the context of MET, identifying the zone of ZPD [43] of students is useful. A key aspect is pinpointing the challenge point [44] during the training process (see Figure 3). The findings of this study indicate that employing biosignal analysis to measure students' stress levels can assist instructors in providing effective guidance. Moreover, the results show a clear correlation between stress management and training performance. These outcomes underscore the necessity for innovative and effective training methods in the maritime domain (see Figure 4), emphasizing their pivotal role in enhancing the overall quality of maritime training.

In particular, this research investigated the effectiveness of training programmes conducted on bridge simulators, which are a vital component of contemporary MET [89]. Simulations provide an immersive and realistic training environment that mirrors real-life situations, enabling trainees to gain practical experience within a safe and controlled setting that reduces the risk of accidents and injuries [148]. Moreover, the use of maritime training simulators enables trainees to experience a broad range of scenarios that may otherwise be too dangerous or expensive to replicate in real life, thereby enhancing their ability to manage risks and respond effectively to emergency situations [149]. Simulation training also offers trainers an opportunity to assess trainees' performance within a controlled environment, providing valuable feedback for areas requiring improvement [150]. In addition, simulation training can be used to promote teamwork, communication and leadership skills, which are essential in the maritime industry [14].

In addition to the traditional training methods in the simulator bridge, the present study introduced new technological advancements in the field of MET, including the application of wearable technology, biosignal data analysis and ML techniques. Objective data collection methods, such as surveys, experiments and case studies, provided a comprehensive understanding of the effectiveness of wearable technology in this context. Incorporating ML techniques and statistical methods into data analysis strengthened the research outcomes. The use of the SAGAT query as an objective tool for measuring situational awareness enhanced the study's ability to identify the impact of experience on navigational tasks and, ultimately, to improve training outcomes.

Finally, the study presented practical implications for the maritime industry, including the adoption of rapid training methods and a decision-quality rating scale to objectively evaluate participants' decision-making processes, providing a more comprehensive understanding of decision-making in the maritime industry. Thus, this research provided valuable insights and practical solutions for enhancing safety and learning outcomes in the maritime industry.

4.3.3 Limitations of the work

Despite the promising findings, this work has certain limitations. First, the sample size used in the study was relatively small, which may restrict the generalisability of the results. Therefore, further research with a larger population is necessary to validate the findings and ensure their applicability to a broader range of individuals. Moreover, it is noteworthy that all data utilised in the maritime training experiments were obtained from a single university, specifically from a cohort of students. Hence, incorporating data from multiple universities would not only strengthen the study's outcomes but also lend greater credence to its conclusions. Unfortunately, due to the logistical challenges associated with the ongoing pandemic, procuring additional data from other academic institutions was not feasible.

Second, it should be noted that the experimental scenario only captures a snapshot of the entire maritime education programme scenario. Conducting a longitudinal study, which follows the same group of participants over an extended period, could have provided more comprehensive and accurate results. This is because a longitudinal study is useful for studying changes and development in individuals and allows for the measurement of changes over time. However, due to constraints on the resources and time available for the PhD project, it was not feasible to conduct such a study. In addition, technical issues, such as limited battery life and the high cost of acquiring an adequate number of sensors, further complicated the implementation of a longitudinal study.

Third, while the current study provides a foundation for understanding the potential use of wearable sensors for evaluating stress levels and training performance in maritime navigation, further investigations are required to develop an automated assessment system to adjust training intensity based on the trainee's stress levels. However, due to constraints associated with the PhD project, this task had to be left for future work.

Moreover, the study only utilised one type of wearable device in the maritime sector, which raises questions about the sufficiency of this approach. Incorporating additional wearable sensors could provide more comprehensive and nuanced insights into training performance in this context. A multi-sensor framework beyond a single type of wearable sensor could enhance training performance by providing a more accurate and holistic evaluation of the trainee's performance. Regrettably, due to the limited time and budgetary constraints associated with the PhD project, it was not feasible to include multiple wearable sensors in the current study.

4.3.4 Implications of the work

The findings of this study suggest that wearable technology, such as AR technology and biosignal monitoring, can be effective tools for enhancing learning outcomes and improving performance and safety in the maritime industry. In the realm of MET, universities and institutions follow a study plan based on the STCW convention set by the IMO. As illustrated in Figure 20, the compliance matrix is an administrative document developed by universities or institutions in accordance with the STCW convention. The compliance matrix describes how the MET study programmes fulfil the requirements from the STCW code, including the knowledge, understanding and skills required for each competency, the methods needed to demonstrate competence, the criteria necessary for competence evaluation, the topics covered by the competency and the assessment methods utilised. The compliance matrix is certified by the Norwegian Maritime Authority (NMA), and it is a vital part of the certification for universities or institutions to commence maritime training programmes. Course plans are then designed according to the compliance matrix of universities or institutions, and lecturers or instructors develop lecture plans that can vary depending on individual instructors or lecturers.

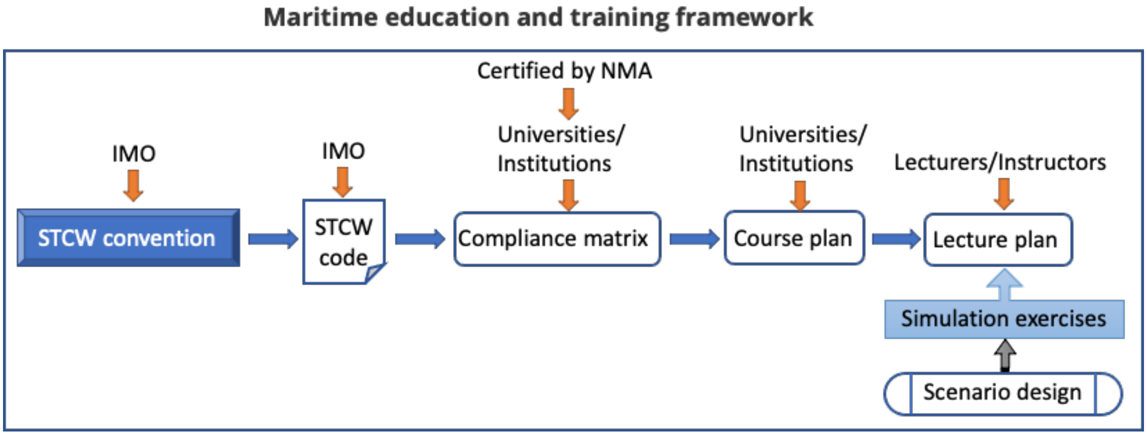


Figure 20: An example of maritime education and training hierarchical work. Note that IMO stands for International Maritime Organization, NMA stands for Norwegian Maritime Authority and STCW stands for International Convention on Standards of Training, Certification and Watchkeeping for Seafarers.

Enhancing simulation exercises and training scenarios can lead to more efficient training outcomes. For example, the MET programme at UiT mandates that each student complete eight simulation training exercises every semester during the first 2 years of study, resulting in 32 simulation training exercises. The duration of each simulation exercise can range from 0.5 to 4 hours, depending on its complexity. While instructors design and assess the content of these exercises before use, many of them might not have been updated for years, even if different instructors had taken over the courses.

Despite the established protocols, designing effective simulation exercises and training scenarios that cater to the needs of all students is a challenging task due to the varying mental capacities and learning paces of individual students. Instructors are required to identify students who are unable to meet the training requirements and gain an understanding of the factors that contribute to their inability to achieve the desired training outcomes. This process requires a deep understanding of the cognitive and behavioural aspects of learning and the ability to tailor training approaches to suit individual needs. In particular, instructors should be able to identify and address individual learning difficulties as well as recognise when a student is not receiving new knowledge or learning effectively. Furthermore, a good instructor may possess extensive sea experience but not necessarily a strong academic background. For instance, they may know how to design a training scenario to improve students’ performance in complex situations, but they may not comprehend certain terms when asked to train students’ in SA or decision-making skills. Therefore, the current study can help improve MET in the following areas:

The use of wearable sensors and the analysis of biosignal data can offer insight into the stress levels of students and guide instructors in adjusting the training intensity to optimise learning outcomes. These objective measures can enhance the effectiveness of training and improve performance and safety in the maritime industry. Thus, this study contributes to the development of a more thorough understanding of the relationship between stress and training outcomes, with the aim of improving the performance and safety of maritime training.

In addition, the inclusion of the SAGAT query in maritime training can assist instructors in identifying students' deficiencies during navigation tasks. SAGAT query is a well-established and objective method of assessing situational awareness, which is crucial for decision-making in complex and dynamic environments [151]. Although it has traditionally been used in aviation training [152], a modified version of SAGAT query can be adapted to maritime training. SAGAT queries are focused on specific data or data criteria that correspond to the three levels of situational awareness (perception, comprehension, and projection). By measuring and analysing students' scores at each level, instructors can identify their weaknesses and tailor their training accordingly. The use of SAGAT in MET can thus enhance the effectiveness of training and contribute to the development of more competent maritime professionals.

Furthermore, the research proposes approaches for delivering project-aimed rapid training that enables seafarers to quickly acquire the necessary knowledge and skills for their specific roles. By reducing the time required for training, this approach allows students to gain on-the-job experience faster, thus reducing costs and significantly improving the efficiency of maritime training programmes. Based on the current results, instructors can design the training scenario in the following two steps: the first step focuses on project-aimed skills acquisition, while the second step aims to improve the weak areas of students' learning.

Finally, the study emphasises the importance of user satisfaction in the adoption and success of AR and wearable technology in training and learning. AR glasses can enhance the effectiveness of training programmes and increase user engagement.

In summary, this dissertation offers innovative solutions for enhancing learning outcomes in the maritime industry, leading to improved performance outcomes, enhanced safety and better training programmes. Furthermore, this research provides valuable insights and implications for the broader field of training and learning.

4.3.5 Contributions to the research field

The research outlined in this dissertation makes several contributions to the field of enhanced learning in the maritime sector. Currently, in MET, performance assessment is often evaluated subjectively by instructors, which can lead to unreliable, invalid and unfair evaluations [153]. The research offers methods and tools used in a new way for measuring training performance, SA, decision-making skills and stress levels, such as the SAGAT query and ML-based biosignal classification [151], [152]. These methods provide objective data for assessing performance and identifying areas for improvement in training programmes, resulting in more effective and personalised training plans.

Stress is a common concern in the maritime domain, as it is often seen as a harmful factor for safety. Stress can negatively impact maritime navigation safety by distracting attention, memory retrieval, and decision-making [154]. Although many researchers have studied stress, the research presented in this dissertation takes a unique approach by investigating the 'comfort zone', where the level of learning and response is optimal [50]. As mentioned in Section 2.2, it is common to use self-reported stress levels. Some researchers have tried to quantify stress levels objectively by measuring cortisol levels in saliva or blood samples [100]. However, these methods are either unreliable or too complicated and inefficient. In this PhD work, we used biosignal data to obtain objective stress levels. By analysing biosignal data, researchers and practitioners can gain a deeper understanding of the physiological responses to stress and develop more effective strategies for managing stress in the workplace. The research provides a basis for developing personalised training plans based on objective data on performance and stress levels. This approach can lead to more effective training programmes that focus on improving individual skills and addressing specific areas for improvement. The use of biosignal data analysis to objectively measure stress levels is a novel and useful contribution to the field.

In this dissertation, deep learning has proven to be a powerful tool for generating meaningful representations of data. Particularly when dealing with small sample data, deep learning, coupled with HOC feature selection, can effectively address this challenge. Other ML algorithms can also provide good accuracy when the data are appropriately processed.

The objective of this dissertation is to enhance the efficiency and effectiveness of maritime training programmes using cutting-edge technology and data-driven approaches. By combining AR technology, wearable sensors, rapid training techniques and ML models, the proposed methods have the potential to improve learning outcomes and assist navigators in handling complex situations in the maritime sector, thereby leading to improved safety and optimised training outcomes.

Chapter 5

5 Conclusion and future work

5.1 Concluding remarks

The scope of the work presented in this dissertation is defined by the four research questions described in the introduction. How the conclusions that can be drawn from the five papers supporting this dissertation provide an answer to these research questions is explained below.

RQ1: Can wearable technology, for example AR technology, be used in hands-on training satisfactorily?

The conclusion that can be drawn from the work presented in Paper I is that wearable technology, specifically AR-based training using Microsoft HoloLens, can be used for hands-on training and the satisfaction of users with this technology. This question is important as it lays the groundwork for the subsequent papers' focus on enhancing learning outcomes in the maritime sector.

RQ2: How does experience affect maritime navigation tasks?

- **RQ2.1**: What methods can be used to measure trainees' performance in maritime navigational tasks, specifically SA, during maritime navigation, and how does the experience impact the navigator's SA?
- **RQ2.2**: Does experience affect the stress levels of seafarers in maritime navigation tasks?

The conclusion that can be drawn from Papers II and III is that the use of the SAGAT query to measure SA is challenging but achievable. Training and practice can improve SA performance. The study suggests that there may be a relationship between workload, stress and SA in maritime navigation, with experts experiencing less stress and having better SA scores than novices (students). However, the SAGAT query can be improved for measuring the higher level of SA, and larger studies are needed to confirm these findings. Nevertheless, the study's results may contribute to the development of an automated assessment system for evaluating SA performance in maritime navigation, which can ultimately enhance safety and efficiency in the maritime industry.

RQ3: How can efficient training progress be made during stressful maritime tasks?

The conclusion that can be drawn from Paper IV is that project-aimed rapid training can give participants enough knowledge to some degree to make efficient decisions in stressful and critical situations.

Different training methods can affect the decision-making models applied by the participants. Participants who have received conventional teaching over a longer period are able to apply their knowledge and skills at a deeper level when faced with unfamiliar and critical situations compared to those who have received project-aimed rapid training. In addition, the study contributes to the field of MET by demonstrating the effectiveness of project-aimed rapid training, using a decision-quality rating scale to evaluate decision-making and assessing the impact of training on stress levels through subjective and objective stress measures.

RQ4: Can we build a system to analyse the objective stress levels of navigators based on biosignal data so that it can improve training outcomes in maritime training programmes?

The conclusion drawn from Paper V is that biosignal data can be used to objectively assess stress levels in navigators during maritime training programmes and this assessment can have a positive impact on training outcomes and performance. The combination of ML algorithms for analysing biosignal data and statistical methods for analysing questionnaire data allowed for a comprehensive analysis of stress levels and their impact on training outcomes.

More generally, the work presented in this dissertation has contributed to the development of methods for enhanced learning to handle complex situations in the maritime sector. The studies presented in this dissertation have advanced our understanding of the relationship between stress and training outcomes in the maritime industry. It concludes that wearable technology can be used in hands-on training and user satisfaction is generally positive. In addition, the study supports the idea that experience affects maritime navigational tasks, such as SA. The results demonstrated the importance of considering stress levels in maritime training programmes, as stress can significantly impact safety and performance. Also, the proposed conceptual model in Paper V highlights the relationship between stress and safety factors and provides a framework for future research in this area.

5.2 Future work

In the future, it would be valuable to continue exploring the relationship between stress and training outcomes and further develop and test biosignal data-based training systems. The model is shown in Figure 21.

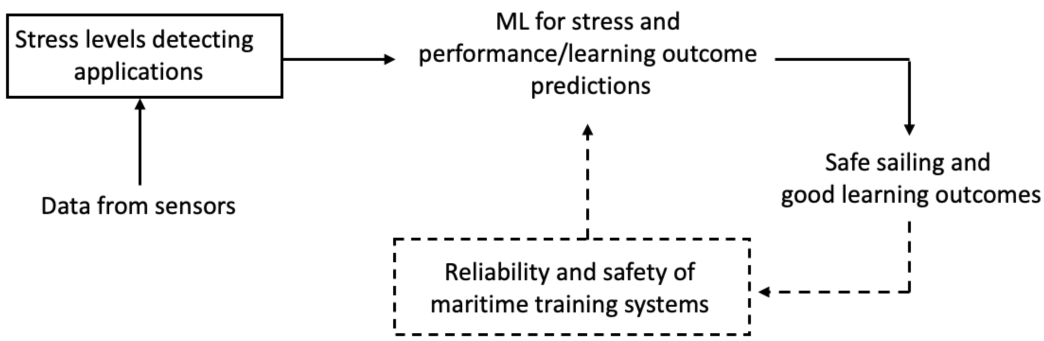


Figure 21 : Future work for a reliable and safe maritime training system.

Biosignal data-based training systems comprise a relatively new approach to improving the performance and safety of maritime training. The system uses various physiological signals, such as HR, skin conductance, and respiration rate, to measure trainees' stress levels during training sessions. These signals are then used to provide real-time feedback to the trainees and instructors, allowing them to adjust the training programme according to the trainee's stress level.

One of the main advantages of biosignal data-based training systems is that they provide a more objective measure of stress levels than traditional self-reporting methods. This allows for a more accurate assessment of stress levels and can help instructors adapt the training programme to better suit the needs of each trainee. In addition, these systems can help improve maritime training safety by identifying when trainees may be at an increased risk of making errors due to high stress levels. This information can then be used to adjust the training programme, such as reducing the complexity of the task or providing more support, to reduce the risk of errors and accidents. Furthermore, by providing real-time feedback, these systems can help trainees develop better stress management strategies and improve their overall performance. This can be achieved by providing training in stress management techniques, such as relaxation techniques or cognitive-behavioural strategies, which can help to reduce stress levels and improve performance.

The process of implementing such a system, including the development of a real-time stress-level-detecting application and field testing in various scenarios with sufficient biosignal data, is left for future work.

References

[1] M. Prussi, N. Scarlat, M. Acciaro, and V. Kosmas, "Potential and limiting factors in the use of alternative fuels in the European maritime sector", *J. Clean. Prod.*, vol. 291, p. 125849, Apr. 2021, doi: 10.1016/j.jclepro.2021.125849.

[2] "'Shipping: indispensable to the world' selected as World Maritime Day theme for 2016". https://imopublicsite.azurewebsites.net/en/MediaCentre/PressBriefings/Pages/47-WMD-theme-2016-.aspx (accessed May 09, 2023).

[3] J. Oh, S. Park, and O.-S. Kwon, "Advanced navigation aids system based on augmented reality", *Int. J. E-Navig. Marit. Econ.*, vol. 5, pp. 21–31, Dec. 2016, doi: 10.1016/j.enavi.2016.12.002.

[4] K. Wróbel, J. Montewka, and P. Kujala, "Towards the assessment of potential impact of unmanned vessels on maritime transportation safety", *Reliab. Eng. Syst. Saf.*, vol. 165, pp. 155–169, Sep. 2017, doi: 10.1016/j.ress.2017.03.029.

[5] B. Wu, T. Cheng, T. L. Yip, and Y. Wang, "Fuzzy logic based dynamic decision-making system for intelligent navigation strategy within inland traffic separation schemes", *Ocean Eng.*, vol. 197, p. 106909, Feb. 2020, doi: 10.1016/j.oceaneng.2019.106909.

[6] M. E. Manuel, "Vocational and academic approaches to maritime education and training (MET): Trends, challenges and opportunities", *WMU J. Marit. Aff.*, vol. 16, no. 3, pp. 473–483, Sep. 2017, doi: 10.1007/s13437-017-0130-3.

[7] Y. Lau and A. K. Y. Ng, "The motivations and expectations of students pursuing maritime education", *WMU J. Marit. Aff.*, vol. 14, no. 2, pp. 313–331, Oct. 2015, doi: 10.1007/s13437-015-0075-3.

[8] A. G. Seamon, "12 The Transformation of ship handling and navigation training", in *Human Factors in Simulation and Training*, Boca Raton: CRC Press, 2008, pp. 231–238.

[9] 'Introduction to IMO'. https://www.imo.org/en/About/Pages/Default.aspx (accessed Feb. 21, 2023).

[10] IMO, *STCW (Standards of Training, Certification, & Watchkeeping for Seafarers)*, 4th ed., 2017 edition. London: International Maritime Organization, 2018.

[11] J. Lisowski, H. Śniegocki, S. Rajamaki, and T. Salmenhaara, *The Finnish and Polish credit transfer systems for maritime studies*. 1999.

[12] T. Kim and S. Mallam, "A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations", *WMU J. Marit. Aff.*, vol. 19, no. 2, pp. 163–181, Jun. 2020, doi: 10.1007/s13437-020-00203-1.

[13] B. Belev, M. Dimitrova, and A. Meczkowska-Christiansen, "Refresher training in maritime qualification", *IOP Conf. Ser. Earth Environ. Sci.*, vol. 172, no. 1, p. 012022, Jun. 2018, doi: 10.1088/1755-1315/172/1/012022.

[14] A. M. Wahl and T. Kongsvik, "Crew resource management training in the maritime industry: A literature review", *WMU J. Marit. Aff.*, vol. 17, no. 3, pp. 377–396, Sep. 2018, doi: 10.1007/s13437-018-0150-7.

[15] T. Kim *et al.*, "The continuum of simulator-based maritime training and education", *WMU J. Marit. Aff.*, vol. 20, no. 2, pp. 135–150, Jun. 2021, doi: 10.1007/s13437-021-00242-2.

[16] G. Emad and W. M. Roth, "Contradictions in the practices of training for and assessment of competency: A case study from the maritime domain", *Educ. Train.*, vol. 50, no. 3, pp. 260–272, Jan. 2008, doi: 10.1108/00400910810874026.

[17] M. Mazzarino, "Costs and financing of maritime education and training in Europe: analysis and policy implications", *Transit. Stud. Rev.*, vol. 12, no. 1, pp. 147–160, Jul. 2005, doi: 10.1007/s11300-005-0042-3.

[18] S. Vujičić, N. Hasanspahić, A. Gundić, and L. Maglić, "Analysis of factors influencing the effectiveness of MET instructors", *WMU J. Marit. Aff.*, vol. 21, no. 4, pp. 549–570, Dec. 2022, doi: 10.1007/s13437-022-00271-5.

[19] A. Alop, "The challenges of the digital technology era for maritime education and training", in *2019 Eur. Navi. Conf. (ENC)*, Apr. 2019, pp. 1–5. doi: 10.1109/EURONAV.2019.8714176.

[20] E. M. Smith Johnson, "Exploring the effects of technology and innovation on changing market requirements and the evolving maritime curriculum: A Jamaican perspective", *Worldw. Hosp. Tour. Themes*, vol. 12, no. 1, pp. 69–79, Jan. 2020, doi: 10.1108/WHATT-10-2019-0065.
[21] H. A. Oltedal and M. Lützhöft, "The human contribution", in *Managing Maritime Safety*, Routledge, 2018.
[22] J. Reason, *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Boca Raton, FL: CRC Press, 2017.
[23] G. Makransky and S. Klingenberg, "Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-WEIRD sample", *J. Comput. Assist. Learn.*, vol. 38, no. 4, pp. 1127–1140, 2022, doi: 10.1111/jcal.12670.
[24] E. Markopoulos, J. Lauronen, M. Luimula, P. Lehto, and S. Laukkanen, "Maritime safety education with VR Technology (MarSEVR)", in *2019 10th IEEE Int. Conf. Cog. Infocom (CogInfoCom)*, pp. 283–288, 2019. doi: 10.1109/CogInfoCom47531.2019.9089997.
[25] H. Kim, Y.-T. Kwon, H.-R. Lim, J.-H. Kim, Y.-S. Kim, and W.-H. Yeo, "Recent advances in wearable sensors and integrated functional devices for virtual and augmented reality applications", *Adv. Funct. Mater.*, vol. 31, no. 39, p. 2005692, 2021, doi: 10.1002/adfm.202005692.
[26] F. Sanfilippo, "A multi-sensor fusion framework for improving situational awareness in demanding maritime training", *Reliab. Eng. Syst. Saf.*, vol. 161, pp. 12–24, May 2017, doi: 10.1016/j.ress.2016.12.015.
[27] 'Wearable Experience for Knowledge Intensive Training | WEKIT Project | Fact Sheet | H2020', *CORDIS | European Commission*. https://cordis.europa.eu/project/id/687669 (accessed Dec. 21, 2022).
[28] E. Demirel, "Maritime education and training in the digital era", *Univers. J. Educ. Res.*, vol. 8, no. 9, pp. 4129–4142, Sep. 2020, doi: 10.13189/ujer.2020.080939.
[29] V. Gekara, "Understanding attrition in UK maritime education and training", *Glob. Soc. Educ.*, vol. 7, no. 2, pp. 217–232, Jun. 2009, doi: 10.1080/14767720902908190.
[30] K. I. Spenner, "Skill: Meanings, methods, and measures", *Work Occup.*, vol. 17, no. 4, pp. 399–421, Nov. 1990, doi: 10.1177/0730888490017004002.
[31] D. Sampson and D. Fytros, "Competence models in technology-enhanced competence-based learning", in *Handbook on Information Technologies for Education and Training*, H. H. Adelsberger, Kinshuk, J. M. Pawlowski, and D. G. Sampson, Eds., in International Handbooks on Information Systems. Berlin, Heidelberg: Springer, pp. 155–177, 2008. doi: 10.1007/978-3-540-74155-8_9.
[32] P. H. Cheney, D. P. Hale, and G. M. Kasper, "Knowledge, skills and abilities of information systems professionals: past, present, and future", *Inf. Manage.*, vol. 19, no. 4, pp. 237–247, Nov. 1990, doi: 10.1016/0378-7206(90)90033-E.
[33] E. A. Fleishman, D. P. Costanza, and J. Marshall-Mies, "Abilities", in *An occupational information system for the 21st century: The development of O*NET*, Washington, DC, US: American Psychological Association, pp. 175–195, 1999. doi: 10.1037/10313-010.
[34] R. E. Snow and D. F. Lohman, "Toward a theory of cognitive aptitude for learning from instruction.", *J. Educ. Psychol.*, vol. 76, no. 3, p. 347, 1984, doi: 10.1037/0022-0663.76.3.347.
[35] A. Sharma and T. Kim, "Exploring technical and non-technical competencies of navigators for autonomous shipping", *Marit. Policy Manag.*, vol. 49, no. 6, pp. 831–849, Aug. 2022, doi: 10.1080/03088839.2021.1914874.
[36] V. V. Thai and G.-T. Yeo, "Perceived competencies required for container shipping logisticians in Singapore and South Korea", *Int. J. Logist. Manag.*, vol. 26, no. 2, pp. 334–355, Jan. 2015, doi: 10.1108/IJLM-02-2014-0031.
[37] J. Hill, T. Brand, T. Dobbins, and Str. Ltd, "The training system: Developing crew competence within the virtual and real world", *Hum. Factors*, 2016.
[38] A. N. Leont'ev, "On Vygotsky's creative development", in *The Collected Works of L. S. Vygotsky: Problems of the Theory and History of Psychology*, R. W. Rieber and J. Wollock, Eds., in Cognition and Language. Boston, MA: Springer US, pp. 9–32, 1997. doi: 10.1007/978-1-4615-5893-4_2.
[39] "Vygotsky's zone of proximal development and scaffolding", Nov. 03, 2022. https://www.simplypsychology.org/zone-of-proximal-development.html (accessed Apr. 27, 2023).

[40] Y. Abtahi, "Who/what is the more knowledgeable other?", *For the Learning of Mathematics,* vol.34, no.3, pp.14-15, 2014.

[41] M. Cicconi, "Vygotsky meets technology: A reinvention of collaboration in the early childhood mathematics classroom", *Early Child. Educ. J.*, vol. 42, no. 1, pp. 57–65, Jan. 2014, doi: 10.1007/s10643-013-0582-9.

[42] V. K. Zaretskii, "The zone of proximal development", *J. Russ. East Eur. Psychol.*, vol. 47, no. 6, pp. 70–93, Nov. 2009, doi: 10.2753/RPO1061-0405470604.

[43] S. Chaiklin, "The zone of proximal development in Vygotsky's analysis of learning and instruction", in *Vygotsky's Educational Theory in Cultural Context*, A. Kozulin, B. Gindis, S. M. Miller, and V. S. Ageyev, Eds., in Learning in Doing: Social, Cognitive and Computational Perspectives. Cambridge: Cambridge University Press, pp. 39–64, 2003. doi: 10.1017/CBO9780511840975.004.

[44] M. A. Guadagnoli and T. D. Lee, "Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning", *J. Mot. Behav.*, vol. 36, no. 2, pp. 212–224, Jul. 2004, doi: 10.3200/JMBR.36.2.212-224.

[45] R. G. Marteniuk, "7 - Cognitive information processes in motor short-term memory and movement production", in *Motor Control*, G. E. Stelmach, Ed., Academic Press, pp. 175–186, 1976. doi: 10.1016/B978-0-12-665950-4.50012-2.

[46] M. A. Guadagnoli and C. P. Bertram, "Optimizing practice for performance under pressure", *Int. J. Golf Sci.*, vol. 3, no. 2, pp. 119–127, Dec. 2014.

[47] C. E. Morris *et al.*, "Effect of a simulated tactical occupation stressor and task complexity on mental focus and related physiological parameters", *Int. J. Ind. Ergon.*, vol. 66, pp. 200–205, Jul. 2018, doi: 10.1016/j.ergon.2018.03.006.

[48] M. M. Maymand, F. Shakhsian, and F. S. Hosseiny, "The effect of stress on flight performance", 2012.

[49] B. Mayer, S. Helm, M. Barnett, and M. Arora, "The impact of workplace safety and customer misbehavior on supermarket workers' stress and psychological distress during the COVID-19 pandemic", *Int. J. Workplace Health Manag.*, vol. 15, no. 3, pp. 339–358, Jan. 2022, doi: 10.1108/IJWHM-03-2021-0074.

[50] P. Hancock and J. Warm, "A dynamic model of stress and sustained attention", *J. Hum. Perform. Extreme Environ.*, vol. 7, no. 1, Jun. 2003, doi: 10.7771/2327-2937.1024.

[51] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation", *J. Comp. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, 1908, doi: 10.1002/cne.920180503.

[52] R. A. Cohen, "Yerkes–Dodson law", in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan, Eds., New York, NY: Springer, pp. 2737–2738, 2011. doi: 10.1007/978-0-387-79948-3_1340.

[53] P. G. Nixon, "The human function curve. With special reference to cardiovascular disorders: part I", *The Practitioner*, vol. 217, no. 1301, pp. 765–770, Nov. 1976.

[54] R. Frank and R. Kodikal, "Role stress study: An effective tool for employee engagement", Aug. 2017.

[55] J. R. Jepsen, Z. Zhao, C. Pekcan, M. Barnett, and W. M. A. van Leeuwen, "Risk factors for fatigue in shipping, the consequences for seafarers' health and options for preventive intervention", in *Maritime Psychology: Research in Organizational & Health Behavior at Sea*, M. MacLachlan, Ed., Cham: Springer International Publishing, pp. 127–150, 2017. doi: 10.1007/978-3-319-45430-6_6.

[56] S. K. Basak, "A framework on the factors affecting to implement maritime education and training system in educational institutions: A review of the literature", *Procedia Eng.*, vol. 194, pp. 345–350, Jan. 2017, doi: 10.1016/j.proeng.2017.08.155.

[57] P. A. Hancock, "In defense of the maximal adaptability model", *Physiol. Behav.*, vol. 252, p. 113844, Aug. 2022, doi: 10.1016/j.physbeh.2022.113844.

[58] P. A. Hancock and I. Vasmatzidis, "Effects of heat stress on cognitive performance: the current state of knowledge", *Int. J. Hyperthermia*, vol. 19, no. 3, pp. 355–372, Jan. 2003, doi: 10.1080/0265673021000054630.

[59] G. R. Emad and S. Ghosh, "Identifying essential skills and competencies towards building a training framework for future operators of autonomous ships: a qualitative study", *WMU J. Marit. Aff.*, Apr. 2023, doi: 10.1007/s13437-023-00310-9.

[60] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems", *Hum. Factors*, vol. 37, no. 1, pp. 32–64, Mar. 1995, doi: 10.1518/001872095779049543.

[61] M. R. Endsley and D. J. Garland, "Theoretical underpinnings of situation awareness: A critical review", *Mahwah NJ: Lawrence Erlbaum Assoc.*, vol. 1, no. 1, pp. 3–21, 2000.

[62] C. Chauvin, J. P. Clostermann, and J.-M. Hoc, "Situation awareness and the decision-making process in a dynamic situation: avoiding collisions at sea", *J. Cogn. Eng. Decis. Mak.*, vol. 2, no. 1, pp. 1–23, Mar. 2008, doi: 10.1518/155534308X284345.

[63] A. Sharma, S. Nazir, and J. Ernstsen, "Situation awareness information requirements for maritime navigation: A goal directed task analysis", *Saf. Sci.*, vol. 120, pp. 745–752, Dec. 2019, doi: 10.1016/j.ssci.2019.08.016.

[64] E. Salas, C. Prince, D. P. Baker, and L. Shrestha, "Situation awareness in team performance: implications for measurement and training", *Hum. Factors*, vol. 37, no. 1, pp. 123–136, Mar. 1995, doi: 10.1518/001872095779049525.

[65] C. Chauvin, J. P. Clostermann, and J.-M. Hoc, "Impact of training programs on decision-making and situation awareness of trainee watch officers | Elsevier Enhanced Reader", *Safety Science*, vol. 47, no. 9, pp. 1222–1231, 2009, doi: 10.1016/j.ssci.2009.03.008.

[66] A. Lee Chang *et al.*, "Comparison between simulation-based training and lecture-based education in teaching situation awareness. A randomized controlled study", *Ann. Am. Thorac. Soc.*, vol. 14, no. 4, pp. 529–535, Apr. 2017, doi: 10.1513/AnnalsATS.201612-950OC.

[67] E.-R. Saus, B. H. Johnsen, J. Eid, P. K. Riisem, R. Andersen, and J. F. Thayer, "The effect of brief situational awareness training in a police shooting simulator: An experimental study", *Mil. Psychol.*, vol. 18, no. sup1, pp. S3–S21, Jan. 2006, doi: 10.1207/s15327876mp1803s_2.

[68] C. Chauvin, S. Lardjane, G. Morel, J.-P. Clostermann, and B. Langard, "Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS", *Accid. Anal. Prev.*, vol. 59, pp. 26–37, Oct. 2013, doi: 10.1016/j.aap.2013.05.006.

[69] M. R. Endsley and D. J. Garland, *Situation Awareness Analysis and Measurement*. CRC Press, 2000.

[70] R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design", in *Situational Awareness*, Routledge, pp. 111-128, 2017.

[71] W. L. Waag and M. R. Houck, "Tools for assessing situational awareness in an operational fighter environment", *Aviat. Space Environ. Med.*, vol. 65, pp. A13–A19, 1994.

[72] M. R. Endsley, "Measurement of situation awareness in dynamic systems", *Hum. Factors*, vol. 37, no. 1, pp. 65–84, Mar. 1995, doi: 10.1518/001872095779049499.

[73] M. A. Vidulich, "Testing the sensitivity of situation awareness metrics in interface evaluations", in *Situation Awareness Analysis and Measurement*, CRC Press, p. 230, 2000.

[74] A. R. Pritchett and R. J. Hansman, "Use of testable responses for performance-based measurement of situation awareness", in *Situation Awareness Analysis and Measurement*, CRC Press, p. 189, 2000.

[75] P. Salmon, N. Stanton, G. Walker, and D. Green, "Situation awareness measurement: A review of applicability for C4i environments", *Appl. Ergon.*, vol. 37, no. 2, pp. 225–238, Mar. 2006, doi: 10.1016/j.apergo.2005.02.001.

[76] S. Hasanzadeh, B. Esmaeili, and M. D. Dodd, "Measuring construction workers' real-time situation awareness using mobile eye-tracking", pp. 2894–2904, May 2016, doi: 10.1061/9780784479827.288.

[77] P. O'Meara *et al.*, "Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper", *Int. Emerg. Nurs.*, vol. 23, no. 2, pp. 94–99, Apr. 2015, doi: 10.1016/j.ienj.2014.11.001.

[78] M. R. Endsley, "Situation awareness in aviation systems", in *Handbook of aviation human factors*, in Human factors in transportation. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 257–276, 1999.

[79] M. C. Wright, J. M. Taekman, and M. R. Endsley, "Objective measures of situation awareness in a simulated medical environment", *BMJ Qual. Saf.*, vol. 13, no. suppl 1, pp. i65–i71, Oct. 2004, doi: 10.1136/qshc.2004.009951.

[80] G. Robert, J. Hockey, A. Healey, M. Crawshaw, D. G. Wastell, and J. Sauer, "Cognitive demands of collision avoidance in simulated ship control", *Hum. Factors*, vol. 45, no. 2, pp. 252–265, Jun. 2003, doi: 10.1518/hfes.45.2.252.27240.

[81] C. Hetherington, R. Flin, and K. Mearns, "Safety in shipping: The human element", *J. Safety Res.*, vol. 37, no. 4, pp. 401–411, Jan. 2006, doi: 10.1016/j.jsr.2006.04.007.

[82] K. I. Øvergård, L. J. Sorensen, S. Nazir, and T. J. Martinsen, "Critical incidents during dynamic positioning: operators' situation awareness and decision-making in maritime operations", *Theor. Issues Ergon. Sci.*, vol. 16, no. 4, pp. 366–387, Jul. 2015, doi: 10.1080/1463922X.2014.1001007.

[83] R. Szeligowski, "Cognifying the OODA Loop: Improved maritime decision making", *Gravely Naval Research Group, Naval War College Newport United States,* 2018. Accessed: Feb. 06, 2023. [Online]. Available: https://apps.dtic.mil/sti/citations/AD1057893

[84] M. R. Hafeznia, H. Pirdashti, and Z. Ahmadipour, "An expert-based decision making tool for enhancing the consensus on Caspian Sea legal regime", *J. Eurasian Stud.*, vol. 7, no. 2, pp. 181–194, Jul. 2016, doi: 10.1016/j.euras.2015.10.009.

[85] O. Arslan and I. D. Er, "A SWOT analysis for successful bridge team organization and safer marine operations", *Process Saf. Prog.*, vol. 27, no. 1, pp. 21–28, 2008, doi: 10.1002/prs.10209.

[86] C. E. Zsambok and G. Klein, *Naturalistic Decision Making*. Psychology Press, 2014.

[87] Y. Hanoch, "'Neither an angel nor an ant': Emotion as an aid to bounded rationality", *J. Econ. Psychol.*, vol. 23, no. 1, pp. 1–25, Feb. 2002, doi: 10.1016/S0167-4870(01)00065-4.

[88] C. Wilson and H. Dowlatabadi, "Models of decision making and residential energy use", *Annu. Rev. Environ. Resour.*, vol. 32, no. 1, pp. 169–203, 2007, doi:10.1146/annurev.energy.32.053006.141137.

[89] S. Nazir, S. Jungefeldt, and A. Sharma, "Maritime simulator training across Europe: A comparative study", *WMU J. Marit. Aff.*, vol. 18, no. 1, pp. 197–224, Mar. 2019, doi: 10.1007/s13437-018-0157-0.

[90] L. C. Main, A. Wolkow, and T. P. Chambers, "Quantifying the physiological stress response to simulated maritime pilotage tasks: the influence of task complexity and pilot experience", *J. Occup. Environ. Med.*, vol. 59, no. 11, p. 1078, Nov. 2017, doi: 10.1097/JOM.0000000000001161.

[91] G. L. Butler, G. J. M. Read, and P. M. Salmon, "Understanding the systemic influences on maritime pilot decision-making", *Appl. Ergon.*, vol. 104, p. 103827, Oct. 2022, doi: 10.1016/j.apergo.2022.103827.

[92] M. Grabowski, "Research on wearable, immersive augmented reality (WIAR) adoption in maritime navigation", *J. Navig.*, vol. 68, no. 3, pp. 453–464, May 2015, doi: 10.1017/S0373463314000873.

[93] A. Van Gastel, S. Scataglini, S. Zelck, L. Denteneer, H. V. Bossche, and S. Verwulgen, "Towards wearable technology assisting in training and risk assessment against musculoskeletal disorders for maritime workers", in *Advances in Simulation and Digital Human Modeling*, J. L. Wright, D. Barber, S. Scataglini, and S. L. Rajulu, Eds., in Lecture Notes in Networks and Systems. Cham: Springer International Publishing, pp. 368–376, 2021. doi: 10.1007/978-3-030-79763-8_44.

[94] D. R. Seshadri *et al.*, "Wearable sensors for monitoring the physiological and biochemical profile of the athlete", *Npj Digit. Med.*, vol. 2, no. 1, Art. no. 1, Jul. 2019, doi: 10.1038/s41746-019-0150-9.

[95] F. Canento, A. Fred, H. Silva, H. Gamboa, and A. Lourenço, "Multimodal biosignal sensor data handling for emotion recognition", in *2011 IEEE Sensors*, Oct. 2011, pp. 647–650. doi: 10.1109/ICSENS.2011.6127029.

[96] M. F. Rizwan, R. Farhad, F. Mashuk, F. Islam, and M. H. Imam, "Design of a biosignal based stress detection system using machine learning techniques", in *2019 Int. Conf. Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 364–368, Jan. 2019. doi: 10.1109/ICREST.2019.8644259.

[97] L. Ciabattoni, F. Ferracuti, S. Longhi, L. Pepa, L. Romeo, and F. Verdini, "Real-time mental stress detection based on smartwatch", in *2017 IEEE Int. Conf. Consumer Electronics (ICCE)*, pp. 110–111, 2017. doi: 10.1109/ICCE.2017.7889247.

[98] A. Carotenuto, I. Molino, A. M. Fasanaro, and F. Amenta, "Psychological stress in seafarers: a review", *Int. Marit. Health*, vol. 63, no. 4, Art. no. 4, 2012.

[99]K. Tam, R. Hopcraft, T. Crichton, and K. Jones, "The potential mental health effects of remote control in an autonomous maritime world", *J. Int. Marit. Saf. Environ. Aff. Shipp.*, vol. 5, no. 2, pp. 40–55, Apr. 2021, doi: 10.1080/25725084.2021.1922148.

[100] D. H. Hellhammer, S. Wüst, and B. M. Kudielka, "Salivary cortisol as a biomarker in stress research", *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 163–171, Feb. 2009, doi: 10.1016/j.psyneuen.2008.10.026.

[101] Ibadurrahman, K. Hamada, Y. Wada, J. Nanao, D. Watanabe, and T. Majima, "Long-term ship position prediction using automatic identification system (AIS) data and end-to-end deep learning", *Sensors*, vol. 21, no. 21, Art. no. 21, Jan. 2021, doi: 10.3390/s21217169.

[102] X. Chen, Y. Liu, K. Achuthan, and X. Zhang, "A ship movement classification based on automatic identification system (AIS) data using convolutional neural network", *Ocean Eng.*, vol. 218, p. 108182, Dec. 2020, doi: 10.1016/j.oceaneng.2020.108182.

[103] S. Mao, E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "An automatic identification system (AIS) database for maritime trajectory prediction and data mining", in *Proceedings of ELM-2016*, J. Cao, E. Cambria, A. Lendasse, Y. Miche, and C. M. Vong, Eds., in Proceedings in Adaptation, Learning and Optimization. Cham: Springer International Publishing, pp. 241–257, 2018. doi: 10.1007/978-3-319-57421-9_20.

[104] K.-I. Kim and K. M. Lee, "Deep learning-based caution area traffic prediction with automatic identification system sensor data", *Sensors*, vol. 18, no. 9, Art. no. 9, Sep. 2018, doi: 10.3390/s18093172.

[105] F. Wild, R. Klemke, P. Lefrere, M. Fominykh, and T. Kuula, "Technology acceptance of augmented reality and wearable technologies", in *Immersive Learning Research Network*, D. Beck, C. Allison, L. Morgado, J. Pirker, F. Khosmood, J. Richter, and C. Gütl, Eds., in Communications in Computer and Information Science. Cham: Springer International Publishing, pp. 129–141, 2017. doi: 10.1007/978-3-319-60633-0_11.

[106] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, in CHI ’88. New York, NY, USA: Association for Computing Machinery, pp. 213–218, 1988. doi: 10.1145/57167.57203.

[107] K. Norman, B. Shneiderman, and B. Harper, "QUIS: The questionnaire for user interaction satisfaction". Human-Computer interaction lab, University of Maryland [Online]. Accessed: May 03, 2023. Available: https://www.cs.umd.edu/hcil/quis/.

[108] T. Olsson, "Concepts and subjective measures for evaluating user experience of mobile augmented reality services", in *Human Factors in Augmented Reality Environments*, W. Huang, L. Alem, and M. A. Livingston, Eds., New York, NY: Springer, pp. 203–232, 2013. doi: 10.1007/978-1-4614-4205-9_9.

[109] S. Ssemugabi and R. de Villiers, "A comparative study of two usability evaluation methods using a web-based e-learning application", in *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, in SAICSIT ’07. New York, NY, USA: Association for Computing Machinery, pp. 132–142, 2007. doi: 10.1145/1292491.1292507.

[110] J. Annett, "Hierarchical task analysis", in *Handbook of Cognitive Task Design*, CRC Press, 2003.

[111] E. Salas, K. A. Wilson, C. S. Burke, and D. C. Wightman, "Does crew resource management training work? An update, an extension, and some critical needs", *Hum. Factors*, vol. 48, no. 2, pp. 392–412, Jun. 2006, doi: 10.1518/001872006777724444.

[112] M. R. Endsley, "Direct measurement of situation awareness: validity and use of SAGAT", in *Situational Awareness*, Routledge, 2011.

[113] D. Sharek, "A useable, online NASA-TLX tool", *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 55, no. 1, pp. 1375–1379, Sep. 2011, doi: 10.1177/1071181311551286.

[114] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research", in *Advances in Psychology*, P. A. Hancock and N. Meshkati, Eds., in Human Mental Workload, vol. 52. North-Holland, pp. 139–183, 1988. doi: 10.1016/S0166-4115(08)62386-9.

[115] E. Galy, J. Paxion, and C. Berthelon, "Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving", *Ergonomics*, vol. 61, no. 4, pp. 517–527, Apr. 2018, doi: 10.1080/00140139.2017.1369583.

[116] C. D. Spielberger, "State-trait anxiety inventory for adults", 1983, doi: 10.1037/t06496-000.

[117] K. N. Fountoulakis *et al.*, "Reliability and psychometric properties of the Greek translation of the state-trait anxiety inventory form Y: Preliminary data", *Ann. Gen. Psychiatry*, vol. 5, no. 1, p. 2, Jan. 2006, doi: 10.1186/1744-859X-5-2.

[118] A. Bryman, "Research design", in *Social Research Methods*, 5th ed.Oxford University Press, pp. 39–72, 2016.

[119] D. Zwillinger, "Chapter 7: Probability and statistics", in *Standard Mathematical Table and Formulae*, 31st ed., New York: CRC Press LLC, 2003.

[120] S. Glen, "Kruskal Wallis H test: definition, examples, assumptions, SPSS", *StatisticsHowTo.com: Elementary Statistics for the rest of us!* [Online]. Accessed: Jun. 15, 2023. Available: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/kruskal-wallis/

[121] M. Hollander, D. A. Wolfe, and E. Chicken, "Distribution-Free Test for General Alternatives (Kruskal–Wallis)", in *Nonparametric Statistical Methods*, 3rd ed.Wiley, p. 204, 2013 [Online]. Accessed: Jan. 14, 2023. Available: https://www.wiley.com/en-us/Nonparametric+Statistical+Methods%2C+3rd+Edition-p-9780470387375

[122] H. J. Keselman, A. R. Othman, R. R. Wilcox, and K. Fradette, "The new and improved two-sample t test", *Psychol. Sci.*, vol. 15, no. 1, pp. 47–51, Jan. 2004, doi: 10.1111/j.0963-7214.2004.01501008.x.

[123] T. Sakai, "Two sample t-tests for IR evaluation: Student or Welch?", in *Proc. 39th Int. ACM SIGIR Conf. Res. Dev. Inform. Retrieval*, New York, NY, USA: Association for Computing Machinery, pp. 1045–1048, 2016. doi: 10.1145/2911451.2914684.

[124] B. L. Welch, "On the comparison of several mean values: An alternative approach", *Biometrika*, vol. 38, no. 3/4, pp. 330–336, 1951, doi: 10.2307/2332579.

[125] A. Sherstyuk, D. Vincent, B. Berg, and A. Treskunov, "Mixed reality manikins for medical education", in *Handbook of Augmented Reality*, B. Furht, Ed., New York, NY: Springer, pp. 479–500, 2011. doi: 10.1007/978-1-4614-0064-6_23.

[126] P. A. Dickstein, J. K. Spelt, and A. N. Sinclair, "Application of a higher order crossing feature to non-destructive evaluation: a sample demonstration of sensitivity to the condition of adhesive joints", *Ultrasonics,* vol. 29, no. 5, pp. 355-365, 1991.

[127] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings", *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010, doi: 10.1109/TITB.2009.2034649.

[128] B. Kedem, "Higher-order crossings in time series model identification", *Technometrics*, vol. 29, no. 2, pp. 193–204, May 1987, doi: 10.1080/00401706.1987.10488210.

[129] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, MA: MIT Press, 2016.

[130] J. Gu *et al.*, "Recent advances in convolutional neural networks", *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[131] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. New York, NY: Wiley, 2001.

[132] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization", in *Adv. Neural Inform. Process Syst.,* 2011, vol.24 [Online]. Accessed: Jan. 15, 2023. Available: https://proceedings.neurips.cc/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

[133] S. K. Ong, Y. Shen, J. Zhang, and A. Y. C. Nee, "Augmented reality in assistive technology and rehabilitation engineering", in *Handbook of Augmented Reality*, B. Furht, Ed., New York, NY: Springer, 2011, pp. 603–630. doi: 10.1007/978-1-4614-0064-6_28.

[134] J. Nilsson, A. C. E. Ödblom, J. Fredriksson, and A. Zafar, "Using augmentation techniques for performance evaluation in automotive safety", in *Handbook of Augmented Reality*, B. Furht, Ed., New York, NY: Springer, 2011, pp. 631–649. doi: 10.1007/978-1-4614-0064-6_29.

[135] E. Ras, F. Wild, C. Stahl, and A. Baudet, "Bridging the skills gap of workers in industry 4.0 by human performance augmentation tools: Challenges and roadmap', in *Proc. 10th Int. Conf. PErvasive Technol. Related to Assistive Environ.*, New York, NY, USA, pp. 428–432, 2017. doi: 10.1145/3056540.3076192.

[136] C. Perey, F. Wild, K. Helin, M. Janak, P. Davies, and P. Ryan, "Advanced manufacturing with augmented reality", in *2014 IEEE Inter. Symp. Mixed and Augmented Reality (ISMAR)*, Munich, Germany, pp. 1–1, Sep. 2014. doi: 10.1109/ISMAR.2014.6948518.

[137] U. Yıldırım, E. Başar, and Ö. Uğurlu, "Assessment of collisions and grounding accidents with human factors analysis and classification system (HFACS) and statistical methods", *Saf. Sci.*, vol. 119, pp. 412–425, Nov. 2019, doi: 10.1016/j.ssci.2017.09.022.

[138] M. R. Grech, T. Horberry, and A. Smith, "Human error in maritime operations: Analyses of accident reports using the leximancer tool", *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 46, no. 19, pp. 1718–1721, Sep. 2002, doi: 10.1177/154193120204601906.

[139] L. Norros and K. Hukki, "Utilization of information technology in navigational decision-making", in *Co-operative Process Management: Cognition And Information Technology*, CRC Press, 2003, pp. 77–88.

[140] C. H. Allen, "Risk of collision", in *Farwell's Rules of the Nautical Road*, 8th ed., Annapolis, MD: Naval Institute Press, 2004, pp. 207–241.

[141] H. Sampson and M. Thomas, "The social isolation of seafarers: causes, effects, and remedies", *Int. Marit. Health*, vol. 54, no. 1–4, pp. 58–67, Jan. 2003.

[142] S.-G. Gug, J.-H. Yun, D. Harshapriya, and J.-J. Han, "A Prefatory study on the effects of alcohol on ship manoeuvring, navigational and decision-making abilities of navigators", *J. Navig.*, vol.75, no.5, pp. 1069–1081, Apr. 2022, doi: 10.1017/S0373463322000133.

[143] S. W. Hystad and J. Eid, "Sleep and fatigue among seafarers: The role of environmental stressors, duration at sea and psychological capital", *Saf. Health Work*, vol. 7, no. 4, pp. 363–371, Dec. 2016, doi: 10.1016/j.shaw.2016.05.006.

[144] A. Carotenuto *et al.*, "The Psychological General Well-Being Index (PGWBI) for assessing stress of seafarers on board merchant ships", *Int. Marit. Health*, vol. 64, no. 4, Art. no. 4, 2013, doi: 10.5603/IMH.2013.0007.

[145] H.-J. Jensen and M. Oldenburg, "Objective and subjective measures to assess stress among seafarers", *Int. Marit. Health*, vol. 72, no. 1, Art. no. 1, 2021, doi: 10.5603/IMH.2021.0007.

[146] K. V. Størkersen, A. Laiou, T. O. Nævestad, and G. Yannis, "Production and protection. Seafarers' handling of pressure in gemeinschaft and gesellshaft", in *Safety and Reliability – Safe Societies in a Changing World*, Boca Raton: CRC Press, 2018.

[147] K. Starcke and M. Brand, "Decision making under stress: A selective review", *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1228–1248, Apr. 2012, doi: 10.1016/j.neubiorev.2012.02.003.

[148] R. Hanzu-Pazara, E. Barsan, P. Arsenie, L. Chiotoroiu, and G. Raicu, "Reducing of maritime accidents caused by human factors using simulators in training process", *J. Marit. Res.*, vol. 5, no. 1, Art. no. 1, Apr. 2008.

[149] K. Hjelmervik, S. Nazir, and A. Myhrvold, "Simulator training for maritime complex tasks: An experimental study", *WMU J. Marit. Aff.*, vol. 17, no. 1, pp. 17–30, Mar. 2018, doi: 10.1007/s13437-017-0133-0.

[150] J. Ernstsen and S. Nazir, "Performance assessment in full-scale simulators – A case of maritime pilotage operations", *Saf. Sci.*, vol. 129, p. 104775, Sep. 2020, doi: 10.1016/j.ssci.2020.104775.

[151] M. R. Endsley, "A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM", *Hum. Factors*, vol. 63, no. 1, pp. 124–150, Feb. 2021, doi: 10.1177/0018720819875376.

[152] T. Nguyen, C. P. Lim, N. D. Nguyen, L. Gordon-Brown, and S. Nahavandi, "A review of situation awareness assessment approaches in aviation environments", *IEEE Syst. J.*, vol. 13, no. 3, pp. 3590–3603, Sep. 2019, doi: 10.1109/JSYST.2019.2918283.

[153] E. Demirel and D. Bayer, "A study on the assessment of sea training as an integral part of maritime education and training", *Online J. Qual. High. Educ.*, vol. 3, pp. 12–24, Jul. 2016.

[154] V. R. LeBlanc, "The effects of acute stress on performance: Implications for health professions education", *Acad. Med.*, vol. 84, no. 10, p. S25, Oct. 2009, doi: 10.1097/ACM.0b013e3181b37b8f.

Paper I

User satisfaction in augmented reality-based training using Microsoft HoloLens

H. Xue, P. Sharma, F. Wild

Computers, vol. 8, no. 1, p. 9

https://doi.org/10.3390/computers8010009

computers

MDPI

Article

User Satisfaction in Augmented Reality-Based Training Using Microsoft HoloLens

Hui Xue [1,*], Puneet Sharma [1] and Fridolin Wild [2]

1 Department of Technology and Security/Faculty of Science and Technology/Tromsø Campus, UiT The Arctic University of Norway, 9037 Tromsø, Norway; puneet.sharma@uit.no

2 Department of Computing and Communication Technologies/Faculty of Technology, Design and Environment/Headington Campus, Oxford Brookes University, Oxford OX33 1HX, UK; wild@brookes.ac.uk

* Correspondence: hui.xue@uit.no; Tel.: +47-776-46-625

Received: 4 December 2018; Accepted: 18 January 2019; Published: 25 January 2019

check for updates

Abstract: With the recent developments in augmented reality (AR) technologies comes an increased interest in the use of smart glasses for hands-on training. Whether this interest is turned into market success depends at the least on whether the interaction with smart AR glasses satisfies users, an aspect of AR use that so far has received little attention. With this contribution, we seek to change this. The objective of the article, therefore, is to investigate user satisfaction in AR applied to three cases of practical use. User satisfaction of AR can be broken down into satisfaction with the interaction and satisfaction with the delivery device. A total of 142 participants from three different industrial sectors contributed to this study, namely, aeronautics, medicine, and astronautics. In our analysis, we investigated the influence of different factors, such as age, gender, level of education, level of Internet knowledge, and the roles of the participants in the different sectors. Even though users were not familiar with the smart glasses, results show that general computer knowledge has a positive effect on user satisfaction. Further analysis using two-factor interactions showed that there is no significant interaction between the different factors and user satisfaction. The results of the study affirm that the questionnaires developed for user satisfaction of smart glasses and the AR application performed well, but leave room for improvement.

Keywords: augmented reality; Microsoft HoloLens; AR application; user experience; user satisfaction

---

1. Introduction

Augmented Reality (AR) means enhancing the user's perception "with additional, artificially generated sensory input to create a new experience including, but not restricted to, enhancing human vision by combining natural with digital offers" [1]. Augmented Reality typically has three characteristics [2]: (1) AR combines the virtual with the real world; (2) objects are registered from both the real and virtual world in one coordinate system; and (3) the interaction between the objects of both worlds is possible in real time.

Hands-on training is important for many disciplines and professions, such as medical workers, mechanics, technicians, electricians, engineers, sailors, pilots, and firefighters. In the past decade, AR has been increasingly employed for a number of training applications, such as medical education [3], rehabilitation engineering [4], automotive safety [5], task assistance [6], and manufacturing [7].

For the successful adoption of AR-based training across different domains, one of the key factors is user satisfaction. User satisfaction is defined as a combination of different factors associated with the usage of the AR application and the associated delivery device [8]. These factors include: a feeling of powerfulness and achievement; an efficient use of time, effort, and other resources; meaningful content; a better insight into the training environment; a natural interaction; a feeling of amazement;

performance that exceeds expectations; playfulness; the invoking of positive feelings and pleasing memories; immersion and engagement; a transparent interaction; the feeling of participation in a community; a sense of privacy of the user's content; inspiration, encouragement, and motivation; and, finally, artistic creativity [8].

The main objective of this study was to test and observe user satisfaction in using AR applications and using AR glasses. The method for evaluating included questionnaires and interviews. The AR app used in this evaluation, therefore, has two parts: one is the expert recording the experience in the workplace, and the other is the novices training on work-related procedures using said recordings. In this study, we evaluated the following research hypotheses: to find if experts and students are satisfied with the prototype application, to see if the application can increase interest in learning new skills, and to evaluate if the users find the application easy to use.

The rest of this paper is organized as follows. First, we turn to the state of the art, summarizing what the research has found so far with respect to AR user interaction, AR user satisfaction, and questionnaires used for evaluating user satisfaction. Next, the AR app used in the trials is described. Subsequently presented are the research methodology and a summary of the information of the participants, devices, design of trial tasks, and evaluation methods. Finally, findings and results are illustrated, and the discussion and conclusion are given at the end.

2. State of the Art

2.1. AR User Interaction

AR technologies provide a different user experience than that of, for example, mobile phone apps. The user interacts with the surrounding real world, combining inputs from the environment with digital augmentations. Popular examples include PokemonGO and SnapChat. These types of apps certainly brought the term "augmented reality" into the spotlight [9]. With the advent of consumer-grade AR glasses, different types of AR user interactions are becoming necessary. For example, a user who is wearing Microsoft HoloLens can communicate diagrams and other types of graphics directly embedded into the environment to a different, remote user (see Figure 1).

Figure 1. With Microsoft HoloLens, a user connects the wires with remote assist (Used with permission from Microsoft Corporation) [10].

2.2. AR User Satisfaction and Questionnaires for Evaluating User Satisfaction

AR technology has evolved from offline to online, from static devices to mobile devices, and from desktop and mobile to wearable devices [11]. Consequently, with AR development over the past decade or so, special attention has been drawn to the maximization of AR user satisfaction. AR user satisfaction is dependent on both the design of the user interface (UI) and the choice of the AR hardware. Personalization of AR glasses can lead to greater AR user satisfaction [12]. AR apps designed for a good user experience result in a more overall satisfied AR user. This applies to AR navigation apps, AR health apps, AR education apps, and certain AR smart glasses games [13].

There are several concepts and subjective measures for evaluating the user experience of AR services. With regards to the user, satisfaction questionnaires are common tools used to evaluate a user's experience. One such tool—the Questionnaire for User Interaction Satisfaction (QUIS)—is designed to assess users' subjective satisfaction with specific aspects of the human–computer interface [14]. The results of QUIS facilitate new developments by addressing reliability and validity problems found using its satisfaction measurements. Therefore, the measure is highly reliable across many types of interfaces.

QUIS consists of a demographic questionnaire, a six-scale measure of overall system satisfaction, and hierarchically organized measures. The measures include the following specific interface factors [14]: screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation. Each area is measured by a seven-point scale according to the user's overall satisfaction with the interface and the above factors [14].

3. The AR Application

The AR application consists of two modes:*recorder* and *player*. This AR application is part of the work from WEKIT (Wearable Experience for Knowledge Intensive Training) project.

The *recorder* is designed for capturing an expert's workplace experience and combining it with technical documentation associated with a given scenario. The *player* is used to reenact the scenario to verify the recordings and usually employed to train a novice for the scenario.

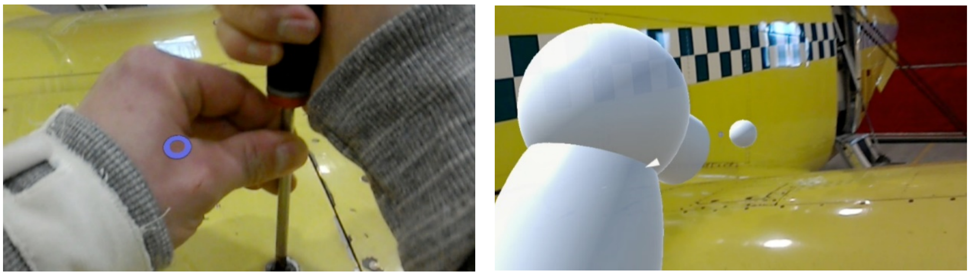
To capture an expert's experience, a set of transfer mechanisms were defined by Limbu et al. [15]. The so-called transfer mechanisms allow us to map the key aspects of an expert's performance to low level data and subsequent sensors. For more details on the different sensor components and their integration, please see the work by Sharma et al. [16]. The recorder (as shown in Figure 2 [17]) consists of a radial menu that allows us to select different options for capturing diverse annotations such as: pictures, videos, text annotations (for adding text information to different objects in the environment), audio, ghost hands (to capture the locations and movements of user's hands) and 3D models (useful for performing the task).

Trainers can use a so-called "ghost track" to record their own position and indoor movement, while at the same time recording voice explanations. When replaying such recording to the trainees, the holographic "ghost" representation of the expert provides more intuitive guidance on where to be, where to focus, and what to do than merely reading about the task to be learned in a manual using text and illustration. Figure 3 shows an example of such ghost track recording and replay for an aircraft maintenance task. The app was recording the expert when he was maintaining the aircraft (Figure 3a [18]). After recording, in the replay, as shown in Figure 3b [18], we can see a representation of the expert's position and his or her hand position (represented by the white sphere).

The *player* is the mode designed for trainees to learn how to do procedural operations (kind of "do-torial" mode). The app executes AR learning experience models (IEEE standard association, working group p1589), thus allows loading different learning and training activities. Activities can be transferred from device to device as well as from place to place, using a calibration marker to recalculate the relative positions of all points of interest, while utilizing 3D environmental mapping to provide stable projections.

The WEKIT player starting screen is shown in Figure 4 [17]. Once the task starts, the first action step and its associated augmentations are shown on the smart glasses display. From the perspective of the users, this typically means that the visual annotations overlay onto their unimpeded real-world view (optical see-through). Step by step, they guide the user through the learning task at hand. Gesture commands, voice commands, and the hardware clicker are all available when using the app. Figure 5 [19] shows an example of the WEKIT player in action. When the sensors on the HoloLens detect the particular tangible object, the virtual button is displayed in front of the trainee, while instruction on handling and movement are given at the same time.

Figure 2. User interface of the recording mode. Image from the WEKIT consortium in 2017 [17].



(**a**) (**b**)

Figure 3. A ghost track in WEKIT Recorder mode: (**a**) recording a ghost track; and (**b**) ghost track replay. Image from [18].

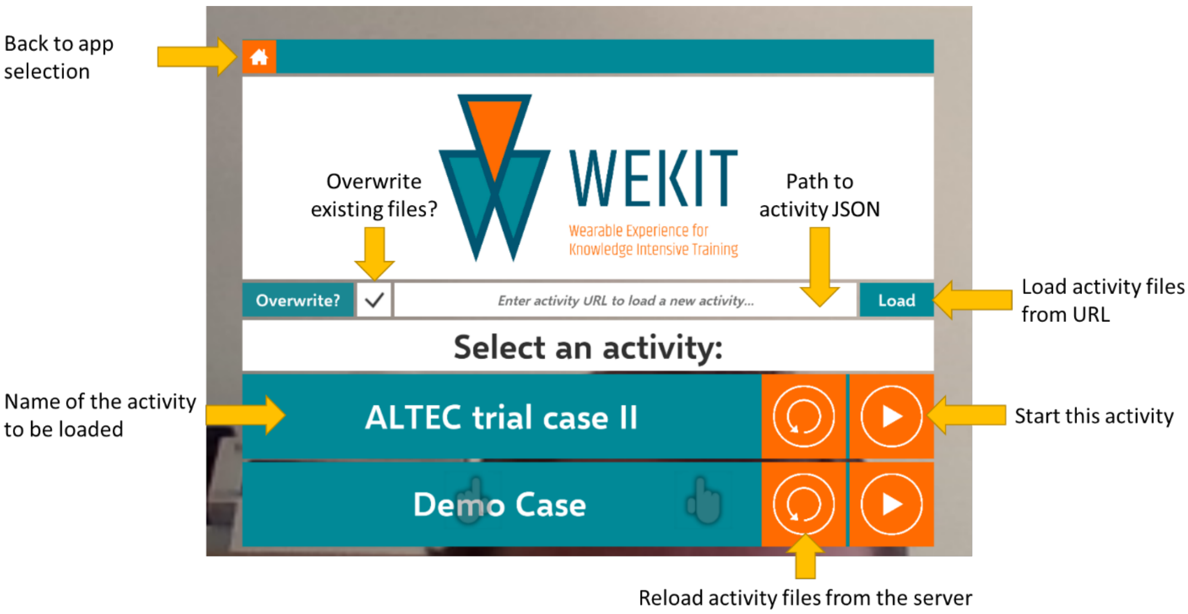


Figure 4. Starting screen in WEKIT (Wearable Experience for Knowledge Intensive Training) player mode. Image from Jaakko Karjalainen, WEKIT consortium in 2017 [17].

Figure 5. Example of user interface of WEKIT Player mode. Image from [19].

4. Research Design/Experiment Methodology

4.1. Participants

To evaluate the satisfaction of the user interaction and the smart glasses user experience, the WEKIT application was deployed in three different pilot testing scenarios: aviation, medical imaging, and space. In total, 142 people participated in the three scenarios, 55 in aviation, 48 in medical imaging, and 39 in space. Moreover, in the experiments, the test population was divided into two main groups, experts and students, respectively. A total of 47 experts (8 females and 39 males) with a high level of technological competency in their respective fields were recruited. A total of 95 learners (23 females and 72 males) from the three different fields voluntarily participated in the trials. The majority of the participants (68) were in the 18–24 age group, followed by 48 of the participants in the range between 25 and 34. Most of the participants had moderate or better computer knowledge and Internet knowledge, expressed on a five-point Likert scale ranging from very poor, poor, moderate, good, to very good. All participants gave written consent for their participation in the trials.

4.2. Material and Apparatus

The trial used the Microsoft HoloLens as wearable AR glasses for assessing the user's satisfaction with AR training. There are two parts in the WEKIT technology platform [20] deployed on HoloLens. One is a recorder for capturing expert experience and the other one is a player for presenting the expert's experience to the trainees. During the trial, all interactions and manipulations were done by using gesture and voice command only.

4.3. Trial Design/Task

The trial tasks were separated into three different areas, as mentioned in Section 4.1. Tasks in the Aeronautics use case were performed at Lufttransport, Norway. The scenario used for the aeronautics use case was a pre-flight inspection consisting of checking and securing different items such as baggage, exits, locks, and checking the status of components such as landing gears, brakes, engine switches, battery, and fuel. The experts comprised of maintenance apprentices, skilled workers (mechanics), and technicians working on base maintenance at Lufttransport. The novice group comprised of student volunteers from UiT The Arctic University of Norway [18]. Figure 6 shows a novice engaging in the pre-flight inspection task. Experts had been using the different types of annotations to create the required instruction for the training procedure, which then was provided to the trainee in the player mode of the AR app. The novice followed the instructions in order to complete the task in the cockpit.

The pre-flight inspection scenario consisted of the steps shown in Table 1.

Table 1. Steps of the pre flight inspection scenario for Beechcraft B200 [18].

No.	Cabin/Cockpit	Action	Content
1.	Baggage	Secure	Ensure that the baggage compartment and net is secured.
2.	Emergency Exit	Secure and unlocked	Emergency exit handle must be in the secured position and the lock must be in the unlocked position.
3.	Control locks	Remove and stowed	The control locks must be removed and stowed.
4.	Trim Tabs Exit	Set to "0"	Including elevator trim tab, aileron trim tab, elevator trim tab.
5.	Condition levers	Fuel cut-off	Must be set to the fuel cut-off position.
6.	Landing gear control	Down	Must be in down position.
7.	Parking brake	Set	If required, ensure that the parking brake is set on.
8.	Ignition and engine start switches	Ensure off	Must be in the off position.
9.	Battery	Check for minimum 23 V	Turn on the battery master switch. Check for minimum 23V on the voltmeters by pushing the push-to-test knobs on the voltmeters.
10.	Fuel quantity	Check	Check the fuel quantity in main fuel tanks. Move and hold the "fuel quantity"-switch to auxiliary position and check the fuel quantity in auxiliary fuel tanks.

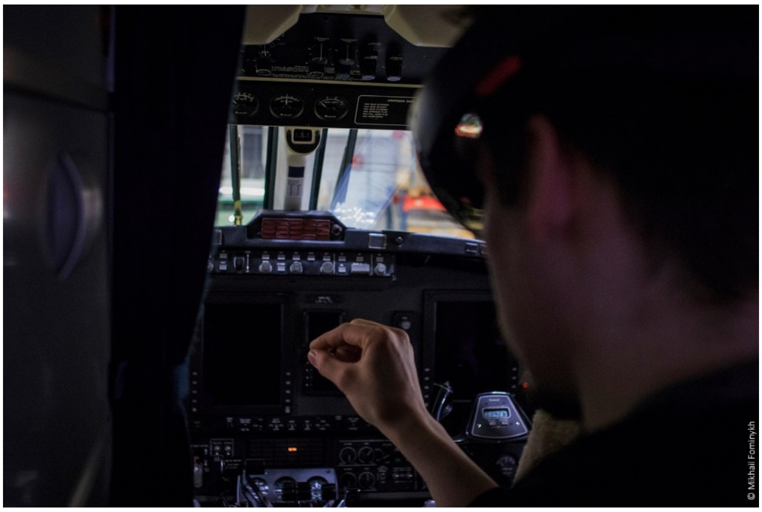


Figure 6. Maintenance Engineer in the cockpit of a Beechcraft B200 King Air model. Image: Mikhail Fominykh, WEKIT consortium in 2017 [21].

The medical task involved imaging and diagnostic workers and was conducted at EBIT (Esaote's Healthcare IT Company) in Genoa, Italy [22]. This task was for training medical students and radiologist apprentices on using MyLab8, an ultrasound machine produced by ESAOTE [23]. Similar to the trial at Lufttransport, the users executed the steps of the procedure using the player mode of the application. The scenario for the medical use case was to perform a particular ultrasound examination to analyze a patient's condition. The patient was a paid actor. During the task, the novice doctors needed to combine data from different sources in order to arrive at the correct diagnosis. As for the holographic training instruction, the guidance was set up for the player mode again using experts, adding the step-by-step description needed to guide the trainee through the full scanning and assessment procedure. The novice doctors then wore the HoloLenses and tried to perform the examination. The tasks are shown in Table 2. In Figure 7, we can see a novice performing a task by positioning the probe in the target direction and taking measurements using the player application.

Table 2. Steps of diagnostic training of radiology students performing an ultrasound examination [23].

No.	Ultrasound Equipment	Action	Content
1.	Probe	Choose	Choose the proper probe. Point to the linear probe and listen to the audio annotation explaining which probe to select and why, and how to hold it (with a raised edge).
2.	Button	Select the mode	Point to the "B/M" button to select the correct mode.
3.	Probe	Transversal position	Position the probe in a transverse direction.
4.	Probe	Longitudinal direction	Position the probe in a longitudinal direction.
5.	Button	Change the mode	Change the mode to Color Mode.
6.	Button	Choose button	Position the center line in the middle of the artery.
7.	Button	Change the mode	Change the mode to Doppler Mode. If required, ensure that the parking brake is set on.
8.	Circle button	Pointing	Point to Circle button highlighted in the following figure.
9.	Measure button	Measure	Choose correct button to start reading.
10.	Trackball button	Measure	Position the cursor over the highest peak in the curve, then click the left trackball button to set the first data point. Repeat for the lowest point in the graph.
11.	Image button	Snapshot	Take a snapshot with the measure.

Figure 7. A radiologist conducting ultrasound training. Image: WEKIT consortium in 2017 [21].

The space task that was conducted at the facilities of ALTEC in Turin, Italy and it involved training astronauts on how to install a Temporary Stowage Rack (TSR). The TSR installation is a procedure that astronauts have to perform on the International Space Station (ISS) [24]. Similar to the trials at the other two organizations, experts designed the training scenario, while a larger number of trainees then executed the scenario on the player application. The evaluation of the expert's experience was conducted using the recorder mode of the app as well as the player, while the trainees used only the player mode. The steps for this procedure are as follows. First, trainees were asked to scan the working area to create the 3D model of the environment, and then to identify the six seat track studs location on the structure, the position of the ball bearing and brackets. Next, they were instructed to fix the six studs in specific locations. Finally, they were asked to extract TSR, deploy it, and fix it to the correct places. The novices performed the task based on the recorded content. Table 3 shows the details of the steps. In Figure 8, we can see a participant of the trials performing a task in a replica module of the International Space Station.

Table 3. Steps of the installation of the Temporary Stowage Rack (TSR) in the Automated Transfer Vehicle (ATV) Part Task Trainer [24].

No.	Training Facility	Action	Content
1.	Rack	Identify	Identify the TSR interfaces.
2.	Rack	Localize and fix	Install the 6 studs in the correct position.
3.	CTB	Localize	Extract the TSR from Cargo Transfer Bag (CTB) and deploy it.
4.	Label	Orient	Orient the TSR with the label on top.
5.	AR glasses	Note	Show note regarding the step: "Start with rear side attachments. Do not tighten rear straps or front straps will be short.".
6.	Rack	Connect	Connect the rack straps to the corresponding Automated Transfer Vehicle (ATV) interfaces.
7.	Experts/Trainees	Change position	Move to the front side of the TSR.
8.	Rack	Fix	Fix the TSR straps to the studs.
9.	Rack	Fix	Fix the TSR straps to the brackets.
10.	Straps	Check	Check the tightness of the straps and tighten as needed.

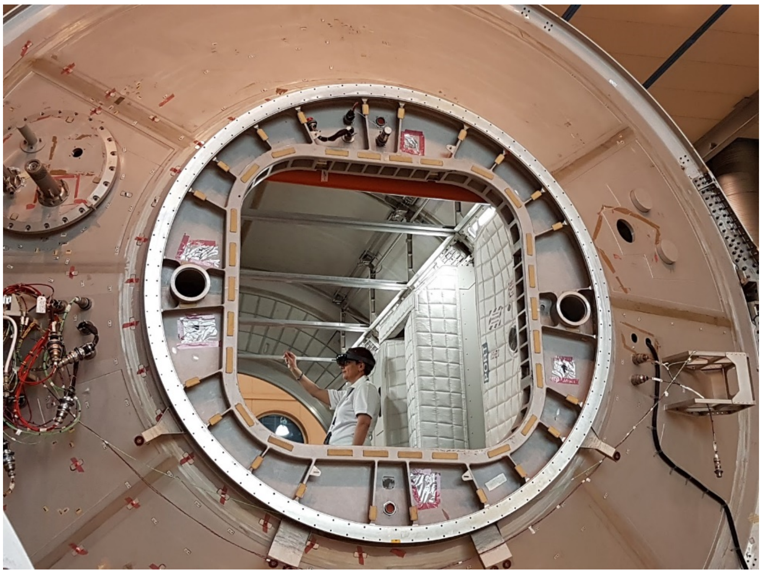


Figure 8. Astronaut trainer in a replica training module of the international space station. Image: WEKIT consortium in 2017 [21].

4.4. Smart Glasses User Satisfaction (SGUS)

The Smart Glasses User Satisfaction (SGUS) questionnaire was created for the WEKIT trials. It is a tool designed to assess users' subjective satisfaction with smart glasses. SGUS is a method and measure to scrutinize aspects, such as an enhanced perception of the environment, interaction with the augmented environment, implications of location and object awareness, the user-created AR content, and the new AR features that users typically use [8]. The general objective of the questionnaire is to understand the potential end users' central expectations of AR services with smart glasses, especially from an experiential point of view [8]. In this study, the smart glasses used for the different use cases were Microsoft HoloLens. SGUS measures subjective satisfaction on the basis of different features associated with user satisfaction, such as the content and interaction with the content. SGUS is based on evaluation criteria for web-based learning [25] and statements evaluating the user experience of mobile augmented reality services [8]. Some of the items from the table "Evaluation criteria for web-based learning—a framework" [25] and table "Examples of formative subjective statements with regard to the value and overall goodness of the service in terms of the UX category in question" [8] were picked and modified. SGUS consists of 11 items (statements) on a seven-point Likert scale (1–7) [18]. The 11 statements include three categories of evaluation criteria, which are general interface usability criteria, AR interaction-specific criteria for an educational AR app, and learner-centered effective learning [25].

4.5. Questionnaire for User Interface Satisfaction (QUIS)

The Questionnaire for User Interaction Satisfaction (QUIS) measures subjective satisfaction with specific aspects of the interface and interaction between the user and the AR application [26]. In this study, QUIS was modified for AR glasses, i.e., HoloLens. There are five sections in the "User Evaluation of an Interactive Computer System" (see Appendix A) [26]. We picked some items from this questionnaires for our study: all the questions in the *overall reactions to the software* section; No. 44 and 46 in *screen section*; No. 50 and 51 in *terminology and system information*; No. 54, 55, and 57 in *learning* section; and No. 60 and 61 in *system capabilities* section. All of these are directly used, except No. 54, which was modified to AR glasses to adapt this study. The rest of the items were not applicable for our setting, therefore were not used. Hence, a questionnaire with 15 items was used. To maintain consistency with the survey in other sections, each item was mapped to a numeric value of 1–7 instead of the nine-point scale.

4.6. Procedure

As most participants had no experience with AR glasses, at the beginning of the trial, they were asked to familiarize themselves with the AR glasses, i.e., HoloLens. To do this, gesture training with HoloLens was done before they started using the application. The application comprised a scenario that the participants had to complete in a particular use case setting. The content of the application was generated by experts in that specific use. After the participants completed all the tasks, they were provided with the QUIS and SGUS questionnaires to complete.

5. Results/Findings

5.1. Descriptive Statistics

In this section, we report on descriptive statistics for the smart glasses user interaction and the interaction satisfaction. We organize the findings alongside the investigation of eight hypotheses, with the summary of these shown in Table 4.

Hypothesis 1. *Does gender matter? In Science and Engineering, gender is not balanced and there are fewer women than men [27]. Gender stereotypes can affect use of established technologies. We therefore investigated whether the influence on user satisfaction of these new media will be moderated by gender.*

Hypothesis 2. *Does age matter? Studies imply that younger people embrace new technologies more easily [28]. Since we used AR glasses and applications for training, we wanted to know whether age affects user satisfaction.*

Hypothesis 3. *Are experts more tech savvy? It is likely that experts have more experience with technology applications, as in general they also have more domain-specific knowledge and skills. We assumed that they* *would be more able to grasp the app concept, thus be more satisfied with the interaction. The novices, however, may have less knowledge and skills, hence, may find the app difficult to use.*

Hypothesis 4. *Does education matter? Higher levels of education go hand in hand with higher levels of ICT skills. It is justified to hypothesize that the educational level predicts satisfaction.*

Hypothesis 5. *Does computer knowledge matter? Higher levels of ICT and media skills typically involve transfer skills. The AR smart glasses headset used, Microsoft Hololens, is a stand-alone device. We need basic computer knowledge to use it. Those with better computer knowledge might find it easy to use, and hence, give a higher score in terms of user satisfaction.*

Hypothesis 6. *Does Internet knowledge matter? In analogy to computer skills, one can expect Internet skills to influence the user satisfaction levels in a positive manner.*

Hypothesis 7. *Are there differences in satisfaction levels between the participants of the three test-beds? The trials involved three different learning tasks, in three different environments, with three different groups of participants. As all three trials are about training a particular procedure, there are no differences identified across test-beds.*

Hypothesis 8. *Is there any interaction between the above-mentioned factors?*

Table 4. Summary of the hypotheses.

#	Description	Expectation
H1	Gender	Men are more satisfied with the user interaction than women.
H2	Age	Younger participants give a higher score.
H3	Experts vs. novices	Experts have higher satisfaction levels.
H4	Education level	Higher education users have higher satisfaction levels.
H5	Computer knowledge level	Users with better computer knowledge might be more satisfied.
H6	Internet knowledge level	Might have influence.
H7	Three different test-beds	Might have different results.
H8	Above seven factors	There might be interactions between factors.

5.1.1. SGUS

As mentioned before, SGUS has 11 items. The summation of the score for the 11 items is the SGUS score. As shown in Table 5, we provide data the following data: n (number of participants), mean, standard deviation, minimum value, Q1 (the first quartile: "middle" value in the first half of the rank-ordered data set), median, Q3 (the third quartile: "middle" value in the second half of the rank-ordered data set), and maximum value for the variables gender, education level, roles, and organizations. Based on these results, it is clear that the mean scores are similar across the different levels associated with the variables.

Table 5. Descriptive statistics of the Questionnaire for Smart Glasses User Satisfaction (SGUS).

Variable	Level	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Gender	Female	31	58.74	7.96	43	54.5	58	64.5	72
	Male	111	58.49	8.45	20	54	60	64.5	72
Role	Experts	47	56.98	8.83	33	49.5	58	64	72
	Students	95	59.32	7.99	20	55	60	65	74
Education level	Upper secondary school or lower	45	57.98	7.92	33	54	57	64	70
	Bachelor's or higher	97	58.8	8.52	20	55	60	65	74
Organization	Space(1)	39	59.54	9.46	20	57	61	65.5	71
	Medicine(2)	48	58.69	7.43	38	54.75	59	64	72
	Engineering(3)	55	57.71	8.26	33	52	57	64	74

5.1.2. QUIS

Similarly, the overall Questionnaire for User Interface Satisfaction (QUIS) score was calculated by summation of the score for the 15 QUIS items. Summary data for all questions in QUIS are presented in Table 6. The 15 items were designed independently from each other. These items aim to investigate the satisfaction of users with different aspects of the interface, including usability and user experience in using AR applications.

Table 6. Descriptive statistic of the Questionnaire for Smart Glasses User Satisfaction (QUIS).

Variable	Level	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Gender	Female	30	75.94	11.44	55	69	76.5	82.75	98
	Male	103	76.99	13.23	18	71	78	86	103
Role	Experts	43	76.28	12.06	49	69	77	85.5	97
	Students	90	77.01	13.21	18	72	78	85	103
Education level	Upper secondary school or lower	43	75.14	13.82	18	69.5	75	85	95
	Bachelor's or higher	90	77.56	12.30	33	71	78	85.75	103
Organization	Space(1)	39	76.67	12.44	33	72	77	86	96
	Medicine(2)	42	80.50	9.71	55	75	80	85.75	97
	Engineering(3)	52	73.85	14.61	18	66	74.5	83.5	103

5.2. Correlation

In this section, we discuss correlation for SGUS and correlation for QUIS.

5.2.1. Correlation of SGUS

Spearman's correlation coefficient, ρ, measures the strength and direction of association between two ranked variables in the range $[-1, 1]$. Based on the 11 items, the results of Spearman's rank correlation are shown in Table 7: the first value of each row represents Spearman's correlation coefficient, and the second value of each row represents the p value. It can be seen that almost all items are statistically significant ($p < 0.05$) and have a low positive correlation. This implies that all the items are independent.

In the study of SGUS, each of the items investigates a different aspect of the user experience. For the analysis, the overall averages for all items were calculated. Figure 9 shows the plot of the average score from individual items. The box in the plot depicts the answer of 50% of the participants, with the line in the middle indicating the median. The dotted lines span the 95% confidence interval. Outliers are depicted with black dots. The connected red dots indicate the medians. The results imply that most of the participants had a good conception of what is real and what is augmented when using AR-glasses (GL5). The participants indicated that the system and content helped them to accomplish the task quite well (GL7) and their attention was captivated in a positive way (GL6). The provided content was also seen as contextually meaningful (GL2). However, performing the task with AR glasses was experienced as less natural (GL9 and GL4), and following and understanding the task phases (GL8 and GL10–11) was not very easy [18]. The results were very much in line across the three.

Table 7. Spearman's rank coefficient of correlation for SGUS: the first value of each row represents Spearman's correlation coefficient, and the second value of each row represents the p value.

	GL1	GL2	GL3	GL4	GL5	GL6	GL7	GL8	GL9	GL10	GL11
GL1	1	0.316 0.000 **	0.209 0.013	0.269 0.001	0.164 0.053	0.301 0.000 **	0.270 0.001	0.323 0.000 **	0.285 0.001	0.376 0.000 **	0.336 0.000 **
GL2		1	0.335 0.000 **	0.371 0.000 **	0.239 0.005	0.308 0.000	0.345 0.000 **	0.227 0.007	0.287 0.001	0.354 0.000 **	0.398 0.000 **
GL3			1	0.487 0.000 **	0.172 0.041	0.270 0.001	0.444 0.000 **	0.312 0.000 **	0.320 0.000 **	0.265 0.002	0.289 0.001
GL4				1	0.121 0.154	0.293 0.000 **	0.376 0.000 **	0.337 0.000 **	0.492 0.000 **	0.226 0.008	0.243 0.004
GL5					1	0.260 0.002	0.178 0.036	0.170 0.046	0.026 0.763	0.062 0.468	0.166 0.052
GL6						1	0.416 0.000 **	0.387 0.000 **	0.453 0.000 **	0.317 0.000	0.403 0.000 **
GL7							1	0.490 0.000 **	0.500 0.000 **	0.453 0.000 **	0.435 0.000 **
GL8								1	0.563 0.000 **	0.442 0.000 **	0.390 0.000 **
GL9									1	0.455 0.000 **	0.364 0.000 **
GL10										1	0.558 0.000 **
GL11											1

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

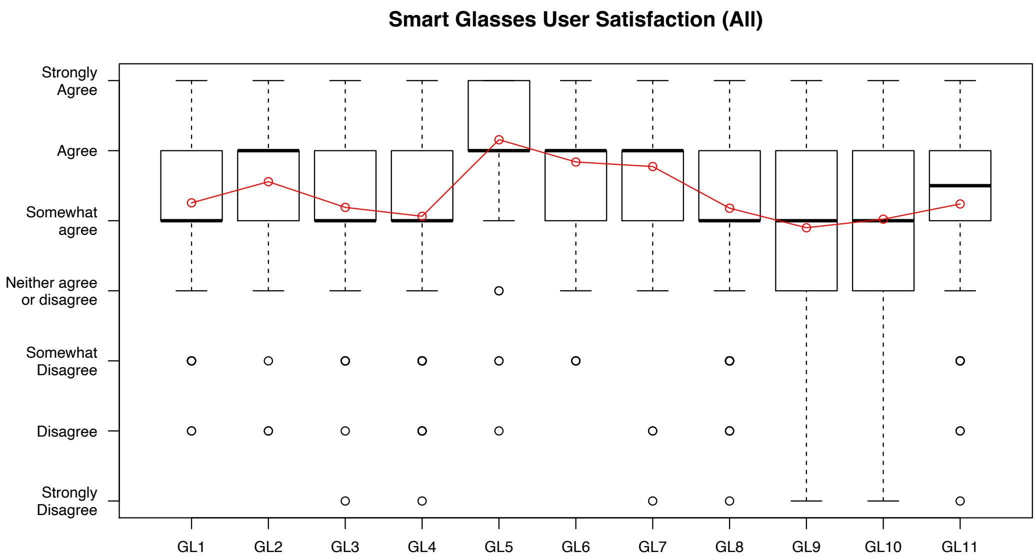


Figure 9. Plot of SGUS score for each item.

5.2.2. Correlation of QUIS

The correlation for QUIS is based on 15 items. The results of Spearman's rank correlation are shown in Table A1 (see Appendix A). The values in the table have the same meaning as in Table 7. The results are similar to those of SGUS; most of the items are statistically significant ($p < 0.05$) and have a low positive correlation. This implies that most of the items are independent.

In the study of QUIS, each of the items investigated different aspects of the user experience. For the analysis, the overall average from all items was calculated. Figure 10 shows the plot of the average score from individual items, and the description of the plot is the same as that of the SGUS plot. The results imply that most of the participants agree that learning to operate the AR glasses (QS13) seemed to be rather easy, and the overall enthusiasm towards the system seemed (QS1 and QS5) to be very positive. The characters on the screen were relatively easy to read (QS9). The means of QS3, QS4, QS6, QS7, and QS8 indicate that the system was experienced as rigid, unreliable, and slow, which may cause frustration [18].

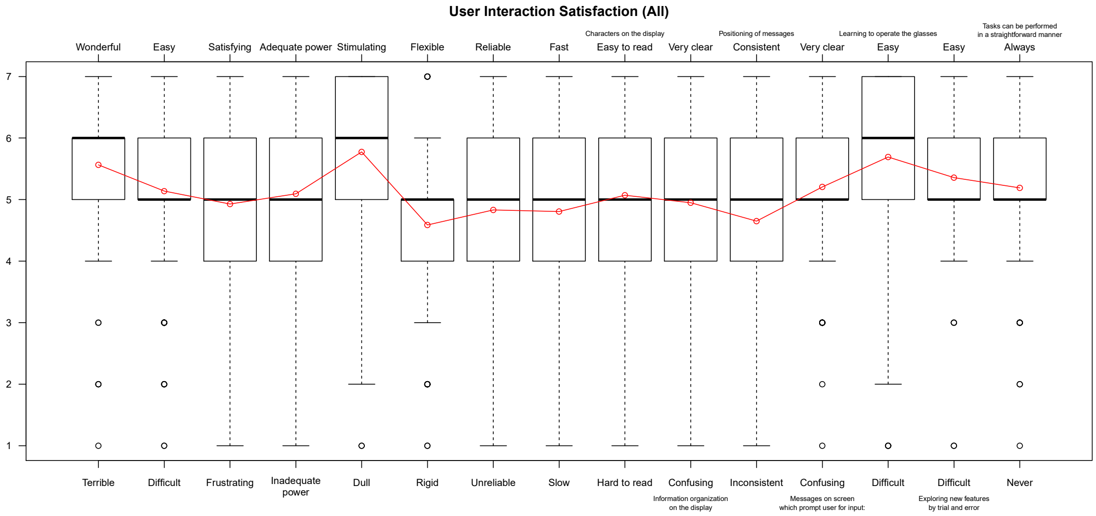


Figure 10. Plot of QUIS score for each item.

5.3. *Analysis of Variance and Interaction Plots*

The participants were described by seven factors: gender, age, role, education skill level, computer knowledge level, Internet knowledge level, and organization. Each factor was divided by two levels, except for organizations, which were in three levels. Please note that none of the participants claimed that they have a poor or very poor Internet knowledge level. The following section discusses the analysis of variance (ANOVA) of QUIS and of SGUS. In this ANOVA study, SGUS and QUIS scores were investigated for using the application on the AR glasses with six independent variables, i.e., the relationships between: age distribution, gender, roles, highest level of education, organization, and computer knowledge. Therefore, there were six main effects and 57 interactions. We were interested in whether there is a relationship between the satisfaction levels (measured by the questionnaire) and these factors.

5.3.1. ANOVA of SGUS

In this study, we investigated whether the age, gender, roles, computer knowledge level, or different organizations have an effect on the satisfaction of using AR glasses. To determine this, we needed to look at the simple main effects: the main effect of one independent variable (e.g., age) at each level of another independent variable (e.g., for students and for experts).

Figure 11 shows the main effects of the six factors. Participants with different computer knowledge levels have the greatest differences in the SGUS results. This means that the participants with good computer knowledge and poor computer knowledge gave different scores for user satisfaction. The results show that participants with good or very good computer knowledge were, in general, more satisfied with the smart glasses application, and there is a significant effect from computer knowledge levels (F value = 8.87, $p = 0.003$). The result implies that the SGUS score was affected by the effects of good computer knowledge.

Table 8 shows the summary results of the linear model of the independent variables. The estimate for the model intercept is 54.688 and the coefficient measuring the slope of the relationship with computer knowledge level is 4.324. There is strong evidence that the significance of the model coefficient is significantly different from zero: as the computer skill level increases, so does the satisfaction. The information about the standard errors of these estimates is also provided in the Coefficients table. In the result of the multiple regression model, only 8.8% of the variance in the SGUS scores is explained by each of the factors (Multiple R-squared is 0.088). There is no statistically significant factor that explains the variation in the SGUS scores (overall p value is 0.08).

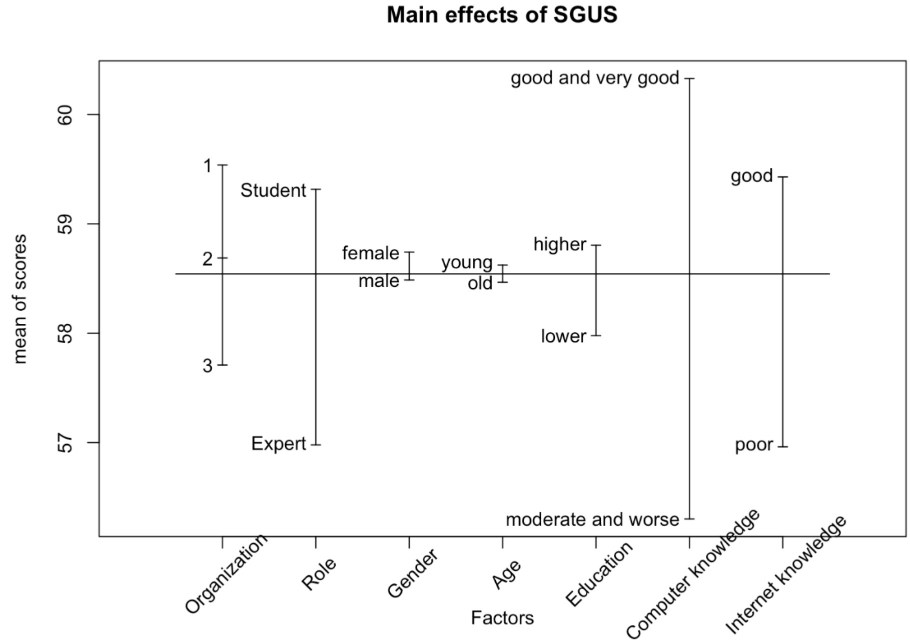


Figure 11. Main effects of SGUS.

Table 8. Results of the linear model of the independent variables.

Source of Variation	Estimate	Std. Error	t Value	Pr (>|t|)
(Intercept)	54.688	2.652	20.622	$<2 \times 10^{-16}$ ***
Medicine	−1.604	1.835	−0.874	0.384
Engineering	−0.996	1.906	−0.523	0.602
Role	2.862	1.624	1.762	0.080
Gender	1.250	1.756	0.712	0.478
Age	0.563	1.634	0.344	0.731
Education level	−0.147	1.716	−0.086	0.932
Computer skill	4.324	1.452	2.978	0.003

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

To investigate the interaction, it was interesting to find whether the SGUS score depends on an interaction between good computer knowledge and the other factors. The two-factor interaction plot is shown in Figure 12. The following are the findings from the plot:

- Female participants with good computer knowledge have a higher SGUS score than males with good computer knowledge; both females and males with moderate and worse computer knowledge have nearly the same, lower SGUS score (Figure 12a).

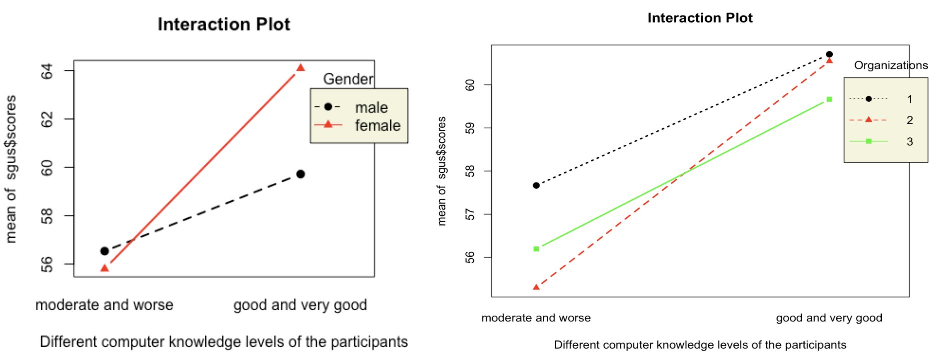
- Participants from medicine with good computer knowledge tended toward a higher SGUS score than participants from engineering, and there is no significant difference between them and the participants with good computer knowledge from astronautics and medicine (Figure 12b).
- There is no significant interaction between participants with different computer knowledge levels from astronautics and engineering (Figure 12b).
- There is no significant interaction between students and experts with different computer knowledge levels (Figure 12c).
- Participants younger than 25 years old with good computer knowledge tended toward a higher SGUS than participants older than 25 years old; however, participants younger than 25 years old with a moderate and worse computer knowledge level tended toward a lower SGUS score (Figure 12d).
- Participants with secondary school or lower education level and good computer knowledge tended toward a higher SGUS score than participants with a bachelor's or higher education level and good computer knowledge level. However, participants with secondary school or lower education level and moderate and worse computer knowledge tended toward a lower SGUS score than participants with a bachelor's or higher education level and moderate and worse computer knowledge level (Figure 12e).

From the result of the ANOVA table (Table 9), there is insufficient evidence of statistical significance for two-factor interactions, since all p values are higher than 0.05.

Table 9. ANOVA results for SGUS with regard to organization, role, and computer knowledge level (reducing factors).

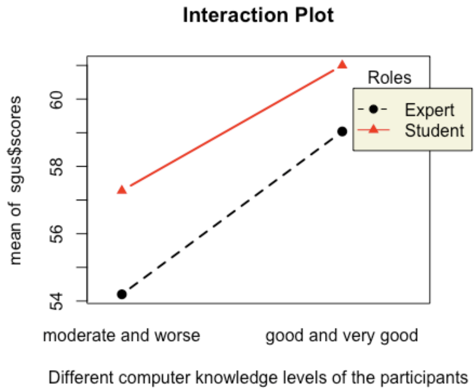
Source of Variation	Df	Sum Sq	Mean Sq	F Value	Pr ($>\lvert F\rvert$)
Organization	2	77.9	38.95	0.576	0.563
Role	1	184.4	184.39	2.729	0.101
Gender	1	0.2	0.19	0.003	0.958
Age	1	4.2	4.16	0.062	0.805
Education level	1	0.0	0.02	0.000	0.988
Computer knowledge	1	589.3	589.31	8.723	0.004 **
Education level: Computer knowledge	1	65.0	64.98	0.962	0.329
Gender: Computer skill	1	121.5	121.49	1.798	0.182
Organization: Computer knowledge	2	28.9	14.47	0.214	0.807
Age: Computer knowledge	1	11.6	11.60	0.172	0.679
Roles: Computer knowledge	1	28.6	28.55	0.423	0.517
Residuals	128	8647.7	67.56		

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

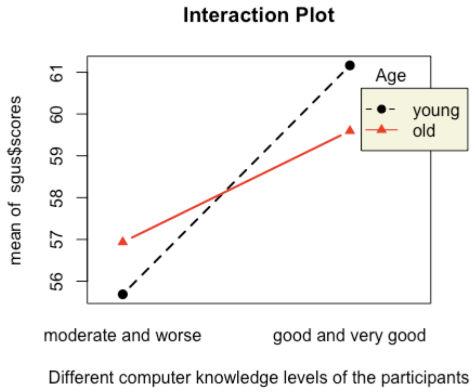


(**a**) (**b**)

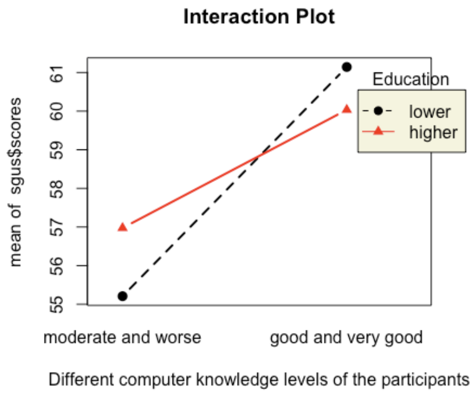
Figure 12. *Cont.*



(c)



(d)



(e)

Figure 12. Interaction effects plots for SGUS: (**a**) different computer knowledge levels with different genders of the participants; (**b**) different computer knowledge levels with different organizations of the participants; (**c**) different computer knowledge levels with different roles of the participants; (**d**) different computer knowledge levels with different age groups of the participants; and (**e**) different computer knowledge levels with different education levels of the participants.

5.3.2. ANOVA of QUIS

In this section, the effect of the six independent variables (age, gender, roles, computer knowledge level, and different organizations) on user interaction satisfaction is reported. Satisfaction includes specific aspects of the interface, usability, and user experience of the AR application.

A total of 133 participants were used for this part of the study and completed the questionnaire. The simple main effects are shown in Figure 13. The results obtained by using the ANOVA in Table 10 indicate that the significance of the two-factor interaction of computer knowledge levels and organizations is not supported since all p values are more than 0.05. Table 10 also shows that the computer knowledge levels and different organizations have a significant effect on QUIS (p value is 0.008 for computer knowledge levels and 0.041 for different organizations).

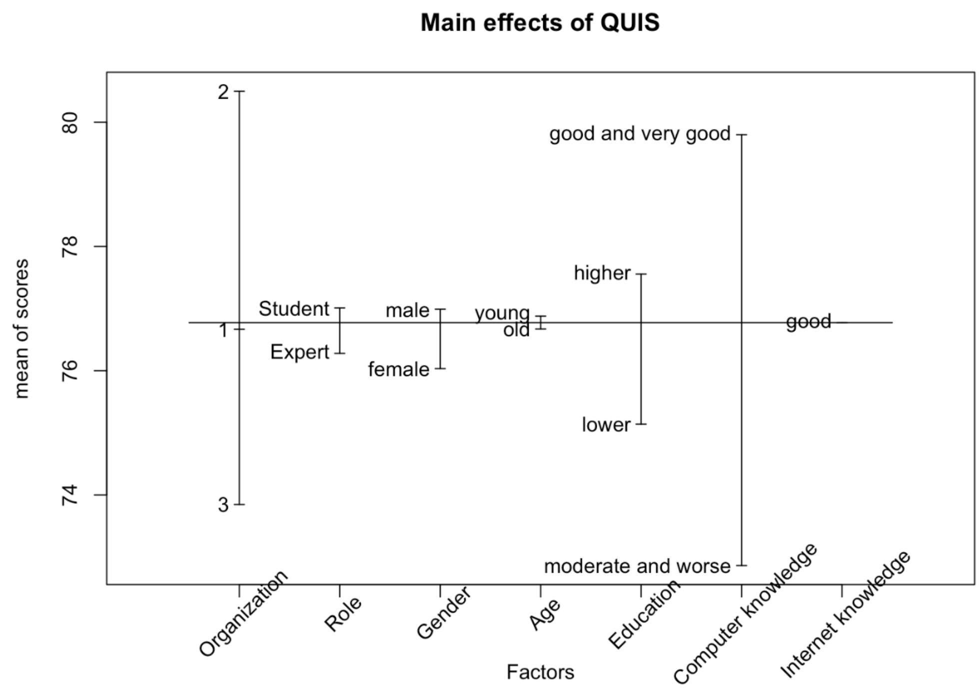


Figure 13. Main effects of QUIS.

Table 10. ANOVA results for QUIS with regard to organization, role, and computer knowledge level (reducing factors).

Source of Variation	Df	Sum Sq	Mean Sq	F Value	Pr (>|F|)
Organization	2	1029.3	514.65	3.279	0.041 *
Role	1	10.4	10.37	0.066	0.798
Gender	1	90.3	90.31	0.575	0.450
Age	1	5.8	5.79	0.037	0.848
Education level	1	32.0	32.02	0.204	0.652
Computer knowledge	1	1138.1	1138.14	7.251	0.008 **
Education level: Computer knowledge	1	165.5	165.55	1.055	0.307
Gender: Computer skill	1	449.7	449.74	2.865	0.093
Organization: Computer knowledge	2	0.9	0.46	0.003	0.997
Age: Computer knowledge	1	28.2	28.18	0.180	0.673
Roles: Computer knowledge	1	31.8	31.84	0.203	0.653
Residuals	119	18679.1	156.97		

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure 14 shows that in all three organizations, participants with moderate or worse computer levels were given lower scores than participants with good and very good computer levels. There are no significant interactions between them.

We selected the factors of organization and computer knowledge level to investigate the interaction between them, and the summary results of the linear model regression (see Table 11) shows that the estimate for the model intercept is 73.533, while there is no significant interaction between them. The information about the standard errors of these estimates is also provided in the coefficients table (Table 11). From the result of the multiple regression model, 10.6% of the variance in QUIS scores is explained by each of the factors (Multiple R-squared is 0.106). There is a statistically significant factor to explain the variation in the QUIS scores (overall p value is 0.0133).

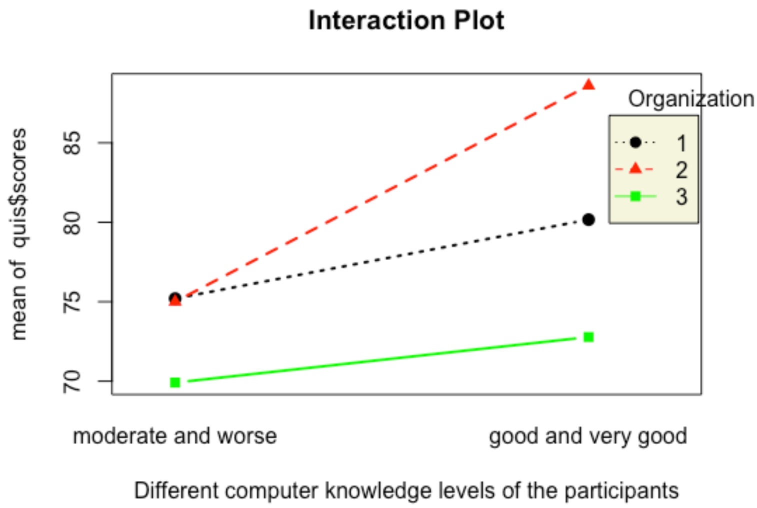


Figure 14. Interaction plot of different computer knowledge levels and the different organizations for QUIS.

Table 11. Summary results of the linear model of the independent variables for QUIS.

Source of Variation	Estimate	Std. Error	t Value	Pr (>|t|)
(Intercept)	73.533	3.188	23.063	$<2 \times 10^{-16}$ ***
Medicine	2.533	4.509	0.562	0.575
Engineering	−2.748	3.951	−0.695	0.488
Computer knowledge	5.092	4.064	1.253	0.213
Medicine: Computer knowledge	1.805	5.686	0.317	0.751
Engineering: Computer knowledge	1.539	5.322	0.289	0.773

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

6. Discussion

This study established a set of norms to be used for the evaluation of satisfaction of using AR glasses and AR applications. The relationship between each questionnaire item shows weak correlation, both in SGUS and in QUIS. Each questionnaire item is designed for evaluating a specific aspect of satisfaction of the smart glasses and AR applications. From the mean score of both questionnaires, we observed that most of the participants are satisfied with the AR glasses and the AR applications. It was found that the system and content helped the participants to accomplish the task quite well and their attention was captivated in a positive way. In other words, the result shows that the user interface is well designed. The user sees "useful information" displayed next to each part.

The main factors age, gender, education level, roles of the participants, and organizations do not have significant effects on the satisfaction of using smart glasses and AR applications. However, computer/Internet knowledge level does influence user satisfaction. Participants who have better computer/Internet knowledge are more satisfied with the smart glasses and AR applications. There is no significant interaction between all these factors. Since most participants have a moderate level or better than moderate level of knowledge using computers and the Internet, it can be predicted that most educated people can easily accept smart glasses and AR applications. The summary of the findings are shown in Table 12.

Table 12. Summary of findings

Hypothesis Number	Description	Accepted/Rejected
H1	Gender matters	Rejected
H2	Age matters	Rejected
H3	Experts and novices will have different level of user satisfaction	Accepted
H4	Education level matters	Rejected
H5	Computer knowledge level matters	Accepted
H6	Internet knowledge level matters	N/A
H7	Three different test-beds might give different results	Accepted in QUIS, Rejected in SGUS
H8	There is interaction effects among all these factors	Accepted in SGUS, Rejected in QUIS

Based on the results associated with the eight hypotheses, we outline the following statements:

Statement 1. Based on the results, we could not identify any gender differences in user satisfaction. It could be a limitation of our experiment set up, as we asked for volunteers, so we ended up with mainly people who were interested in the technology, thus not giving us the option to explore, whether there would be any gender differences in the general population with respect to AR training satisfaction.

Statement 2. Our results suggest that user satisfaction is not influenced by age. A possible explanation for not finding any differences by age could be that the target group had no prior exposure to AR smart glasses, hence age effects of younger people, typically being more open to experimentation of emerging technologies in their home context, could not yet affect the picture.

Statement 3. It is probably to do with our applications. The recorder is a more complicated application, challenging experts in their interaction. Even if, usually, experts would be more technical savvy, in this case, the findings probably reflect more the differences in user friendliness of the applications.

Statement 4. Only the space case had people in higher education. Most participants in the aviation test bed come from upper secondary backgrounds. However, there were no differences found in the impact of education level on user satisfaction. The differences may not be obvious in satisfaction levels, but—judging from observation during trials—there were differences across test-beds with respect to how long it took to explain the applications and their use. The application and the use cases enabled everyone, regardless of whether secondary and tertiary education to use the app.

Statement 5. Computer knowledge possibly matters: Better computer knowledge can drive satisfaction with holographic applications. However, in self assessment tests, users tend to overestimate their computer knowledge [29,30]. This means that it is also possible that user satisfaction levels are not influenced by computer knowledge. It seems that existing knowledge is still relevant. At the same time, this also clearly indicates that the required support and assistance needs to be provided in order to make the introduction of AR applications on smart glasses a success. Not everyone is a digital native.

Statement 6. Internet knowledge matters: All participants in the trial claimed that they have good Internet knowledge and very few people claimed that they have poor Internet knowledge, so there was no chance to observe any differences.

Statement 7. There is no difference between the three test-beds in SUGS: We did not find significant differences between the three test-beds. This indicates that occupation does not have direct influence on satisfaction of the AR glasses. Procedure oriented trainings seem to be covered well. There are some difference between the three test-bed in QUIS. The medicine test-bed have the highest satisfaction of the AR app, while the engineering test-bed gave the lowest scores. The procedures of the tasks might effect the satisfaction of the AR app.

Statement 8. There are no interaction effects for QUIS results but some interaction effects amongst the SGUS results. Young people with good computer knowledge are more satisfied the AR glasses.

People with lower education and good computer knowledge are more satisfied with the AR glasses than the others.

7. Conclusions

This study was started by noting the scarcity of AR applications for hands-on training. As a first step toward incorporating the recorded teaching activities into learning procedures, the AR application was developed on AR glasses. In this work, the Questionnaire for Smart Glasses User Satisfaction (SGUS) and Questionnaire for User Interaction Satisfaction (QUIS) were investigated for augmented reality applications using Microsoft HoloLens.

The results of this study show that the approach is feasible. The experts wore the AR glasses to show the process, and the activities were recorded. The AR applications can facilitate the students to learn the process. The results show that the satisfaction of both teaching and learning are acceptable. The results indicate that satisfaction does increase when participants have higher computer knowledge levels. It also shows that gender, age, education level, and roles of students or experts do not have any effect on user satisfaction.

Author Contributions: H.X., P.S., and F.W. contributed in the conceptualizing, writing and methodology. H.X. and F.W. performed the analysis. H.X. did validation and visualization. P.S. and F.W. helped in the review.

Funding: This study is part of the WEKIT (Wearable Experience for Knowledge Intensive Training) project. It is a European project supported under Horizon 2020 and is aiming to enable smart wearable technology for industrial training.

Acknowledgments: We express sincere appreciation to Kaj Helin, Principal Scientist, Project Manager of VTT Technical Research Centre of Finland, and Timo Kuula, Research Scientist of VTT Technical Research Centre of Finland for their modification of questionnaire QUIS and their design of questionnaire SGUS. We thank all the participants and companies for their valuable contributions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
ATV	Automated Transfer Vehicle
CTB	Cargo Transfer Bag
ISS	International Space Station
UI	user interface
QUIS	Questionnaire for User Interaction Satisfaction
SGUS	Smart Glasses User Satisfaction
TSR	Temporary Stowage Rack
WEKIT	Wearable Experience for Knowledge Intensive Training

Appendix A

Table A1. Spearman's rank coefficient of correlation of QUIS: the first value of each row represents Spearman's correlation coefficient, and the second value of each row represents the p value.

	QS1	QS2	QS3	QS4	QS5	QS6	QS7	QS8	QS9	QS10	QS11	QS12	QS13	QS14	QS15
QS1	1	0.39	0.53	0.50	0.49	0.34	0.53	0.47	0.17	0.37	0.37	0.28	0.31	0.32	0.44
		0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.05	0.000 **	0.000 **	0.001	0.000 **	0.000 **	0.000 **
QS2		1	0.53	0.41	0.34	0.30	0.37	0.43	0.16	0.35	0.25	0.33	0.58	0.50	0.52
			0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.07	0.000 **	0.003	0.000 **	0.000 **	0.000 **	0.000 **
QS3			1	0.56	0.55	0.39	0.49	0.45	0.16	0.33	0.27	0.22	0.35	0.37	0.40
				0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.06	0.000 **	0.001	0.009	0.000 **	0.000 **	0.000 **
QS4				1	0.49	0.23	0.42	0.41	0.18	0.35	0.38	0.30	0.27	0.27	0.40
					0.000 **	0.008	0.000 **	0.000 **	0.04	0.000 **	0.000 **	0.000 **	0.001	0.001	0.000 **
QS5					1	0.22	0.41	0.45	0.14	0.22	0.13	0.14	0.24	0.36	0.34
						0.01	0.000 **	0.000 **	0.11	0.01	0.12	0.10	0.005	0.000 **	0.000 **
QS6						1	0.36	0.26	0.26	0.18	0.25	0.09	0.28	0.33	0.33
							0.000 **	0.002	0.001	0.03	0.003	0.28	0.000 **	0.000 **	0.000 *
QS7							1	0.54	0.17	0.38	0.39	0.28	0.24	0.35	0.44
								0.000 **	0.05	0.000 **	0.000 **	0.001	0.004	0.000 **	0.000 **
QS8								1	0.23	0.40	0.26	0.33	0.26	0.40	0.43
									0.006	0.000 **	0.002	0.000 **	0.002	0.000 **	0.000 **
QS9									1	0.35	0.31	0.31	0.19	0.32	0.24
										0.000 **	0.000 **	0.000	0.024	0.000 **	0.005
QS10										1	0.57	0.45	0.27	0.29	0.44
											0.000 **	0.000 **	0.001	0.001	0.000 **
QS11											1	0.43	0.25	0.33	0.38
												0.000 **	0.003	0.000 **	0.000 **
QS12												1	0.34	0.30	0.42
													0.000 **	0.000 **	0.000 **
QS13													1	0.57	0.48
														0.000 **	0.000 **
QS14														1	0.47
															0.000 **
QS15															1

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

References

1. Wild, F.; Perey, C. Draft standard for Augmented Reality Learning Experience Model. *IEEE Stand. Assoc.* **2018**, unpublished report.
2. Azuma, R.T. A survey of augmented reality. *Teleoperators Virtual Environ.* **1997**, *6*, 355–385. [CrossRef]
3. Sherstyuk, A.; Vincent, D.; Berg, B.; Treskunov, A. Mixed reality manikins for Medical Education. In *Handbook of Augmented Reality*; Springer: New York, NY, USA, 2011; pp. 479–500
4. Ong, S.K.; Shen, Y.; Zhang, J.; Nee, A.Y.C. Augmented Reality in Assistive Technology and Rehabilitation Engineering. In *Handbook of Augmented Reality*; Springer: New York, NY, USA, 2011; pp. 603–630.
5. Nilsson, J.; Ödblom, A.C.E.; Fredriksson, J.; Zafar, A. Using Augmentation Techniques for Performance Evaluation in Automotive Safety. In *Handbook of Augmented Reality*; Springer: New York, NY, USA, 2011; pp. 631–649.
6. Ras, E.; Wild, F.; Stahl, C.; Baudet, A. Bridging the Skills Gap of Workers in Industry 4.0 by Human Performance Augmentation Tools. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments—PETRA '17, Island of Rhodes, Greece, 21–23 June 2017 .
7. Perey, C.; Wild, F.; Helin, K.; Janak, M.; Davies, P.; Ryan, P. Advanced manufacturing with augmented reality. In Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 10–12 September 2014; doi:10.1109/ismar.2014.6948518. [CrossRef]
8. Olsson, T. Concepts and Subjective Measures for Evaluating User Experience of Mobile Augmented Reality Services. In *Human Factors in Augmented Reality Environments*; Springer: New York, NY, USA, 2013; pp. 203–232.
9. Augmented Reality: What Does It Mean for UX? Nielsen Norman Group. Available online: https://www.nngroup.com/articles/augmented-reality-ux/ (accessed on 5 July 2018).
10. Microsoft. Microsoft HoloLens. Microsoft Image Gallery. Available online: https://news.microsoft.com/mshololens_skype_03495_16x9_rgb/ (accessed on 24 January 2019).
11. Rauschnabel, P.A.; Brem, A.; Ro, Y.K. Augmented Reality Smart Glasses: Definition, Conceptual Insights, and Managerial Importance.Unpublished Working Paper, The University of Michigan-Dearborn, College of Business (2015). Available online: https://www.researchgate.net/profile/Alexander_Brem/publication/279942768_Augmented_Reality_Smart_Glasses_Definition_Conceptual_Insights_and_Managerial_Importance/links/5721ec2e08aee857c3b5dd6c/Augmented-Reality-Smart-Glasses-Definition-Conceptual-Insights-and-Managerial-Importance.pdf (accessed on 12 July 2018).
12. Rauschnabel, P.A.; Brem, A.; Ivens, B.S. Who will buy smart glasses? Empirical results of two pre-market-entry studies on the role of personality in individual awareness and intended adoption of Google Glass wearables. *Comput. Hum. Behav.* **2015**, *49*, 635–647. [CrossRef]
13. Wiederhold, B.K. Time to port augmented reality health apps to smart glasses? *Cyberpsychol. Behav. Soc. Netw.* **2013**, *16*, 157–158. [CrossRef]
14. Questionnaire For User Interaction Satisfaction. Available online: http://lap.umd.edu/quis/ (accessed on 29 January 2018).
15. Limbu, B.; Fominykh, M.; Klemke, R.; Specht, M.; Wild, F. Supporting Training of Expertise with Wearable Technologies: The WEKIT Reference Framework. In *Mobile and Ubiquitous Learning. Perspectives on Rethinking and Reforming Education*; Yu, S., Ally, M., Tsinakos, A., Eds.; Springer: Singapore, 2018.
16. Puneet, S.; Roland, K.; Fridolin, W. Experience Capturing with Wearable Technology in the WEKIT project. In Proceedings of the SIGWELL, Munich, Germany, 18–19 October 2018.
17. WEKIT D2.5 WEKIT.one (Final Prototypes). Available online: http://wekit.eu/category/results/deliverables/ (accessed on 28 December 2018).
18. WEKIT D6.4 Implementation of Evaluation Trials in Aeronautics. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.4.pdf (accessed on 7 November 2017).
19. Fominykh, M. D2.4 First Prototype. Available online: http://wekit.eu/d2-4first-prototype/ (accessed on 2 September 2018).
20. WEKIT D1.4 Requirements for Scenarios and Prototypes. Available online: http://wekit.studiohangloose.it/wp-content/uploads/2017/06/WEKIT_D1.4.pdf (accessed on 7 November 2017).
21. WEKIT Community. Available online: https://wekit-community.org/ (accessed on 28 December 2018).

22. WEKIT D6.2 Annex 1 Training Scenario and Evaluation Plan for Engineering. Available online: http://wekit.studiohangloose.it/wp-content/uploads/2017/06/WEKIT_D6.2.pdf (accessed on 9 November 2017).
23. WEKIT D6.5 Implementation of Evaluation Trials in Engineering. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.5 (accessed on 9 November 2017).
24. WEKIT D6.6 Implementation of Evaluation Trials in Space. Available online: http://wekit.eu/wp-content/uploads/2017/09/WEKIT_D6.6 (accessed on 5 June 2018).
25. Ssemugabi, S.; de Villiers, R. A comparative study of two usability evaluation methods using a web-based e-learning application. In Proceedings of the 2007 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries—SAICSIT '07, Port Elizabeth, South Africa, 2–3 October 2007.
26. Chin, J.P.; Diehl, V.A.; Norman, K.L. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 1988; pp. 213–218.
27. Kerkhoven, A.H. Gender stereotypes in science education resources: A visual content analysis. *PLoS ONE* **2016**, *11*, e0165037. [CrossRef] [PubMed]
28. Hall, D.T.; Mirvis, P.H. The new career contract: Developing the whole person at midlife and beyond. *J. Vocat. Behav.* **1995**, *47*, 269–289. [CrossRef]
29. Ballantine, J.A.; Larres, P.M.; Oyelere, P. Computer usage and the validity of self-assessed computer competence among first-year business students. *Comput. Educ.* **2007**, *49*, 976–990. [CrossRef]
30. Patricia, L.M.; Ballantine, J.; Whittington, M. Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities. *Account. Educ.* **2003**, *12*, 97–112.

Paper II

Development of a SAGAT query and simulator experiment to measure situation awareness in maritime navigation

H. Xue, B. M. Batalden, J. F. Røds

In: Stanton, N. (Eds.) Advances in human aspects of transportation. AHFE 2020. *Adv. Intell. Syst. Comput.*, vol. 1212. Springer, Cham.

https://doi.org/10.1007/978-3-030-50943-9_59

Development of a SAGAT Query and Simulator Experiment to Measure Situation Awareness in Maritime Navigation

Hui Xue, Bjørn-Morten Batalden(✉), and Johan-Fredrik Røds

Faculty of Science and Technology, Department of Technology and Safety, UiT The Arctic University of Norway, Pb 6050, Langnes, Tromsø 9037, Norway
{hui.xue, bjorn.batalden, johan-fredrik.rods}@uit.no

Abstract. Many ship collisions and groundings occur due to navigators' erroneous situation awareness (SA). The objective of this study is to develop a method to measure SA for maritime navigation and collision avoidance (SA-MA). This study uses the Situation Awareness Global Assessment Technique (SAGAT) as a basis and tool to assess SA. Both interviews with experts and simulator experiments are used. Ten participants, five navigators with extensive experience, and five second-year students at a nautical science program participate in the simulator experiment. Hierarchical Task Analysis (HTA) is used to map the navigation and collision avoidance tasks as input to the SA queries. The objective measurements collected from the simulator and subject matter experts are used for the SAGAT score. A well-developed SAGAT query and simulator experiment results in a difference in the SA-MA between the experienced navigators and the students with less experience. The study found it is difficult to measure SA-MA, especially for level 2 and 3 SA.

Keywords: Situation Awareness (SA) · SAGAT · Maritime · Navigation · Hierarchical Task Analysis (HTA) · Simulator

1 Introduction

Many ship collisions and groundings occur due to navigators' erroneous situation awareness (SA). In particular, unsafe acts and preconditions for unsafe acts are important causes for ship collisions and groundings [1]. For both ship collisions and groundings, decision-errors and resource management are the two most frequent causes [1]. Grech et al. [2] found that 71% of navigators' errors are SA related problems.

The concept of SA is based on the interaction between the operator and the surrounding environment [3]. There is an underlying assumption that the situation in the working environment can be changed in different ways. For example, it can change fast or slow, significantly or not significantly, obviously or concealed, repeatedly or not repeatedly, planned or unplanned [4]. In the maritime domain, during the different tasks related to sailing a ship, the navigator should be able to adjust and adapt the performance based on the current situation or the change of the situation, taking future development into account. SA of navigators can be improved by training and practising

N. Stanton (Ed.): AHFE 2020, AISC 1212, pp. 468–474, 2020.
https://doi.org/10.1007/978-3-030-50943-9_59

[5, 6]. Therefore, in order to improve the SA of navigators, developing reliable and valid measures of SA has been the focus of this experiment.

The objective of the research is to develop and assess a Situation Awareness Global Assessment Technique (SAGAT) Query to measure SA-MA.

2 Theory of SA and SA in Maritime

Endsley and Jones [7] explain situation awareness as "being aware of what is happening around you and understanding what that information means to you and in the future". It relates to what is important for a task or goal. Several definitions of situation awareness exist [8], but in this paper, we choose to use the definition "Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." [9].

Through human sensory systems, being either vision, audition, vestibular system, the somatic sensory system, gustation or olfaction, it is possible to perceive information about the elements (Proctor and Proctor, 2006) [10]. The use of these sensory systems will vary depending on the domain and type of job or task, and the data necessary to achieve or address level 1 SA can be hard to map in many domains [7]. Endsley [9] identify that SA challenges mostly relate to level 1 SA.

Level 2 SA is to convert the sensory information to create an understanding of the current situation. Regarding level 2 SA for a deck officer on a ship, Sharma et al. (2019) [8] present information elements that refer to the parameter Closest Point of Approach (CPA) and Time to CPA (TCPA) as examples. The novice operator may achieve the same level 1 SA as the experienced operator but not being able to convert this information into level 2 SA, achieving a lower level 2 SA [7].

The last level from the definition, level 3 SA, is how an operator manages to translate the information gathered, and understanding of the current state, into a future state. Endsley and Jones [7], state that good level 3 SA can only be achieved from an operator having a sound level 2 SA and an understanding of the "functioning and dynamics of the system they are working in". To achieve a good level 3 SA requires sound domain understanding and time spent on achieving good level 3 SA is often extensive among experts Endsley and Jones [7]. Sharma et al. [8] refer to extending vectors from targets and radar trials as methods to assist in achieving good level 3 SA for deck officers. Insufficient mental capacity and insufficient knowledge of the domain are two possible reasons for not achieving a good level 3 SA [7].

3 Methodology

This study uses both interviews with experts and a simulator experiment to develop and assess the SAGAT for SA-MA. The interviewees have extensive experience as navigators in both the merchant fleet and the navy. The study uses Hierarchical Task Analysis (HTA) to list the navigation and collision avoidance tasks as input to the SA queries. The simulator experiment uses ten participants, five navigators with extensive experience, and five second-year students at a nautical science program with little experience. The 240° view simulator used for the experiment is equipped with the K-sim Navigation software from Kongsberg Digital. The vessel-model used in the experiment is called BULKC11 (length overall of 90 m and a moulded beam of 14 m). The procedure of developing SAGAT Queries is shown in Fig. 1.

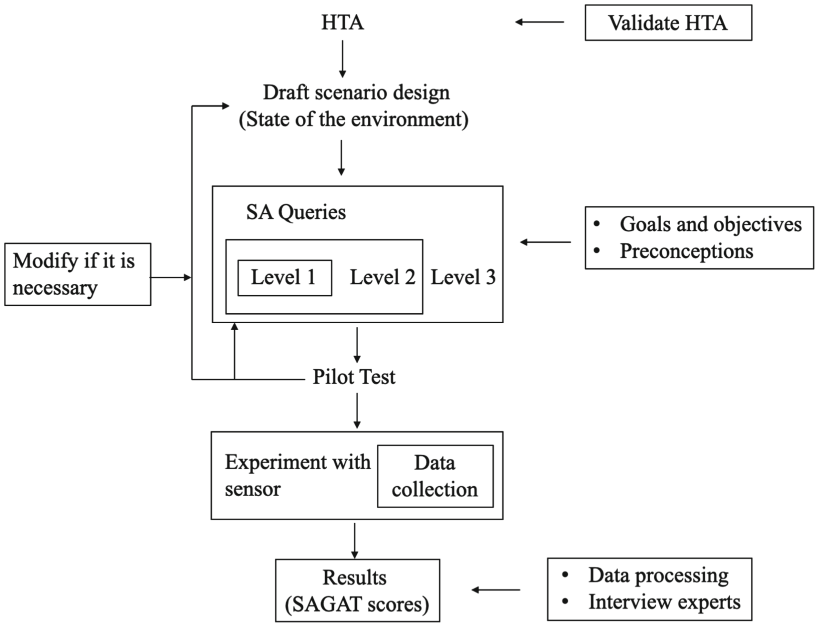


Fig. 1. The procedure of developing the SAGAT Queries.

In this study, the HTA was developed in the navigation tasks. Each task has a goal to achieve. The carry out of an HTA can be adapted to different situation and needs [11]. Before starting the process of developing the HTA, a literature study was conducted to enhance the knowledge regarding the field of research. A draft HTA was then discussed with a subject matter expert for input. The revised HTA was presented to a second subject matter expert and further adjusted and finalised.

SAGAT was initially developed to assess the SA of pilots across the three levels of SA [6, 7]. The procedure of SAGAT comprise of ten steps: (1) Define tasks, (2) Development of SA queries, (3) Selection of participants, (4) Brief participants, (5) Pilot run, (6) Task performance, (7) Freeze the simulation, (8) SA query administration, (9) Query answer evaluation (by a subject matter expert), (10) SAGAT score calculation [5].

The results of the SA requirements analysis are used to develop a set of SA queries for the experiment in the simulator [12]. As a global measure, SAGAT includes queries about all operator SA requirements, including Level 1, Level 2 and Level 3 components [12, 13]. Participants were briefed regarding the purpose of the study and the voyage plan. There were four stops where the simulator was frozen, handing out the queries based on all three levels of SAGAT. For the four stops, each of the stops was conducted within a fixed range on the course line (Fig. 2). In total, it is an approximate forty minutes voyage. An expert completed the same SA Queries with the correct answer on the simulator. The participants' answers were compared to the results of the expert.

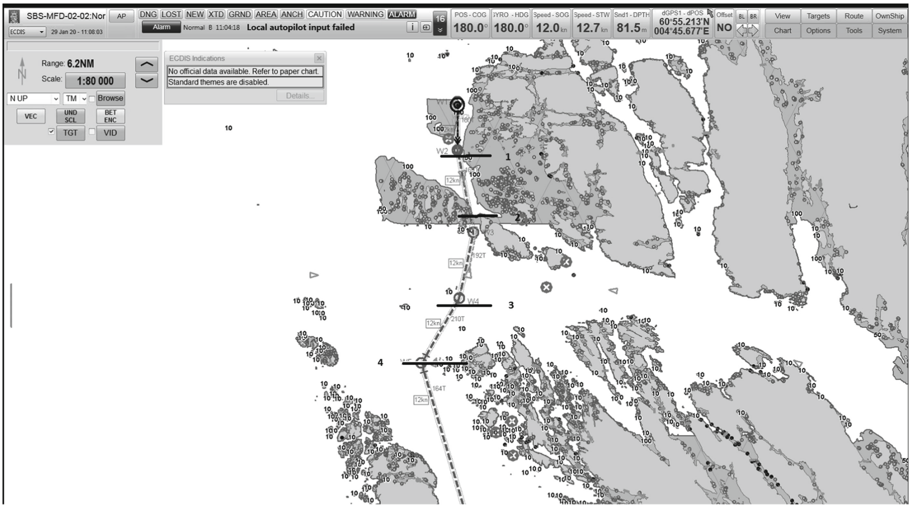


Fig. 2. Chart layout with the route and horizontal lines indicating the four SAGAT stops.

Five expert participants were interviewed face to face (Mean age = 41.8, Stand deviation = 14.0). The interviews were semi-structured with the purpose to elicit expert knowledge and experience from the SAGAT queries and simulator experiment [8, 14]. The average length of the interviews was around 10 min. The participants have an average of 9.7 years' experience as navigators with the longest being 18 years and the shortest being one year.

4 Results

The results were collected from both the students and experts SA scores and interviews with the experts. Each of the results will be presented respectively. The analysis identified that the students and experts have different performance in each of the three SA levels. The results and analysis of the SA scores are presented in Table 1. It was possible to achieve in total a maximum of 45 points on all three levels together. The results of the SA scores show that the students have 42.7% correct in total, while the experts get 46.2% in total. Students score higher than the experts do on level 1 SA while the experts score higher on level 2 SA. For level 3 SA, the students and experts score close to the same, the students scoring 1.8% points above the experts. The scores show that students got better scores than experts on level 1, experts got better scores on level 2, and they have similar scores on level 3.

Table 2 breaks the results further down and show each stop for level 1, 2 and 3 SA. There is no difference between the expert and student on level 3 SA except for the last stop. For level 2 SA, experts scored noticeably higher than the students on all four stops.

Table 1. Results of the SA scores at different levels and different stops.

Role	Level 1	Level 2	Level 3	1st Stop	2nd Stop	3rd Stop	4th Stop	Overall
Student	67.1%	25.9%	30.9%	37.3%	38.5%	48.6%	60.0%	42.4%
Expert	58.8%	44.7%	29.1%	42.7%	46.2%	50.0%	46.7%	46.2%

Table 2. Results of the SA scores at all levels of SA for all four stops.

Role	1st Stop			2nd Stop		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Student	72.0%	16.0%	24.0%	48.0%	36.0%	26.7%
Expert	64.0%	40.0%	24.0%	56.0%	48.0%	26.7%
Role	3rd Stop			4th Stop		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Student	76.7%	26.7%	30.0%	80.0%	20.0%	80.0%
Expert	60.0%	46.7%	30.0%	40.0%	40.0%	60.0%

The interviews with the experts gave valuable insight into possible strengths and weaknesses with the SAGAT queries. The development of SAGAT queries will be scenario- and time-dependent. As an example, two different candidates conducting the same simulator experiment but with slightly different location or time when the query is conducted may have shifted what needs to be focused on by the candidate. Sometimes the time to arrive a certain position is important while for others the position, as a result of speed and time, is important. The experts meant that it is more important to know where two vessels meet than to know the time when they meet. The candidates need to focus on many information sources, and it is less likely that they remember details from all sources when the query is handed out. Most of the experts agreed that for different situations, some information is not important to remember. It is also not always important to remember how many target ships there are when there is not too much traffic.

5 Discussion and Concluding Remarks

From a maritime point of view, it is found to be difficult to measure SA, especially for level 2 and 3. The participants' answers from the queries are analysed by a subject matter expert, which decides if the answers are correct, or within the acceptable range. For level 1 SA, the information necessary to analyse the answers from the queries can be collected directly from the simulator. For level 2 SA and level 3 SA, it is more complex to analyse the answers from the queries. Experts may have to decide if the answers are within an acceptable range based on their experience in the given situations. This may be a challenge since different experts may have different experiences and different opinions. Further, the experts have to combine the information and their experience in order to analyse the situation.

Several explanations may be given to why students score higher on level 1 SA. Firstly, the students may be more focused on the information on the screens but have a lower capacity to understand and utilise the information to comprehend the situation. Secondly, experts have more experience in selecting important information in a situation. Experts may be better at selecting relevant information while the SA query tries to measure too much information. This is in line with Sharma et al. [8]. Thirdly, better memory might also help the student group to get higher scores on level 1 SA since research shows that memory loss is age-related [15, 16]. Fourthly, the students are familiar with the instruments since they practice very often on the simulator while most of the experts are less familiar with the instruments, and rely more on looking out and just collect basic information.

When it comes to level 2 SA, experts got markedly higher scores than students, which may indicate that experts are better in converting the sensory information to the understanding of the current situation. This is in line with the results of Endsley and Jones [7], that novice are not being able to convert this information into level 2 SA. Another explanation could be that some part of the information asked for in the queries did not affect the experts understanding of the current situation. Based on the results of the leve1 3 SA, it indicates that the students and experts have a similar capacity of translating the current state into a future state. This is an unexpected result. It might be that the existing queries are not good at measuring level 3 SA. This is supported by the difference detected from level 1 SA to level 2 SA.

SAGAT queries should perhaps focus more on the required situation rather than specific parameters such as CPA and TCPA. Based on the interviews and the experiments, it is necessary to spend more time on making clearer questions in the queries. The participants need to understand the questions correctly. A well-developed SAGAT query and simulator experiment should result in a difference in the SA-MA between the experienced navigators and the students with less experience. For future studies, it is recommended to create queries that enable utilisation of all tools in the simulator to reduce the subjective assessment. A tested SAGAT for the SA-MA will be a valuable tool for future studies related to maritime navigation and collision avoidance and as a tool to assess the training of the nautical students. With an overall score of between 42.4 and 46.2, there is a need to further develop the SAGAT query for SA-MA.

Acknowledgments. The authors would like to acknowledge the assistance from Ingar Lorentsen and Øyvind Haugseggen at UiT the Arctic University of Norway during data collection on the simulator.

References

1. Yıldırım, U., Başar, E., Uğurlu, Ö.: Assessment of collisions and grounding accidents with human factors analysis and classification system (HFACS) and statistical methods. Saf. Sci. **119**, 412–425 (2019)
2. Grech, M.R., Horberry, T., Smith, A.: Human error in maritime operations: analyses of accident reports using the Leximancer tool. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage Publications, Los Angeles (2002)
3. Endsley, M.R., Garland, D.J.: Situation Awareness Analysis and Measurement. CRC Press, Boca Raton (2000)
4. Koester, T.: Situation awareness and situation dependent behaviour adjustment in the maritime work domain. Hum.-Centered Comput. Cogn. Soc. Ergon. Aspects **3**, 255 (2019)
5. Salas, E., et al.: Does crew resource management training work? An update, an extension, and some critical needs. Hum. Factors **48**(2), 392–412 (2006)
6. Stout, R.J., Salas, E., Fowlkes, J.E.: Enhancing teamwork in complex environments through team training. Group Dyn. Theory Res. Pract. **1**(2), 169 (1997)
7. Endsley, M.R., Jones, D.G.: Designing for Situation Awareness: Understanding Situation Awareness in System Design. CRC Press, Boca Raton (2011)
8. Sharma, A., Nazir, S., Ernstsen, J.: Situation awareness information requirements for maritime navigation: a goal directed task analysis. Saf. Sci. **120**, 745–752 (2019)
9. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference. IEEE (1988)
10. Proctor, R.W., Proctor, J.D.: Sensation and perception. In: Handbook of Human Factors and Ergonomics, pp. 51–88 (2006)
11. Annett, J.: Hierarchical task analysis. Handb. Cogn. Task Des. **2**, 17–35 (2003)
12. Stanton, N.A., et al.: Human Factors Methods: A Practical Guide for Engineering and Design. CRC Press, Boca Raton (2017)
13. Endsley, M.R.: Direct measurement of situation awareness: validity and use of SAGAT. In: Situational Awareness, pp. 129–156. Routledge (2017)
14. Kokar, M.M., Endsley, M.R.: Situation awareness and cognitive modeling. IEEE Intell. Syst. **27**(3), 91–96 (2012)
15. Small, G.W.: What we need to know about age related memory loss. BMJ **324**(7352), 1502–1505 (2002)
16. Luszcz, M.A., Bryan, J.: Toward understanding age-related memory loss in late adulthood. Gerontology **45**(1), 2–9 (1999)

Paper III

Biosignals-based driving skill classification using machine learning: A case study of maritime navigation

H. Xue, B.-M. Batalden, P. Sharma, J. A. Johansen, D. K. Prasad

Appl. Sci., vol. 11, no. 20, p. 9765.

https://doi.org/10.3390/app11209765

applied sciences

Article

Biosignal-Based Driving Skill Classification Using Machine Learning: A Case Study of Maritime Navigation

Hui Xue [1,*], Bjørn-Morten Batalden [1], Puneet Sharma [2], Jarle André Johansen [2] and Dilip K. Prasad [3]

[1] Department of Technology and Safety, UiT The Arctic University of Norway, 9019 Tromsø, Norway; bjorn.batalden@uit.no
[2] Department of Automation and Processing Technology, UiT The Arctic University of Norway, 9019 Tromsø, Norway; puneet.sharma@uit.no (P.S.); jarle.johansen@uit.no (J.A.J.)
[3] Department of Computer Science, UiT The Arctic University of Norway, 9019 Tromsø, Norway; dilip.prasad@uit.no
* Correspondence: hui.xue@uit.no

Abstract: This work presents a novel approach to detecting stress differences between experts and novices in Situation Awareness (SA) tasks during maritime navigation using one type of wearable sensor, Empatica E4 Wristband. We propose that for a given workload state, the values of biosignal data collected from wearable sensor vary in experts and novices. We describe methods to conduct a designed SA task experiment, and collected the biosignal data on subjects sailing on a 240° view simulator. The biosignal data were analysed by using a machine learning algorithm, a Convolutional Neural Network. The proposed algorithm showed that the biosingal data associated with the experts can be categorized as different from that of the novices, which is in line with the results of NASA Task Load Index (NASA-TLX) rating scores. This study can contribute to the development of a self-training system in maritime navigation in further studies.

Keywords: biosignal; maritime navigation; classification; situation awareness (SA); neural network; maritime training

Citation: Xue, H.; Batalden, B.-M.; Sharma, P.; Johansen, J.A.; Prasad, D.K. Biosignal-Based Driving Skill Classification Using Machine Learning: A Case Study of Maritime Navigation. *Appl. Sci.* **2021**, *11*, 9765. https://doi.org/10.3390/app11209765

Academic Editors: YangQuan Chen, Subhas Mukhopadhyay, Nunzio Cennamo, M. Jamal Deen, Junseop Lee and Simone Morais

Received: 26 September 2021
Accepted: 15 October 2021
Published: 19 October 2021

1. Introduction

The classic methodology for maritime training generally involves multiple sensors in addition to simulator for improving situation awareness (SA) in maritime navigation and seafaring skills [1]. The purpose of this study is to determine if a wearable sensor can be used to detect stress changes with skills during a maritime navigation task. We define stress as the task requirement for both experienced seafarers (experts) and novices (students). We collected the biosignal data of subjects for indicating the stress differences under the SA tasks during maritime navigation. Biosignal data including electrodermal activity (EDA), body temperature, blood volume pulse (BVP) and heart rate (HR) are some of the indicators to present the stress level, since stress is the body's reaction to pressure and a physical response to situations in which people feel threatened.

Safe maritime navigation in the Arctic region is challenging because there is less infrastructure, long distances between harbours and harsh weather conditions [2]. However, the safety of the Arctic route is of great significance to the economic development of Scandinavia, and at the same time has an impact on environmental protection and the safe growth of marine life. The sailing route in the Vessel Traffic Services area on the west coast of Norway, north of Bergen, is a typical route for training the seafarers because of its complexity and busy traffic, especially for training SA in the safety of the maritime navigation.

In maritime, the study of situation awareness (SA) has always been an important topic of discussion. Studies show that many ship collisions and groundings occur due to navigators' erroneous SA. Grech et al. [3] found that 71% of navigators' errors can be attributed to SA-related problems. Therefore, training maritime students to improve their SA is one of the most important tasks in maritime education.

In maritime navigation, an experienced navigator can keep track of multiple tasks and deal with more complex situations without losing SA as compared to a novice. For a novice, managing multiple tasks required for navigation can be quite challenging [4]. As navigating a ship can be stressful, managing this stress can bring different results. In this study, we aim to investigate whether there are differences in the biosignals between the experts and novices for a given sailing task.

1.1. Related Research Work

There are only a few works related to performance assessment of SA objectively during maritime training navigation. In the existing literature, survey and interview are usually the common tools for assessing SA. From its conception, SA was defined by Endsley in 1988 as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, the projection of their status in the near future" [5,6]. In simple terms, it can be understood as "being aware of what is happening around you and understanding what information means to you now and in the future" [6]. In maritime operations, the awareness refers to the important information for sailing and being safe on board (particular job or goals in general), and only the situations that relate to the tasks are important to SA. For example, the navigator of a ship must be aware of other ships, the weather, the water, the grounding, and so on. When sailing a ship, during the different tasks, the navigator should be able to adjust and adapt the performance based on the current situation or the change of the situation. In order to improve the training system for SA within maritime navigation, several studies of sensor fusion technology have been used on simulators in the past few years [1,7]. However, as far as we know, use of bio-sensor data with the SA training system is not common. The main contribution of this paper is to provide a possibility for doing such research.

1.2. Objective and Contributions

The main objective of this study is to investigate the differences in the stress levels of experts and novices in SA experiment during a maritime navigation task. In this study, we investigate the following research hypotheses: First, the biosensor data from experienced experts can be distinguished from the biosensor data from students. Second, the stress levels obtained from the biosensor data show a correlation to the NASA-TLX rating results. Third, compared to the novices, the experts feel less stress during a navigation task.

The main contribution can be summarised as follows:

- Maritime transport requires safety and security. SA in maritime is the effective understanding of activity that could impact the security and safety. This study discovers that the stress level varies according to the experience of the seafarers, which matters in the performance of SA during the maritime navigation.
- SA training is common in the aviation domain and the maritime domain. While SA training-related stress level analysis is widely studied for aviation, SA training-related stress level analysis in maritime navigation is less studied. This paper is a pilot case study towards classifying stress level among the expert and novice seafarers. We have used a 'hybrid' convolutional neural network approach in combination with statistical, wavelet and higher-order crossing features to classify the stress level based on biosignals during the maritime navigation. We are first extracting features and then passing those features to a Convolutional Neural Network so we refer to is as a 'hybrid Convolutional Neural Network'.

2. Methodology

In order to study the SA-based stress level analysis, we used the Kongsberg K-Sim navigation platform. Both expert and novice drivers were given the same driving scenario and their biosignals were recorded using Empatica E4 band while they were driving on the simulator. In the next sub-section we elaborate on the information about participants, materials and apparatus used for this study.

2.1. Participants

The trial was performed with 10 healthy male participants. In order to compare the performance and emotion between the experts and novice in maritime navigation tasks, both experts and novices were invited to participate in the experiment. There were five navigators with extensive experience (mean age = 41.8 years, standard deviation = 14.0 years) and five second-year students from a nautical science program with little experience (mean age = 22.8 years, standard deviation = 1.2 years). The experts had an average of 9.7 years' experience as navigators with the longest period being 18 years and the shortest being one year.

2.2. Materials and Apparatus

A maritime navigation task was designed for testing the relationship between the navigating experience and stress. The maritime navigation task was performed on a 240° view simulator. It is equipped with the K-sim Navigation software from Kongsberg Digital. The vessel-model used in the experiment is called BULKC11 (overall length of 90 m and a moulded beam of 14 m). The task consists of two part, one part is sailing, the other part is filing the The Situation Awareness Global Assessment Technique (SAGAT) queries when the simulator screen is frozen. Each participant sailed a 40-minute voyage with four stops. Each section of the sailing lasts approximately 8 to 12 min. During the sailing section, participants had to complete the SAGAT queries in around 15 min (4 stops with an average of 4 min to answer the SAGAT queries). The whole experiment takes approximately 55 min. Figure 1 shows a participant sailing on the simulator.

Figure 1. One of the participants was sailing on the simulator.

During the experiment, each participant wore a wearable device for collecting the biosignal data. In this study, among the diversity of wearable sensors, a medical-grade wearable device, Empatica E4 Wristband (see Figure 2), was chosen for recording the real-time physiological data to conduct in-depth analysis. EDA and PPG sensors were equipped in the E4 Wristband that can simultaneously enable the measurement of sympathetic nervous system activity and heart rate [8]. Following is the description of the sensors in the E4 Wristband:

- **PPG Sensor:** Measures blood volume pulse (BVP), from which heart rate variability can be derived [8];

- **Infrared Thermopile:** Reads peripheral skin temperature [8];
- **EDA Sensor (GSR Sensor):** Measures the constantly fluctuating changes in certain electrical properties of the skin [8];
- **3-axis Accelerometer:** Captures motion-based activity [8].

The data from E4 Wristband such as Electrodermal Activity (EDA), body temperature, blood volume pulse (BVP) and heart rate (HR) are collected and used in the analysis.

Figure 2. Empatica E4 wristband [8].

3. Experiment

In this experiment, we used NASA Task Load Index (NASA-TLX) as a reference. The rating result of NASA-TLX is a subjective measurement evaluated by the participants themselves. The result show that there are different workload and stress level between experts and students. In the light of this result, we hypothesize that it is possible to classify the biosignal data we collected during the sailing task. Hence, we extracted the features of the data and analyzed it by using convolutional neural network(CNN) in deep learning.

3.1. NASA Task Load Index (NASA-TLX)

The NASA Task Load Index (NASA-TLX) was used as an assessment tool to rate the perceived workload in order to assess the performance of the participants [9,10]. The six categories were required to be rated from low to high level, namely, Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. The rating was transferred to a ten-point scaler scores. All the participants were given NASA-TLX after the experiment.

NASA-TLX Rating Results

There are two ways of analysing the NASA-TLX scores: one is a two-step process that needs participants to give both scores for each item and a pairwise comparison score between each pair of items (there will be in total 15 pairwise comparisons); another way is simple and convenient—calculating the average score of the six items for each participant [11,12]. In this study, only the "Raw TLX" was used [13] and the average score was calculated in order to keep the experimental validity [12]. When using the "raw TLX", individual subscales may be dropped if less relevant to the task [14].

Figure 3 shows the average scores from the experts and students. Students considered the workload was higher than expected by experts, however, it was not highly significant. Figure 4 shows the comparison of the raw scores for students, experts and overall. The results show that both students and experts felt that the task was low in physical demand and temporal demand, while it was a highly mentally demanding. There is not much difference between the students and experts in the rating results.

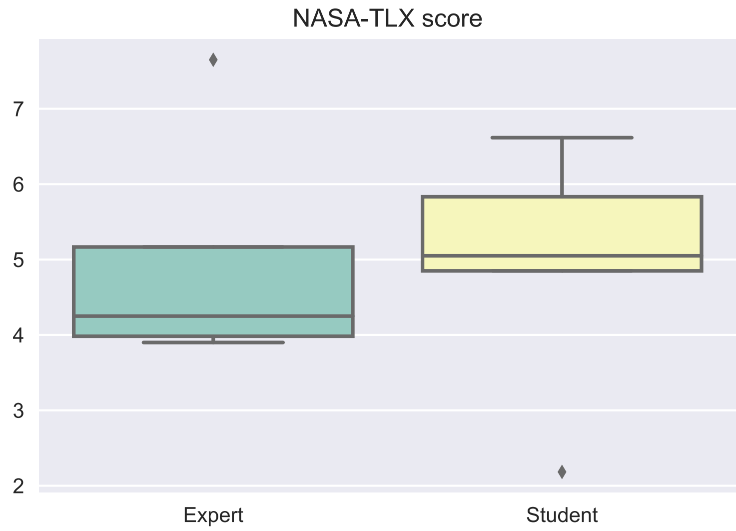


Figure 3. Comparison of the NASA-TLX rating raw score for students, experts.

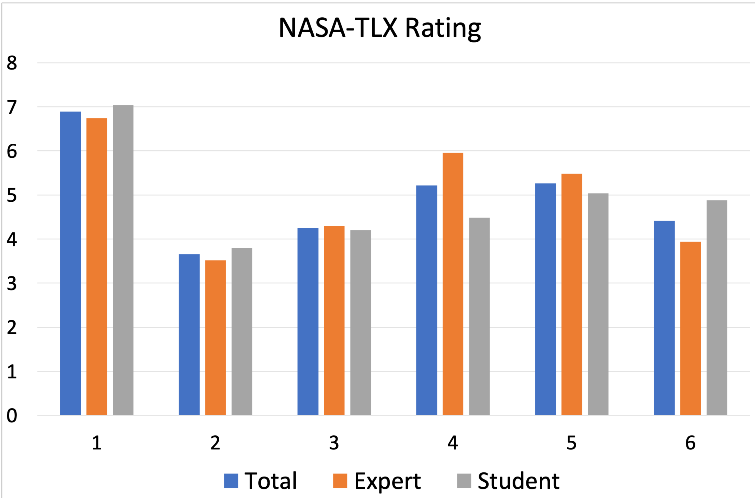


Figure 4. Comparison of the NASA-TLX rating results for students, experts and overall. The number from 1 to 6 presents Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level, respectively.

3.2. Data Pre-Processing

The dataset consists of four signal channels associated with EDA, body temperature (Temp.), BVP and HR. EDA is collected from electrodermal activity sensor measured in µS with frequency of 4 Hz. Body temperature data are measured on the Celsius (°C) scale with a frequency of 4 Hz [15]. BVP data are from photoplethysmograph (PPG) and sampled at 64 Hz [15]. HR data are the average heart rate values per second, derived directly from the BVP analysis [16]. All the signals were downsampled to 1 Hz for data analyses. The data are associated with ten participants, each participant has four sailing sections, the data can be split into forty samples (see Table 1). In order to compare the data with different resolutions, and to make it easier to analyse, we have normalized the downsampled data.

Table 1. The form of the data samples.

No. of Samples	No. of Participants	No. of Sailing Sections	Signal Channel 1	Signal Channel 2	Signal Channel 3	Signal Channel 4
Sample 1	Participant 1	Sailing section 1	EDA data	Temp. data	BVP data	HR data
Sample 2		Sailing section 2	..	..	..	..
. . .		Sailing section 3	..	..	..	..
		Sailing section 4	..	..	..	..
	. . .	. . .	. . .	. . .	. . .	. . .
	Participant 10	Sailing section 1	..	..	..	..
		Sailing section 2	..	..	..	..
Sample 39		Sailing section 3	..	..	..	..
Sample 40		Sailing section 4	EDA data	Temp. data	BVP data	HR data

Normalization of the data is done by calculating the standard normal distribution. The standard normal distribution is the simplest case of the normal distribution when the data are standardized to have a mean of zero and a standard deviation of one [17]. Before calculating the standard normal distribution, the standardized value of the signal data is computed from the mean and standard deviation using the following formula (see Equation (1)) [18]:

$$z_i = \frac{x_i - \bar{x}}{S}, \quad i = 1, 2, \ldots, n \tag{1}$$

where $z_i = \{z_1, z_2, \ldots, z_n\}$ is standardization value of the sample, $x_i = \{x_1, x_2, \ldots, x_n\}$ is the value of the downsampled signals, $\bar{x}$ is the mean of x_i on each file, S is the sample standard deviation, n is the number of the data on each file.

Finally, the standard normal distribution of the data was calculated using Equation (2) [19]:

$$f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}, \quad i = 1, 2, \ldots, n \tag{2}$$

where $z_i = \{z_1, z_2, \ldots, z_n\}$ is standardization value of the sample and n is the number of the data.

All 40 samples were labelled into two categories: expert and novice.

3.3. Classification Features Extraction

The normalized signal data were split into 40 samples as we mentioned in Section 3.2. Next, the features vectors (FVs) were collected from each sample. Since each sample of data has four signal channels, the total number of rows of the data is 160. The FVs include statistical-based feature vectors (SFV), wavelet-based feature vectors (WFV), and higher-order crossings (HOC)-based feature vectors.

3.3.1. Statistical-Based Features

For the statistical-based feature vectors (SFV), for each signal channel in each data sample, we calculated five types of vectors. These are mean vectors, the standard deviation vectors, variance vectors, median skewness vectors and kurtosis vectors. All SFVs were collected by the statistic feature vectors, i.e., (see Equation (3)):

$$SFV = [\mu_{SFV}, \sigma_{SFV}, Var, Sk_{SFV}, Kur_{SFV}] \tag{3}$$

where μ_{SFV} are the mean vectors, σ_{SFV} are the standard deviation vectors, Var are variance vectors, Sk_{SFV} are the median skewness vectors, Kur_{SFV} are kurtosis vectors.

1. Mean vectors

   Mean vectors are a collection of the mean of the normalized sample in each signal channel and defined as :

   $$\mu_{SFV} = [\bar{x}(i)_{EDA}, \bar{x}(i)_{temp}, \bar{x}(i)_{BVP}, \bar{x}(i)_{HR}]^T, \quad i = 1, 2, \ldots, N \tag{4}$$

   where $\bar{x}(i)_{EDA}$ is the mean of the EDA data for each sample, $\bar{x}(i)_{temp}$ is the mean of body temperature data for each sample, $\bar{x}(i)_{BVP}$ is the mean of the BVP data for each sample, $\bar{x}(i)_{HR}$ is the mean of heart rate data for each sample, and $N = 40$ is the number of the data samples.

2. Standard deviation vectors

   The standard deviation vectors σ_{SFV} are defined in the following (see Equation (5)):

   $$\sigma_{SFV} = [S(i)_{EDA}, S(i)_{temp}, S(i)_{BVP}, S(i)_{HR}]^T, \quad i = 1, 2, \ldots, N \tag{5}$$

   where $S(i)_{EDA}$ is the standard deviation of the EDA data for each sample, $S(i)_{temp}$ is the standard deviation of body temperature data for each sample, $S(i)_{BVP}$ is the standard deviation of the BVP data for each sample, $S(i)_{HR}$ is the standard deviation of heart rate data for each sample, $N = 40$ is the number of data samples.

3. Variance vectors

   Variance is the average of the squared differences from the mean. It is the square of the standard deviation. In a similar manner as above, variance vectors are calculated and represented as Var. It can be calculated using the following (see Equation (6)):

   $$Var = \sigma^2_{SFV} = [S(i)^2_{EDA}, S(i)^2_{temp}, S(i)^2_{BVP}, S(i)^2_{HR}]^T, \quad i = 1, 2, \ldots, N \tag{6}$$

   where σ_{SFV} is the standard deviation of each signal for each sample defined in Equation (5).

4. Median skewness vectors

Median skewness is also called Pearson's second skewness coefficient. It is defined as [20]):

$$Sk_2 = \frac{3(\mu - \nu)}{\sigma} \tag{7}$$

where Sk_2 is the Pearson's second skewness coefficient, μ is the mean of the data, ν is the median, and σ is the standard deviation of each signal in each file. This formula compares the mean to the median in a precise way and shows how many standard deviations apart they are [21]. The median skewness vectors Sk_{SFV} are the collection of median skewness of each signal channel for each sample:

$$Sk_{SFV} = [Sk_2(i)_{EDA}, Sk_2(i)_{temp}, Sk_2(i)_{BVP}, Sk_2(i)_{HR}]^T, \quad i = 1, 2, \ldots, N \tag{8}$$

where $Sk_2(i)_{EDA}$ is the median skewness of the EDA data for each sample, $Sk_2(i)_{temp}$ is the median skewness of body temperature data for each sample, $Sk_2(i)_{BVP}$ is the median skewness of the BVP data for each sample, $Sk_2(i)_{HR}$ is the median skewness of heart rate data for each sample, $N = 40$ is the number of samples.

5. Kurtosis vectors

Kurtosis is an important descriptive statistic of data distribution. It describes how much the tails of a distribution differ from the tails of a normal distribution. Is is defined as the fourth central moment divided by the square of the variance [22,23]:

$$Kurtosis(X) = E[(\frac{X - \mu}{\sigma})^4] = \frac{\mu_4}{\sigma^4} \tag{9}$$

where X is the dataset, $Kurtosis(X)$ is the kurtosis value of the normal distribution for the dataset, μ_4 is the fourth central moment, μ is the mean (defined in Equation (4)), and σ is the standard deviation of each signal in each sample defined in Equation (5).

The kurtosis vectors are the collection of kurtosis of each signal channel for each sample:

$$Kur_{SFV} = [Kur(i)_{EDA}, Kur(i)_{temp}, Kur(i)_{BVP}, Kur(i)_{HR}]^T \quad i = 1, 2, \ldots, N \tag{10}$$

where $Kur(i)_{EDA}$ is the kurtosis of the EDA data for each sample, $Kur(i)_{temp}$ is the kurtosis of body temperature data for each sample, $Kur(i)_{BVP}$ is the kurtosis of the BVP data for each sample, $Kur(i)_{HR}$ is the kurtosis of heart rate data for each sample, $N = 40$ is the number of the samples.

3.3.2. Wavelet-Based Features

Wavelet Transform is a powerful tool for analysing and classifying the time series signal data. Daubechies wavelets was selected because it is the most commonly used set of discrete wavelet transforms [24]. Among the extremal phase wavelet of Daubechies family, db4 wavelet was chosen in this study, where the number 4 refers to the number of vanishing moments [25].

In this study, the signal from each sensor collected during each sailing section was subjected to wavelet decomposition into N levels, and the result of the decomposition is divided into two parts as one set (of the N^{th} level) of approximation coefficients (cA) and N set (from 1 to N^{th} level) of and detail coefficient (cD). The cA represents low-frequency signal and the cDs represent high-frequency signal. The original signal usually can be decomposed to several levels, and each layer decomposition coefficients are obtained from the previous decomposition. In other words, the original signal S is decomposed into (see Equation (11)):

$$S = cD1 + cD2 + \ldots + cDN + cAN \tag{11}$$

where S is the dataset, $cD1, cD2, \ldots, cDN$ are high-frequency signal obtained by decomposition from the first layer, the second layer and the N layers respectively, cAN is the low-frequency signal obtained by decomposition of the N^{th} layer.

In the wavelet decomposition, the greater the gain, the more obvious the performance of the different characteristics of noise and signal is, and the more conducive to the separation of the noise and signal. On the other hand, the greater the number of decomposition levels, the greater the distortion of the reconstructed signal, which affects the final denoising effect to a certain extent. In this study, in order to handle the contradiction and choose an appropriate decomposition level, the highest six values of both cA and cD from the first level decomposition were selected. In addition, the mean, standard deviation, entropy of cD and cA are added into the feature vectors.

3.3.3. Higher-Order Crossings (HOC)-Based Features

The higher-order crossings (HOC) method is also often called zero-crossing and level-crossing method [26]. It counts the number of axis-crossing, i.e., the symbol changes in the dataset. In our dataset, we set the zero mean signal data from each sensor collected from each stop of each participant as a series $\{\vec{Z}\} = \{Z_t, t = 1, 2, \ldots, N\}$. The number of crossing of the horizontal axis, is denoted D_1, and it is the same as the number of sign changes in $\{\vec{Z}\}$ [26].

The higher-order crossings are defined by using the difference operator $\triangledown$, and $\triangledown Z_t$ is defined (see Equation (12) [27]):

$$\triangledown Z_t = Z_t - Z_{t-1} \tag{12}$$

For the second order of the difference, it is:

$$\begin{aligned} \triangledown^2 Z_t &= \triangledown(\triangledown Z_t) \\ &= \triangledown(Z_t - Z_{t-1}) \\ &= Z_t - 2Z_{t-1} + Z_{t-2} \end{aligned} \tag{13}$$

Higher orders can be computed in the same manner as above. In general, the k^{th} order difference is (see Equation (14) [28]):

$$\begin{aligned} \triangledown^{k-1} Z_t &= \sum_{i=1}^{k} C_{i-1}^{k-1} (-1)^{i-1} Z_{t+1-i} \\ with \quad C_{i-1}^{k-1} &= \frac{(k-1)!}{(i-1)!(k-i)!} \end{aligned} \tag{14}$$

where $k = 1, 2, \ldots$, and $\triangledown^0$ is the identity.

From $\{\vec{Z}\}$, a binary process, $\{\vec{X}\}$ is defined by (see Equation (15) [27–29]):

$$X_t = \begin{cases} 1, & Z_t \geq 0 \\ 0, & Z_t < 0 \end{cases} \tag{15}$$

Let

$$d_t = \begin{cases} 1, & X_t \neq X_{t-1} \\ 0, & otherwise, \end{cases} \tag{16}$$

where d_t is the indicator. It indicates that there is a symbol change in $\{\vec{X}\}$ when it is 1. The number of crossing of the horizontal axis, D_1, in $\{\vec{Z}\}$ is defined as [27,28]:

$$\begin{aligned} D_1 &= d_2 + d_3 + \ldots + d_N \\ &= \sum_{t=2}^{N} [X_t - X_{t-1}]^2 \end{aligned}$$

where N is the length of the data.

For the k^{th} order, the count of the symbol changes is:

$$\begin{aligned} D_k &= d_2^{(k)} + d_3^{(k)} + \ldots + d_N^{(k)} \\ &= \sum_{t=2}^{N} [X_t(k) - X_{t-1}(k)]^2 \end{aligned}$$

Above all, for the signal channel for each sample, the HOC-based feature vector, FV_{HOC}, is formed as follows (see Equation (17) [28]):

$$FV_{HOC} = [D_1, D_2, \ldots, D_L], \quad 1 < L \leq J \tag{17}$$

where J is the maximum order of the estimated HOC and L is the HOC order chosen in this study.

For our dataset, the FV_{HOC} was extracted from the four signals within a range of order $K = 1, \ldots, 50 (= J)$.

3.4. Deep Learning Model

In this study, deep learning algorithm was applied to classify the data. Convolutional Neural Network (CNN) was the approach employed for classification. Compared with traditional neural networks, the advantage of CNN was obvious, it has fewer parameters to learn for processing high-dimensional data, which helps to accelerate the training speed and reduce the chance of overfitting [30]. The steps of creating and training CNN are described below:

- First, load the dataset and separate the data into training and validation datasets. In this study, 80 percent of the data is used for training and 20 percent for testing. In order to protect against over fitting, cross-validation was applied. Cross-validation is 5-fold.
- Second, define the CNN architecture. For the first layer, the spatial input and output sizes of these convolutional layers are 32-by-32, and the following max pooling layer reduces this to 16-by-16. For the second layer, the spatial input and output sizes of these convolutional layers are 16-by-16, and the following max pooling layer reduces this to 8-by-8. For the next layer, the spatial input and output sizes of these convolutional layers are 8-by-8. The global average pooling layer averages over the 8-by-8 inputs, giving an output of size 1-by-1-by-4 times of initial number of filters. With a global average pooling layer, the final classification output is only sensitive to the total amount of each feature present in the input image, but insensitive to the spatial positions of the features. In the end, add the fully connected layer and the final softmax and classification layers.
- Third, specify the training options. We used Adam (adaptive moment estimation) optimizer, set the maximum number of epochs to 100, mini batch size to 128, and monitored the network accuracy during training by specifying validation data and validation frequency, shuffling the data every epoch, and plotting training progress [31].
- Fourth, train the network using the structure defined by layers, the training data, and the training options.
- Last, predict the labels of the validation data using the trained network, and calculate the final validation accuracy [31].

3.5. Optimization

The performance of CNN depends on an appropriate setting of hyper-parameters, including the batch size, learning rate, activation function, network structure, etc. [32]. Optimizing hyper-parameters yields better behaviour of the training algorithm, since hyper-parameters effect the performance of the training result for the model. Among the techniques of fine-tune machine-learning algorithms, automatic hyper-parameter tuning is an effective and computational power saving method compared to manual grid search. In this process, the next parameter settings is dependent on the performance of previous configurations. Configurations are inferred and decided by the relation between the hyper-parameter settings and model performance [33]. Bayesian optimization for hyper-parameter automatic tuning is one of the frequently used automating tuning hyper-parameters, and we applied it on this dataset for finding a good optimum.

Bayesian optimization approaches use the results of previous configuration performance to constitute a probabilistic model. The probability is the scores given by the hyper-parameters [34]. This model is used as a surrogate function for the objective function for choosing the best hyper-parameters [35]. The surrogate can be easily modeled by Gaussian Process, and a set of hyper-parameters are selected to give the best performance on the surrogate function. These hyper-parameters are applied on the actual objective function. The surrogate model is updated and the previous steps are repeated until it is optimized [35].

Choose Variables to Optimize

In this study, four variables were chosen to optimize using Bayesian optimization, and their search ranges were specified. The four variables are: network section depth, initial learning rate, stochastic gradient descent momentum and L_2 regularization strength. The following is the illustration of the variables:

- Network section depth: Network section depth is the variable which controls the depth of the network.
- Initial learning rate: Select the best initial learning rate.
- Stochastic gradient descent momentum: Momentum adds inertia so that the network can update the parameters more smoothly and reduce the noise inherent in stochastic gradient descent [36].
- L_2 regularization strength: Choose a good value of regularization to prevent overfitting issues.

Optimization variables with properties are as below (see Table 2):

Table 2. Optimizing variable with properties.

Name	Possible Values	Type	Transform	Optimize
Section Depth	[1, 3]	integer	none	1
Initial Learn Rate	[0.01, 1]	real	log	1
Momentum	[0.08, 0.98]	real	none	1
L_2 regularization	$[1.00 \times 10^{-10}, 0.01]$	real	log	1

3.6. Results

The following section presents the results including feature selection and data classification.

3.6.1. Dataset

The dataset consists of statistical based features (5 columns), wavelet-based features (12 columns) and HOC-based features (50 columns). The data collected from 10 participants with 4 sailing sections and 4 different signal channels (EDA, BVP, body temperatures and HR), i.e., the data comprises 160 rows and 67 columns in total.

3.6.2. Feature Selection

In this section, the results of the feature selection and the method are discussed.

From the original dataset, 73 features are extracted, out of which five are statistic-based features, eighteen are wavelet-based features, and fifty are HOC-based features. Among these features, seven different combinations can be made to give different accuracy results. After comparing the results by using the same learning algorithm from a different set of features or combinations, the best results from the combination of the features will be chosen. The learning algorithms random forest and support vector machine (SVM) were chosen for comparing the results.

Table 3. Results of accuracy by using the random forest and SVM (kernel = 'linear') for different amount of features selection.

Feature Type	Feature Amount	Random Forest	SVM
Statistic-based features	5	0.51 (+/− 0.25)	0.54 (+/− 0.20)
Wavelet-based features	18	0.46 (+/− 0.24)	0.49 (+/− 0.17)
HOC-based features	50	0.64 (+/− 0.30)	0.58 (+/− 0.09)
Statistic-based + Wavelet-based features	23	0.54 (+/− 0.17)	0.53 (+/− 0.17)
Wavelet-based + HOC-based features	68	0.57 (+/− 0.39)	0.59 (+/− 0.06)
Statistic-based + HOC-based features	55	0.61 (+/− 0.40)	0.59 (+/− 0.21)
Statistic-based + Wavelet-based + HOC-based features	73	0.62 (+/− 0.21)	0.61 (+/− 0.09)

The accuracy of different combinations of feature selection by using random forest and SVM is listed on Table 3. It shows that HOC-based features give the highest accuracy by using the random forest algorithm, and combining all the 73 features gives the highest accuracy by using SVM algorithm. When using all of the 73 features, the accuracy by using random forest and SVM are very close to each other and the standard deviation of the accuracy is lowest among all the results. Therefore, selecting all the features to do the analyse will give a stable results. We propose a hybrid CNN approach for the classification task, where a combination of extracted features like statistical, wavelet and HOC features will be used instead of raw biosignals for CNN-based classification of stress level of user during maritime navigation.

3.6.3. Results from the Data Classification

The final result for the training accuracy is calculated by applying deep learning using Bayesian optimization on our optimal model. The results are given in Table 4. For better understanding, the histogram of results is shown in Figure 5 as below. The result shows that by selecting all the features, we got the highest accuracy which is 75.5%. The approximate 95% confidence interval (written as "testError95CI") of the generalization error rate are also given in the table. "testError95CI" is the interval resulting from Equation (18):

$$testError - 1.96 \cdot testErrorSE \leqslant testError95CI \leqslant testError + 1.96 \cdot testErrorSE \quad (18)$$

where $testErrorSE$ is the standard error.

The standard error is calculated in Equation (19):

$$testErrorSE = \sqrt{\frac{testError \cdot (1 - testError)}{NTest}} \quad (19)$$

where $Ntest$ is the number of elements in the array of label of testing data.

Table 4. Bayesian Optimization Results for CNN from different feature selections.

Feature Types	Validation Error	testError	testError95CI		Accuracy
Statistic-base + wavelet-based features (method 1)	0.357	0.357	0.193	0.522	64.3%
HOC-base features (method 2)	0.263	0.263	0.119	0.407	73.7%
Statistic-based + Wavelet-based + HOC-based features (method 3)	0.245	0.245	0.101	0.388	75.5%

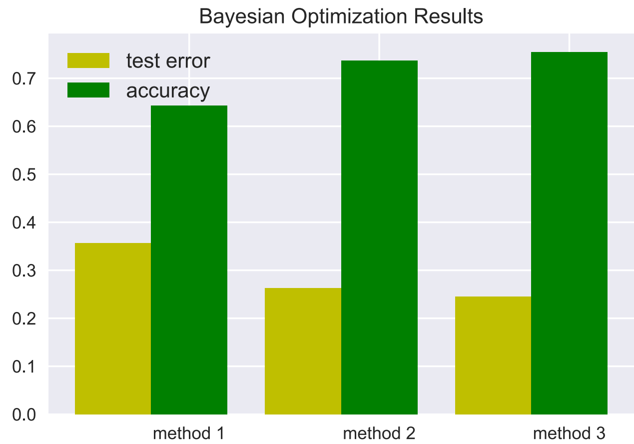


Figure 5. Histogram of Bayesian optimization results for CNN from different feature selections. In this figure, method 1 represents the selection of statistic-based feature and wavelet-based features, method 2 represents the selection of HOC-based features, and method 3 represents the selection of statistic-based feature, wavelet-based features and HOC-based features.

By increasing the number of features by applying CNN Bayesian Optimization, we obtained the highest accuracy. When adding the 50 HOC-based features, the prediction accuracy increased significantly. When using Wavelet-based features, the accuracy was poorer than using other features. Therefore, wavelet-based features do not affect much the results when combined with other features. However, it still gives the best results when adding all of the 73 features. The result presents that feature selection improves the classification accuracy. Nonetheless, we do not claim that it can apply to all types of data, since classification accuracy achieved with different feature reduction strategies is highly sensitive to the type of data [37].

4. Discussion

This study conducted an experiment for finding out the stress differences between experienced experts and students in SA in maritime navigation. The result of data analysis shows that the biosignal data from experts and students can be classified by a certain machine learning algorithm. The result of subjective measurement of workload shows that there is a difference between experts and students. The summary of findings is shown in Table 5.

Table 5. Summary of findings.

Hypothesis Number	Description	Accepted/Rejected
H1	The bio-sensor data from experienced experts can be distinguished from the biosensor data from students	Accepted
H2	The stress levels obtained from the bio-sensor data show a correlation to the NASA-TLX rating results	Accepted
H3	Compared to the novices, the experts feel less stress during a navigation task	Accepted

Based on the results associated with the three hypotheses, we outline the following statements:

Statement 1. Biosignal data are considered as stress monitoring analysis, since stress can be a physical, mental or emotional reaction, and it causes hormonal, respiratory, cardiovascular and nervous system changes. For example, stress can make your heart beat faster, make you breathe rapidly, sweat and tense up. Based on the results of the classification of the biosensor data, we can see that the data from the experts and the students have different patterns. Accuracy of 75.5% is an acceptable result for distinguishing the data. This could have implication for stress difference with maritime navigating skills. Previous research shows that experts obtain better results in SA task [38]. When facing the same task, with different level of skills, the stress level is different.

Statement 2. NASA-TLX contains subjective data which was evaluated by the participants themselves. Results from the NASA-TLX show that experts and students had different evaluation of the workload, which is consistent with the results from classification of biosignal data, i.e., biosignal data show a different pattern for experts and students. This result is also consistent with other research suggesting that mental effort and anxiety are closely related to HRV [39].

Statement 3. The results of NASA-TLX rating show that experts have a smaller workload compared to students. Many research articles show that there is a high correlation between workload and stress. When there is the overload of the work, the stress is increasing [40,41].Therefore, there have implication that experts feel less stress than students. In addition, some other research also shows that higher levels of stress had a negative relationship with work SA [42]. Since experts have better SA score [38], they should be under less stress.

5. Conclusions

In this paper, we propose a deep learning approach using Bayesian optimization for classifying the biosignal data of navigators during the maritime operation. We extracted different types of the features to improve the prediction accuracy. We also compared the objective results to the subjective results, NASA-TLX rating results, that the two results are correlated.

We would like to highlight that number of samples were few in order to make any statistical claims of our findings. Based on our current study in the next step, we plan to experiment further with a larger set of population. Nevertheless, the results of our current data analysis as well as this study will contribute to auto-assessment system for evaluating the SA performance in maritime navigation.

Author Contributions: Conceptualization, H.X., P.S. and B.-M.B.; methodology, H.X., B.-M.B. and D.K.P.; software, H.X. and D.K.P.; validation, H.X., B.-M.B. and D.K.P.; formal analysis, investigation, resources and data curation, H.X., B.-M.B. and D.K.P.; writing—original draft preparation, H.X.; writing—review and editing, H.X., B.-M.B., J.A.J., P.S. and D.K.P.; supervision, B.-M.B. and D.K.P. All authors have read and agreed to the published version of the manuscript.

Funding: There is no funding reported for the research. The resources and publication supports are provided by UiT The Arctic University of Norway.

Institutional Review Board Statement: All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted by the Department of Technology and Safety, UiT The Arctic University of Norway, and the project was approved by NSD Norsk Senter For Forskningsdata (Norwegian center for research data).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The transformed feature data from the raw data presented in this study are available on request from the corresponding author. The raw data are the property of Department of Technology and Safety, UiT The Arctic University of Norway, and not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BVP	Blood volume pulse
CNN	Convolutional Neural Network
EDA	Electrodermal activity
FV	Features vector
HOC	Higher order crossing
HR	Heart rate
NASA-TLX	NASA Task Load Index
SA	Situation awareness
SAGAT	Situation Awareness Global Assessment Technique
SFV	Statistical-base feature vector
SVM	Support vector machine
WFV	Wavelet-based feature vector

References

1. Sanfilippo, F. A multi-sensor fusion framework for improving situational awareness in demanding maritime training. *Reliab. Eng. Syst. Saf.* **2017**, *161*, 12–24. [CrossRef]
2. Marchenko, N.; Andreassen, N.; Borch, O.J.; Kuznetsova, S.; Ingimundarson, V.; Jakobsen, U. Arctic shipping and risks: Emergency categories and response capacities. *Transnav Int. J. Mar. Navig. Saf. Sea Transp.* **2018**, *12.1*. [CrossRef]
3. Grech, M.R.; Horberry, T.; Smith, A. Human error in maritime operations: Analyses of accident reports using the Leximancer tool. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; Sage Publications Sage CA: Los Angeles, CA, USA, 2002; Volume 46, pp. 1718–1721.
4. TheNauticalInstitute. Situational Awareness-The Sense of Everything, Feb 2020, Issue no. 23. Available online: https://www.nautinst.org/uploads/assets/8f9da438-3486-4b77-a4d79e1af61fe15c/Issue-23-Situational-Awareness.pdf (accessed on 21 August 2020).
5. Endsley, M.R. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society Annual Meeting*; SAGE Publications Sage CA: Los Angeles, CA, USA, 1988; volume 32, pp. 97–101.
6. Endsley, M.R.; Bolte, B.; Jones, D. G. *Designing for Situation Awareness: An Approach to User-Centered Design*; CRC press: Boca Raton, FL, USA, 2003; Chapter 2.
7. Van den Broek, A.; Neef, R.; Hanckmann, P.; van Gosliga, S.P.; Van Halsema, D. Improving maritime situational awareness by fusing sensor information and intelligence. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; pp. 1–8.
8. © 2019 Empatica Inc. E4 Wristband Real-Time Physiological Data Streaming and Visualization. 2020. Available online: https://www.empatica.com/en-int/research/e4/ (accessed on 30 November 2020).
9. Sharek, D. A useable, online NASA-TLX tool. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications Sage CA: Los Angeles, CA, USA, 2011; Volume 55, pp. 1375–1379.
10. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
11. Rubio, S.; Díaz, E.; Martín, J.; Puente, J.M. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Appl. Psychol.* **2004**, *53*, 61–86. [CrossRef]
12. Bustamante, E.A.; Spain, R.D. Measurement invariance of the Nasa TLX. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications Sage CA: Los Angeles, CA, USA, 2008; Volume 52, pp. 1522–1526.

13. Hart, S.G. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; Sage publications Sage CA: Los Angeles, CA, USA, 2006; Volume 50, pp. 904–908.
14. Colligan, L.; Potts, H.W.; Finn, C.T.; Sinkin, R.A. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *Int. J. Med. Inf.* **2015**, *84*, 469–476. [CrossRef] [PubMed]
15. © 2019 Empatica Inc. Data Export and Formatting from E4 Connect, January 2020. Available online: https://support.empatica.com/hc/en-us/articles/201608896-Data-export-and-formatting-from-E4-connect- (accessed on 12 May 2020).
16. © 2019 Empatica Inc. E4 Data-HR.csv Explanation, January 2020. Available online: https://support.empatica.com/hc/en-us/articles/360029469772-E4-data-HR-csv-explanation (accessed on 12 May 2020).
17. Patel, J.K.; Read, C.B. *Handbook of the Normal Distribution*; Dekker, M., Ed.; International Biometric Society: New York, NY, USA, 1982.
18. Bohm, G.; Zech, G. *Introduction to Statistics and Data Analysis for Physicists*; Desy: Hamburg, Germany, 2010; Volume 1, p.23.
19. Lane, D.M.; Scott, D.; Hebl, M.; Guerra, R.; Osherson, D.; Zimmer, H. *An Introduction to Statistics*; Rice University: Houston, TX, USA, 2017; pp. 261–268.
20. Weisstein, E.W. "Pearson's Skewness Coefficients." From MathWorld–A Wolfram Web Resource, 23 July 2020. Available online: https://mathworld.wolfram.com/PearsonsSkewnessCoefficients.html(accessed on 28 July 2020).
21. Doane, D.P.; Seward, L.E. Measuring skewness: A forgotten statistic? *J. Stat. Educ.* **2011**, *19*. [CrossRef]
22. Kokoska, S.; Zwillinger, D. *CRC Standard Probability and Statistics Tables and Formulae*; CRC Press: Boca Raton, FL, USA, 2000.
23. Pearson, K. IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* **1916**, *216*, 429–457.
24. Akansu, A.N.; Haddad, P.A.; Haddad, R.A.; Haddad, P.R. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*; Academic press: Cambridge, MA, USA, 2001.
25. Daubechies, I. CBMS-NSF regional conference series in applied mathematics. *Ten Lect. Wavelets* **1992**, *61*. [CrossRef]
26. Dickstein, P.; Spelt, J.; Sinclair, A. Application of a higher order crossing feature to non-destructive evaluation: A sample demonstration of sensitivity to the condition of adhesive joints. *Ultrasonics* **1991**, *29*, 355–365. [CrossRef]
27. Kedem, B. Higher-order crossings in time series model identification. *Technometrics* **1987**, *29*, 193–204. [CrossRef]
28. Petrantonakis, P.C.; Hadjileontiadis, L.J. Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *14*, 186–197. [CrossRef] [PubMed]
29. Kedem, B.; Yakowitz, S. *Time Series Analysis by Higher Order Crossings*; IEEE Press: New York, NY, USA, 1994; p. 19.
30. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.
31. © 1994–2020 The MathWorks, Inc. Create Simple Deep Learning Network for Classification, 2020. Available online: https://se.mathworks.com/help/deeplearning/ug/create-simple-deep-learning-network-for-classification.html (accessed on 25 November 2020).
32. Sameen, M.I.; Pradhan, B.; Lee, S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena* **2020**, *186*, 104249. [CrossRef]
33. Pham, V. Bayesian Optimization for Hyperparameter Tuning, 2016. Available online: https://arimo.com/data-science/2016/bayesian-optimization-hyperparameter-tuning/ (accessed on 5 May 2020).
34. Dewancker, I.; McCourt, M.; Clark, S. Bayesian optimization for machine learning: A practical guidebook. *arXiv* **2016**, arXiv:1612.04858.
35. Koehrsen, W. A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning, 2018. Available online: https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f (accessed on 5 May 2020).
36. © 1994–2020 The MathWorks, Inc. Deep Learning Using Bayesian Optimization, 2020. Available online: https://se.mathworks.com/help/deeplearning/ug/deep-learning-using-bayesian-optimization.html (accessed on 8 May 2020).
37. Janecek, A.; Gansterer, W.; Demel, M.; Ecker, G. On the relationship between feature selection and classification accuracy. In *New Challenges for Feature Selection in Data Mining and Knowledge Discovery*; PMLR: McKees Rocks, PA, USA, 2008; pp. 90–105.
38. Xue, H.; Batalden, B.M.; Røds, J.F. Development of a SAGAT Query and Simulator Experiment to Measure Situation Awareness in Maritime Navigation. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, New York, NY, USA, 24–28 July 2020; Springer: Cham, Switzerland, 2020; pp. 468–474.
39. Thayer, J.F.; Friedman, B.H.; Borkovec, T.D. Autonomic characteristics of generalized anxiety disorder and worry. *Biol. Psychiatry* **1996**, *39*, 255–266. [CrossRef]
40. Lundberg, U.; Hellström, B. Workload and morning salivary cortisol in women. *Work. Stress* **2002**, *16*, 356–363. [CrossRef]
41. Steptoe, A.; Cropley, M.; Griffith, J.; Kirschbaum, C. Job strain and anger expression predict early morning elevations in salivary cortisol. *Psychosom. Med.* **2000**, *62*, 286–292. [CrossRef] [PubMed]
42. Sneddon, A.; Mearns, K.; Flin, R. Stress, fatigue, situation awareness and safety in offshore drilling crews. *Saf. Sci.* **2013**, *56*, 80–88. [CrossRef]

Paper IV

A study on the effects of rapid training method on ship handling, navigation and decision-making skills under stressful situations

H. Xue, J.-F. Røds, Ø. Haugseggen, A. J. Christensen, B.-M. Batalden, O. T. Gudmestad

Submitted to *The Journal of Navigation*

A Study on the Effects of Rapid Training Method on Ship Handling, Navigation and Decision-Making Skills under Stressful Situations

Hui Xue[1,*], Johan-Fredrik Røds[1], Øyvind Haugseggen[1], Anders Johan Christensen[1], Bjørn-Morten Batalden[1], Ove T.Gudmestad[1,2]
[1]Department of Technology and Safety, UiT The Arctic University of Norway, P.O. Box 6050 Langnes, Tromsø, 9037, Norway
[2]Department of Mechanical and Structural Engineering and Material Science, University of Stavanger, Stavanger, 4036, Norway

*Corresponding author. E-mail: hui.xue@uit.no

Abstract

Navigational safety is one of the important focuses of Maritime Education and Training (MET), and the quality of MET is the key to cultivating competent officers at sea. The objective of this study is to better understand the effects of a rapid training method on ship handling and navigation in restricted waters, as well as decision-making skills under stressful situations. Tests were carried out in a simulator-based maritime training environment to explore the decision-making skills of maritime students in stressful situations under different training levels and methods. This study compares routine maritime training and task-aimed rapid training in improving manoeuvring, navigational and decision-making skills, and examines the training outcomes. The data used in this study is based on a comparison of the task performance and stress levels of the two groups of students, using simulator-based training results from a designed scenario. Through the results, the training outcomes of decision-making skills and maritime operation performance are analysed by applying a specific decision-making model. In addition, the impact of students' stress levels was analysed, both subjectively and objectively. The article concludes with a set of recommendations for the design of future MET. The research is helpful for enhancing decision-making skills in maritime training programmes and for understanding how learning in simulator-based maritime training environments can be improved.

Keywords: navigation, simulator-based maritime training, decision-making, rapid training

1. Introduction

On the basis of Safety System Project theory, the navigation system is basically a 'ship-human-environment' (Inoue, 2000; Xiufeng et al., 2005). Hence, approximately 85% of maritime accidents are accounted for by navigation accidents (i.e., collision and grounding) (Jaeyong, et al., 2016), caused by human errors such as mistakes in ship handling and impropriate decision-making (Wróbel et al. 2017; Wu et al., 2020). In addition, studies found that incompetent officers have frequently contributed to ship accidents. Therefore, navigational safety is one of the important focuses of Maritime Education and Training (MET), and the quality of MET has drawn more attention from both academics and employers in the shipping industry (Bao et al., 2021). Good quality MET is critical for seafarers to acquire knowledge and skills to manage risks, solve problems and complete operations safely and efficiently, and thus ensures the safety of life at sea (Basak, 2017). However, there are challenges involved in MET, in that it requires high investment and incurs high running costs (Markopoulos et al., 2019), as well as being time-consuming. An important challenge facing MET is how to train a certain skill in an efficient and cost-effective way in a safe environment.

It is well known that the tradition of using simulator-based maritime training has a long history in MET (Kim et al., 2021). Simulators provide a non-risk environment for trainees to practise what they have learned from the classroom. Simulator-based training is an effective way to simulate scenarios that may occur on board, for trainees to obtain practical experience and hopefully be capable of handling unexpected scenarios in future offshore work. Throughout the specific simulation exercises, technical, procedural and operational skills are acquired by trainees and therefore their capabilities can be improved. After training, trainees can have a better understanding of the required decision-making process and how different actions will affect a situation (Markopoulos et al., 2019); thus, they learn how to prioritize in challenging traffic and emergency operations and situations.

In navigation tasks, decision-making skills are the key to safe sailing (Nooros and Hukki, 2003). For instance, in the task of collision avoidance, the navigator must decide which means are appropriate (radar, visual means, Automatic Identification System, etc.) in the situation. Afterwards, he/she must determine (decide) whether the risk of collision exists and which action should be taken (Allen, 2004. P.217). Environmental stress is also one of the dominant factors that cause accidents at sea (Sampson and Thomas, 2003; Hetherington et al., 2006; Gug et al., 2022). Working at sea is inherently stressful (Hystad and Eid, 2016; Carotenuto et al., 2013; Jensen and Oldenburg, 2021), especially when the situation changes and many decisions must be made under pressure (Størkersen et al., 2018). In addition, many decision-making situations themselves can trigger stress responses. Therefore, stress can affect decision-making under varying degrees of uncertainty, while changing the underlying decision-making mechanism (Starcke and Brand, 2012). As a result, high stress levels undoubtedly cause faults in decision-making, and this can be dangerous at sea.

As we know, training in decision-making skills is challenging, due, for instance, to ill-structured environments (Klein, 1997), and the maritime field can be described as such an environment. Therefore, this leads to maritime education and training (MET) being a challenging field of education in general. The environment at sea can change rapidly, and decision-makers do not always find themselves in a familiar and predictable situation. Therefore, developing skills in decision-making is an essential subject in MET. The objective of the present study is to better understand the effect of training in decision-making skills in a simulator-based MET environment and to explore the decision-making skills of maritime students in stressful situations under different training levels and methods.

MET is costly (Sampson, 2004); hence, it is important to ensure it is effective and efficient. In addition to the expensive training equipment (simulators), devices and laboratories for maritime practice, teaching costs, such as wages, academic staff costs, administration and support staff costs, account for a large part of the budget (Cicek and Er, 2008). Therefore, it is necessary to continuously update and upgrade the contents of MET education (Čampara et al., 2017). For instance, reducing the training time without compromising training effectiveness or learning more skills during the same MET period are effective ways to reduce the MET costs. From Skill Acquisition Theory (DeKeyser, 2020), we know that learning ship handling and navigation skills involves the learning of habits and skills. This kind of learning is always slow because it requires practice and overcoming mistakes from practising. However, it is possible to speed up this kind of learning if we provide a psychologically safe coaching environment. In the process of learning skills, when mistakes occur, the learner will feel uncomfortable with the temporary incompetence. If learners do not make the same mistake again, they are rewarded. Once learners get consistent rewards for correct responses, the learning progress speeds up and the learning is reliable (Schein and Sloan School of Management, 1992). Therefore, it is possible to have rapid training methods for trainees to acquire skills. However, these are not widely used in maritime training.

Based on the aforementioned literature, it is necessary to perform studies of the training process. This study aims to discover the impact of rapid training methods in MET on ship handling, navigation and decision-making skills' development in a stressful situation. In addition, the learning outcomes from different teaching methods, especially in decision-making competence and communication skills in maritime operations, need to be compared. The innovation of the paper is, firstly, that we explain how the decision-making model has been chosen in different situations in the maritime domain, based on different decision-making models. Secondly, by the use of simulators in MET, a case study is presented to document the development of decision-making skills. In this context, the workload and stress levels were assessed by studying the interaction between the subjective workload, stress level and quality of the decision-making.

This study attempts to formulate answers to the two questions: (1) Can project-aimed rapid training give enough knowledge for participants to make efficient decisions in stressful and critical situations? and (2) Can the training method affect the decision-making model applied by the participants?

The paper is presented as follows: Firstly, a brief introduction of the theoretical basis, including the decision-making model and simulator-based MET, is presented. In Section 3, the designed training scenario is presented, and a customer decision quality rating scale is proposed for evaluating the impact of the MET programme on decision-making. Section 4 presents the results of the training scenario, including the workload assessment, the stress level and the quality of the decision-making. In Section 5, the results are discussed. The final section presents conclusions from the study and future work.

2. Theoretical basis

Decision-making plays an important role in maritime operations (Allen, 2004. P.217) and constitutes the foundation of the present study. At every stage in maritime activities, seafarers make decisions by accessing information, understanding the situation and assessing risks, to make sure that situations are safe, and that activities are performed effectively. Here, we will use the term "risk" to represent a combination of the probability of an unwanted incident and the consequences of the incident. These decisions are not only critical for the continued safety of the ship but also have major implications for the environment and the economy. Therefore, the training in decision-making skills is essential in MET.

Decision models can be used to describe how people decide in realistic settings. There is no unified decision theory, but researchers have proposed different models in different settings (Klein et al., 1993. P.103). One such model is naturalistic decision-making (NDM) (Klein et al., 1993. P.9). During maritime activities at sea in a high-risk environment, some of the decisions must be made under time pressure, where the information is insufficient, and the goals are not well defined. In addition, the decision-makers never have access to all the information. Therefore, it is impossible for them to assess all the possible options and the consequences. Hence, making rational and optimal decisions under time constraints is difficult. NDM does not require the decision-makers to possess the rationality, knowledge and the information-processing capacity to make the decisions. NDM is suitable for situations under limited time, changing context and unstable conditions involving persons with different levels of experience.

Under the NDM framework, the decisions do not need to be the best possible option; the solution should be satisfactory but not the ultimate best. In MET, most of the time, students are given classical decision-making training such as situation awareness training for collision avoidance. In the real world, however, naturalistic decision-making (NDM) is more common than classical decision-making. Unlike the classical decision-making approaches, NDM training is more challenging because it does not rely on decision theory or other formal models; it is based on intuition. In order to make acceptable decisions, a large number of experiences formed in patterns are needed to get different forms of tacit knowledge (Cohen et al., 1998). The process of obtaining these experiences can be speeded up by certain training programmes.

From the NDM framework, the main protocol is the Recognition Primed Decision-making (RPD) mode. This mode describes how people use their experience to make quick and effective decisions in complex situations. It relies on the decision-maker's mental simulation, that is, the decision-maker examines their memory for the situational cues that match previous events as patterns. The RPD mode shows how to implement decisions from four aspects of recognition (plausible goals, relevant cues, expectancies and a series of actions) to generate a plausible course of action (COA) and use mental simulation to evaluate the COA, when people are experiencing a challenging situation (Klein, 1993) (Fig. 1). For the RPD mode, it is most important for the decision-makers to identify a reasonable COA as rapidly as possible. The quality of the decisions is highly dependent on the knowledge, experience and training of the decision-makers.

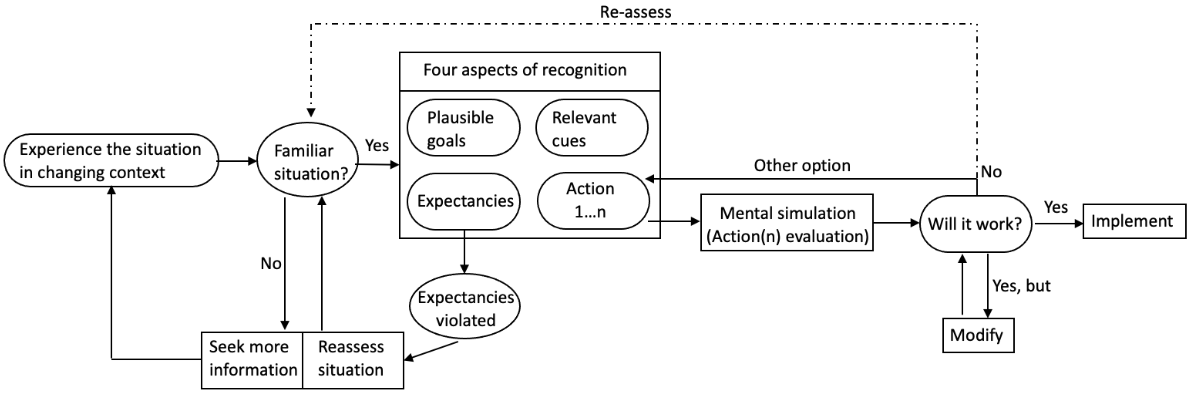


Fig. 1 RPD mode in a complex situation is re-illustrated, based on Klein (1993).

3. Materials and methods

3.1 Participants

A total of 22 (mean age = 22.4 years, standard deviation = 2.04, 5 females and 17 males) undergraduate students in nautical science at UiT The Arctic University of Norway (UiT) voluntarily participated in a simulator based experiment. Some of them have practical experience at sea from part-time jobs. The assessment of participants' skill levels can be found in Appendix A. According to the participants, they were mentally and physically healthy at the time of the experiment. The participants numbered fourteen from the first-grade course and eight from the second-grade course. Students from each grade were randomly divided into groups of two and two persons for handling two tugs of a towing operation. The reason for choosing a towing operation was because it not only is a basic operation in the maritime domain but also provides the environmental conditions that determine the level of ship-handling difficulty, which is an important factor affecting the likelihood of accidents. Beforehand, every participant signed a form consenting to their participation in the experiment.

3.2 Materials and apparatus

The experiment was conducted on two simulator bridges, both with 240° view and equipped with the K-sim Navigation software from Kongsberg Digital. A common instructor station was assigned to both simulator bridges for acting the multiple roles, such as machine, deck, the crew on the towing object, etc. Two types of vessel models were used in the experiment. The vessel model of the towing object was a small bulk carrier, Hagland Saga (HS), with a length between perpendiculars of 90 metres. The vessel model used in the experiment was two tugs with 2 pitch propellers with rudder named SMIT Panama. The three vessels were connected by a line of 200 metres in length.

A medical-grade wearable device, Empatica E4 Wristband, was introduced for collecting the biosignal data relevant to the experiment.

3.3 Experiment design

In this study, the experiment scenario was based on a towing operation as a within-subject factorial. Based on the different training methods, the second-grade students were assigned to the control group, and the first-grade students were assigned to the experiment group. The experiment can be considered a quasi-experiment, as the students were not randomly assigned to the experiment and control groups. Although the absence of random assignment casts some doubt on internal validity, the results of such studies are still compelling because they are not artificial interventions in social life and because their ecological validity appears strong [Bryman, 2012. P. 50]. In each group, every two participants assigned as a team accomplished the towing operation under the designed scenario. The participants were randomly assigned to the two tugboats. The experiment also included several dependent variables for analysing the results: training methods, cognitive workload, stress level, and decision-making.

3.3.1 Scenario

The participants were asked to tow an object (a small bulk carrier, Hagland Saga) near Ryøya island area (south of Tromsø) towards Tromsø (an Arctic city in Norway). Each participant sailed a tug. Tug Bravo was in front of the object to lead the way and tug Charlie was at the back of the object to secure the object. Participants could communicate with each other via maritime VHF (very high frequency) radiocommunication; they could also communicate with the instructor station by using UHF (ultra-high frequency) radiocommunication, so that the other tug would not hear their conversation.

Good weather was chosen for the scenario. The weather condition can be found in Table 1. During the towing operation, failure of both engines would be induced when the tugs were in a critical location where they were going to pass Ryøya island, located south of Kvaløya, southwest of Tromsø. Geographical locations are presented in Table 2, and the 3D view in two different directions of vision can be found in Fig. A1 in Appendix B. With the suggested sailing speed, which is 6 knots, tug Bravo would receive the first failure information from the machine department around 20 minutes after they started. After an approximately 30 seconds to 1 minute time gap, tug Bravo would receive the second failure information, and it would lose all the power to continue sailing.

Table 1 Weather condition applied in the scenario.

	Speed	**Direction**
Wind	6 knots	From 040°
Current	0.5 knots	270° (going to west)

Table 2 Geographical location of the start point, and failure induce location (based on HS).

	Start point	**First failure induced**	**Second failure induced**
Latitude	69°32.948'N	Depends on the current location	Depends on the current location
Longitude	18°38.026'E	18°43E	18°43.5E

3.3.2 Training methods

In this study, the training programme was conducted on the bridge simulators. The reasons for using simulator-based training and the description of the training programme are presented below.

- Why use simulator training?

The use of simulators is a key part of modern maritime training and education. The training of maritime students is based on the regulations found in The International Convention on Standards of Training, Certification, and Watchkeeping for Seafarers (IMO 2018). This regulatory framework (STCW-code, tables A-2/1 and A-2/2) describes the minimum requirements for the students, regarding competence, knowledge, understanding and proficiency. In the tables mentioned above, methods for demonstrating competence are described. For several of the modules described in the code, the available methods for demonstrating competence are by the use of either a vessel or simulators. For practical and economic reasons, the use of simulators is the preferred method in most cases.

Although the training and education programmes might differ in organization and content (Nazir et al., 2019), the use of simulators is still a key part of the training. There are also studies that describe how the use of simulators in the training process can reduce the risk of maritime accidents (Hanzu-Pazara et al., 2008). Maritime simulators can be used for training on a wide spectrum of situations, for instance training on complex tasks (Hjelmervik et al., 2018). It has also been used for training on towing operations as specific tasks (Gudmestad et al., 1995).

- Description of the training programme

The students participating in the experiment are all part of the bachelor programme in nautical science at UiT. During the three-year programme, the students will have, in total, 32 simulator exercises with an instructor present or 96 hours of simulator training for each student. The simulator is also available for self-study exercises, and most of the students will have achieved a significantly higher number of hours in the simulator by the end of their studies. The students participating in the experiment were in the first or second year of their studies. The first-year students had fulfilled half of the simulator exercises (16 exercises or 48 hours, including two examinations with external evaluation) and the second-year students had fulfilled all the simulator exercises (32 exercises or 96 hours, including two examinations with external evaluation).

After the first-year course study, students are able to critically analyse and communicate the interplay between regulations, technology and human factors and their significance for the safety of life, the environment and property at sea. Through the simulator exercise, students practise position determination by terrestrial navigation, coastal navigation and blind navigation. They should have knowledge of using electronic systems for position determination and navigation. They can use and interpret information from on-board meteorological instruments, radar and automatic radar plotting aid (ARPA) and use this information to make decisions for the safety of sailing. They can carry out a safe bridge guard by being able to demonstrate the ability to handle resources, communication, leadership and situational awareness.

After the second-year course study, students can carry out an independent analysis and communicate how navigation procedures and technical equipment in combination affect maritime safety, as well as further identify weaknesses and limitations in the system and find solutions to these. Students have the ability to decide, implement and communicate an optimal use of technical and human resources on board in order to plan and carry out safe, efficient and environmentally friendly maritime transport. Through the set of simulation exercises, students are trained to handle emergency situations, such as no Global Positioning System (GPS) availability and loss of radar. The students must choose the route themselves, learn to deal with other traffic in the trail, deal with the other simulator ships, in addition to other traffic.

Anchoring and towing operations are learned and practised in several simulation training exercises during the second semester of the second-year study. Situation awareness and decision-making skills and communication skills are improved during their second-year studies.

For the first-year students, an extra towing operation training course (project-specific rapid content training) was conducted in a rapid way for them to obtain the skill to fulfill the task. The content of towing operation training was the same as that for the second-year students. Key points, such as towing theory, methods and dealing with emergencies, were covered in a 20-minute video lecture. One-hour hands-on training and practising on the simulator were carried out before the experiment was conducted. After this rapid training, the first-year participants gained the ability to complete the towing operation.

3.3.3 Workload assessment

In this study, a reliable assessment tool, NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988; Sharek, 2011), was employed to access the workload. Six categories, including Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration Level, were rated by the participants after the experiment. For each category, the rating was transferred to a ten-point Likert-type scale from low to high levels, where 0 is low and 10 is high.

3.3.4 Stress level assessment

The stress level was assessed in both subjective and objective ways. The State-Trait Anxiety Inventory (STAI) Form Y-1 (Spielberger, 1983) was used to access the self-assessment of the stress level. Each participant filled in the STAI Y-1 form immediately after leaving the simulator bridge. The STAI Y-1 form has a brief self-rating scale for the assessment of state and trait anxiety. It consists of 20 questions that evaluate the participant's present feeling. STAI scores can be up to 80 and are commonly classified on three levels as "no or low anxiety" (20-37), "moderate anxiety" (38-44), and "high anxiety" (45-80) (Fountoulakis et al., 2006).

The objective stress level can be reflected by changes in the heart rate (HR). The HR increases when people are overwhelmed by stress (Vrijkotte et al., 2000). The reason is that the stress state of the body triggers the release of the hormones, cortisol and adrenaline, which raises the body's blood pressure and causes the HR to increase. In this study, a medical grade biosignal data acquisition device, E4 Wristband, was used to measure the HR data of the participants. A photoplethysmogram (PPG) sensor equipped with the E4 wristband measured blood volume pulse (BVP), from which heart rate variability could be derived. Before a participant entered the simulator, they were asked to sit and relax for 10 minutes, so that the baseline of biosignal data could be collected. Ten minutes was found to be sufficient relaxing time, based on its use in other comparable studies (Ciabattoni et al., 2017; Grewen et al., 2005).

3.3.5 Decision-making (Learning Objectives and Performance Criteria)

The towing operation is a complex task that requires knowledge, experience and cooperation. During the case study, the towing performance, communication skills, decision-making skills and reaction time after the emergency occurred were evaluated. During the experiment, two experts/instructors commented on the participants' performance during the task. For decision-making, we only examined the decisions taken after the emergency was induced.

When the emergency happens during the towing operation, the most important thing for the decision-makers is to decide a safe and reasonable action as soon as possible. The quality of the decisions is not the most important aspect in this case, the efficiency is considered the priority. This is what the RPD mode requires. However, the quality of the decision can still be evaluated afterwards. In this study, an expert rating is proposed for the assessment (see Fig. 2).

From the rating of the decision, we can analyse the impact of the knowledge, experience and training method on the decision-making.

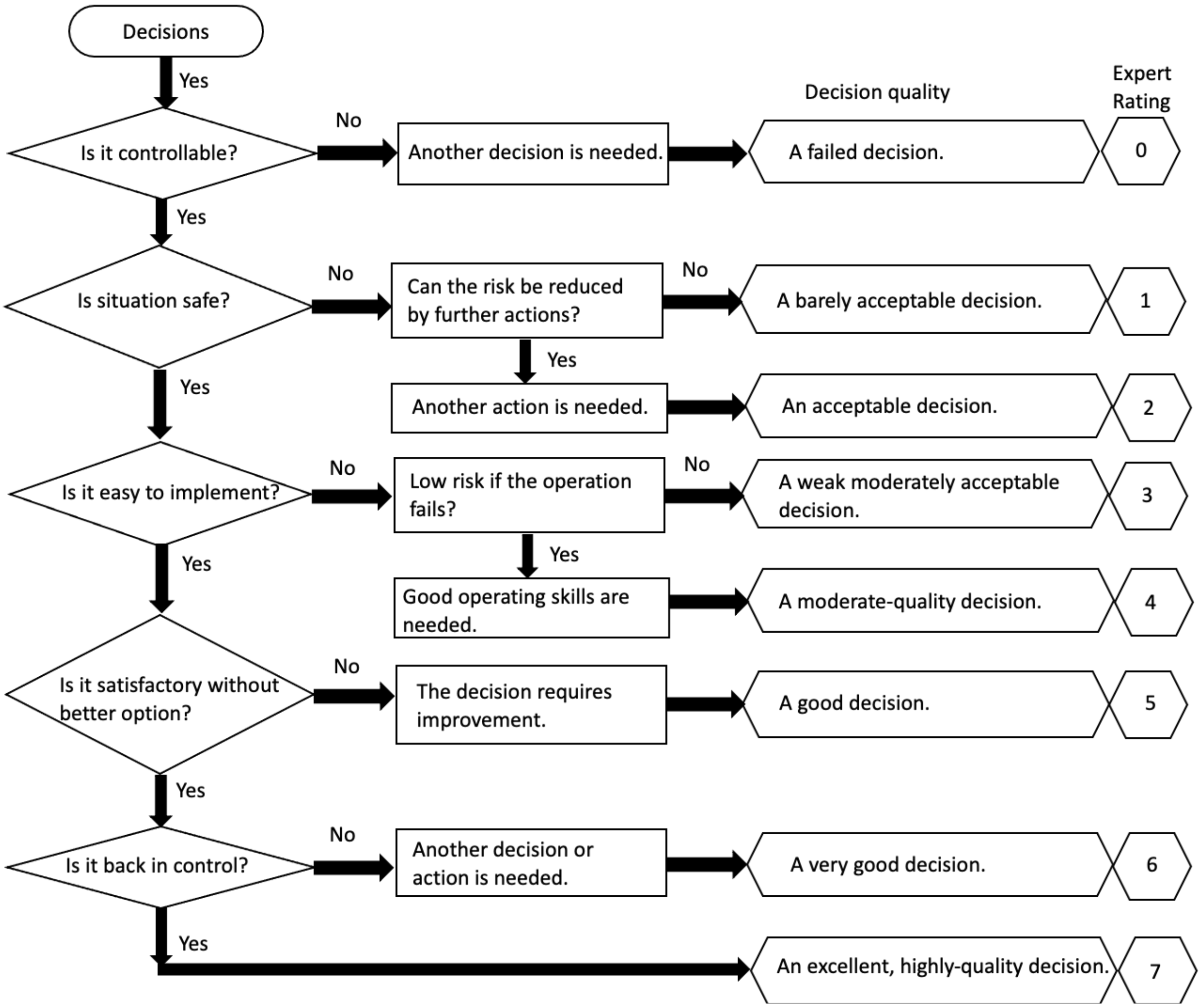


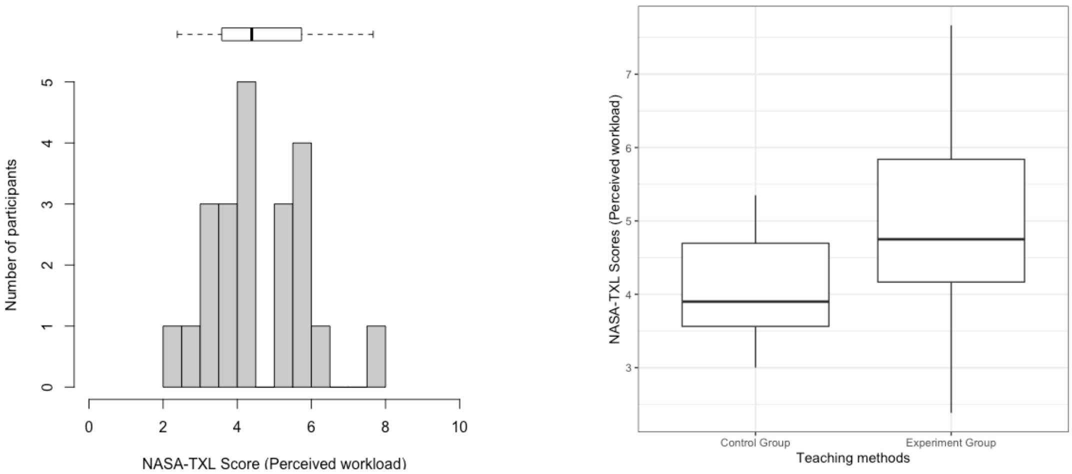
Fig. 2 Proposed customer decision quality rating scale.

In this proposed expert rating scale, there are five levels of requirements reflecting navigational safety. These requirements are presented in the form of questions, and status is judged based on the actual situation after a decision has been made. In order to obtain an accurate rating score, it is important to define the decisions correctly based on the situation. Regarding these requirements, the first level is whether the decision is “controllable”. It is necessary to know that all the vessels involved should be under control, which means that participants will know what to expect after the decision has been made. For example, neither the towed object nor the tugboats should drift. Otherwise, another decision is required, and the first decision will be considered a failed decision. The second level, “safe situation” means that there will not be any incident or accident after the decision-making. For instance, if there is a possibility or tendency to collide, it is considered an unsafe situation. The next level of the requirement is the ship handling skills involved. "Easy to implement" means that it does not require a high standard of ship handling skills to complete the operation. For instance, after the decision is made, the situation is expected to be under control and safe. However, if the navigator does not have sufficient skills to handle the operation, and the situation is not a controllable and safe situation, then if the operation fails and the situation is in low risk category, the rating score to this decision will be higher (a score of 4 in expert rating) than if the situation is in high risk category, which gives a score of 3. Next, we consider whether the decision is "satisfactory". If there is no better option for the decision, we move to the last stage to check whether the situation is "back in control". In this case, it means that the problem is solved, and the emergency is lifted.

4. Results

4.1 Workload

A one-way ANOVA (Analysis of Variance) method was used to find the effects of the different tugboats and different groups on the perceived workload. Results show that there was no statistically significant difference ($F(1,20) = 0.114$, $p > 0.05$) in the perceived workload on the different tugboats during the experiment. (Note that, commonly, if the p-value is higher than 0.05, we say that there is no statistically significant difference between the groups.) In another word, during the experimental task, the participants rated the workload of sailing on the front tugboat and the back tugboat as similar. In addition, participants in the experiment group perceived a higher workload (M = 4.85, SD = 1.44) than those in the control group (M = 4.13, SD = 0.85). However, the result of the one-way ANOVA also shows that there was no statistically significant difference ($F(1,20) = 1.65$, $p > 0.05$) in perceived workload as an effect of teaching methods. The overall perceived workload is 4.59 out of 10 with a standard deviation of 1.28. The summary statistics are depicted in Fig. 3.



(a) (b)

Fig. 3 Perceived workload measured using NASA-TXL. **(a)** The summary statistics (distribution and median) of the NASA-TXL score by all the participants. **(b)** The summary of the minimum, first quartile, median, third quartile, and maximum of NASA-TXL scores in groups.

4.2 Stress level

The subjective stress level was measured by STAI Form Y-1. Similar to the workload, one-way ANOVA was employed to analyse the effect of teaching methods and tugboats on the perceived stress. The results show that participants in the control group perceived higher stress (M = 46.5, SD = 3.78) than in the experiment group (M = 41.0, SD = 6.85), with a statistically significant difference ($F(1,20) = 4.34$, $p = 0.05$). Sailing on different tugboats had no effect on the perceived stress ($F(1,20) = 0.347$, $p > 0.05$). Based on the classification, the control group is considered to have a high anxiety stress level, and the experiment group is considered to experience a moderate anxiety stress level (Fountoulakis et al., 2006).

The objective stress level was measured by the HR of the participants. The data shows that HR increased significantly when the participants were towing (mean HR = 84.9, standard deviation = 8.13) compared to the situation during the relaxing time (mean HR = 76.6, standard deviation = 5.93). A paired t-test was employed to analyse the comparison. The results show a significant increase in HR when participants were performing the towing operation, $t(21) = 5.885$, $p < 0.001$. The HR data shows that participants were stressed under the towing operation, which is in line with the results from the STAI Form Y-1.

4.3 Decision-making and performance

Navigating and towing operations require frequent decision-making. Evaluating each decision along the way is more complicated than you might think because of the information available, number of outcomes, uncertainty in outcome, risk involvement, etc. Therefore, any type of decision is acceptable, as long as navigation is safe. In this study, we only looked at one critical situation and decision-making related to this situation where the emergency occurred. In order to examine the participants' ability to respond to emergencies, a chain of failures occurred in a short time. After the failures were induced, the quality of the decision-making was analysed by the proposed custom decision quality rating scale. For example, one of the towing team did not do anything after they were informed that the front tugboat had engine failure. After a while, they decided to drop anchors in the middle of the sea channel (Fig. A1), one after another, without communication. The situation is uncontrollable since it could be dangerous to all vessels, and the decision was rated as a failed decision. Moreover, the average time of the back tugboat cutting the line was counted as a critical factor in the evaluation. The results are presented in Table 3.

The scenario was the same for all the participants during the experiment. However, based on the speed and location when the failure occurred, participants with different training backgrounds make different decisions. For instance, when the emergency occurred, the front tugboat informed after some time the back tugboat that they had lost engine power. However, regarding the decisions of what to do next, each team made very different decisions. Some of the decisions and consequences are listed in Table 4. Even if participants make the same decision, the consequence may vary, as it is based on the current situation and other factors such as the participant's ship operating skills, experience and previous training. Fig. 4 shows an example of risky decision-making.

Table 3 Results of the reaction time and decision quality for the two groups.

Teaching methods (Groups)	**Back tugboat cut line time (average in minutes)**	**Decision quality rating**
Control group	0.87 min	71.4% (rating 5 out of 7)
Experiment group	3.59 min	47.1% (rating 3.3 out of 7)

Table 4 Different options for decision and the consequences of the decisions.

Decision-maker	**Options for decision (after the second engine failure is induced):**	**Consequences of decisions**
Control group	Front tugboat staff cut the line immediately, while the back tugboat staff need to cut the line and sail to the front.	The towing object may be lost, it starts drifting.
Experiment group	Front tugboat staff do not cut the line and want to be tugged together with the other tugboat.	Pulling both the towing object and the other tug is

		higher risk than tugging only one vessel.
Experiment group and control group	Front tugboat changes direction after the engine failure.	Avoiding being hit by the towed ship and giving a way for the back tugboat to go to the front.
Experiment group	Front tugboat staff decide to launch anchor.	No consequences if it is detached from the towed ship. If it is still connected to the towed ship, it creates an unstable situation.
Experiment group	Back tugboat reduces the speed and tries to stop the towed ship.	Avoid collision.
Experiment group	Back tugboat sails backwards.	Difficult to maintain safety.
Control group	Back tugboat staff cut the line without reducing the speed and sail to the front.	Good ship handling skill is demanded, and an unstable situation is created.
Experiment group	Back tugboat sails between the towing object and the front tugboat.	This represents a short cut, but there is a high risk of collision. Good skill is required.

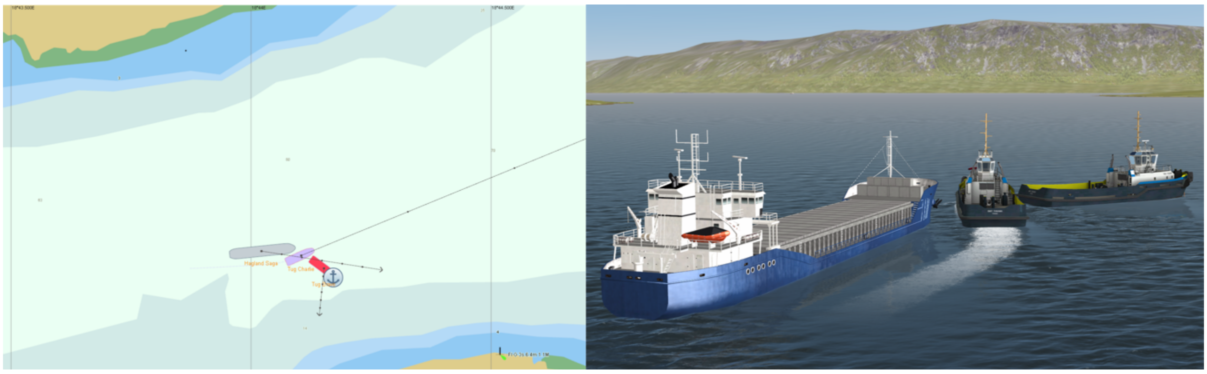


Fig. 4 An example of a risky decision and its corresponding 3D view are shown on the map. One of the participants decided to cross the narrow passage between the tugboat and the disabled ship, an action which requires considerable ship handling skills.

4.4 Correlation

Correlation analyses were performed to determine the interaction among the workload, stress, and quality of decisions as the factors dependent on the teaching methods. The correlations between these variables are listed in Table 5. In order to calculate the correlation, we set the experiment groups who had undergone a rapid training course as index 1 and the control groups who had received conventional training as index 2. The results show that a higher stress level is associated with the period of the participants' study (r = 0.42, p = 0.05). In addition, the quality rating of the decision is also correlated with the length of time the participants have been studying ($r = 0.55$, $p = 0.008$).

Table 5 Correlations between variables; results presented are the Pearson Correlation coefficient, *r*
(Note that, for numbers marked *, then $p \leq .05$ and we have strong correlation.) The matrix is presented in such a way that it is symmetrical.

	1. Teaching Methods	2.NASA-TXL Total	3.Stress Level	4.Decision Quality Rating
1. Teaching Methods (Groups)	-			
2. NASA-TXL Total	-.28	-		
3. Stress Level (STAI Form Y-1)	.42*	-.04	-	
4. Decision Quality Rating	.55*	-.29	.11	-

5. Discussion

In this study, teaching methods were examined and evaluated using decision-making performance in a towing task. The aim was to better understand the effect of training decision-making skills in a simulator-based maritime MET environment. By utilizing the simulator at the university and inviting different groups of participants with different training backgrounds, this study investigated the impact of routine maritime training on improving decision-making skills and how individual decisions are reflected in ship handling skills. In addition, perceived workload and different levels of stress were compared, to test the sensitivity.

The utilization of the NASA-TXL rating as a subjective measurement to measure the towing workload revealed that all participants perceived a similar workload, and that there was no significant difference between those working on the front and back towing tugs. This reflects that the teaching methods did not result in different perceived workloads. Furthermore, the towing operation simulated represents a collective operation with two persons working as a pair, and hence it is difficult to distinguish the workloads of working on different tugboats.

The towing operation is a stressful task. From results of the self-assessment of stress level, the participants in the control group (being in their second year of study) perceived more stress than those in the experiment groups (with limited training in using the simulator). The HR data also shows that all the participants experienced a significant increase in heartbeats per minute from their relaxing time to the time they were performing the towing task. However, there was no significant difference in heart rate increases among participants who received different training methods. This reflects the limitation of HR as an objective measurement of stress in discriminating between the different stress levels.

From the results of the decision-making, the control group is found to have made more homogeneous decisions than the experiment group. This can be explained from the assumption that the participants in the control group had a deeper theoretical knowledge regarding towing operations than the participants from the experiment group. The experiment group had been through a project-specific rapid content training (crash course), consisting of a 20-minute video lecture and a one-hour simulator session, with seven participants participating in each session. The control group had been through several simulator exercises, including three towing operation tasks and lectures over a certain amount of time. The results imply that it is possible to teach students how to handle a tug through such a crash course, but it is difficult to obtain the same level of knowledge, skills and understanding of the situation, compared to regular teaching over a longer period.

The results from analysis of the participants' decision-making show that the model of naturalistic decision-making is suitable for this kind of analysis. Many of the characteristics of naturalistic decision-making, such as high stakes, time stress and uncertain dynamic environments (Orasanu and Connolly, 1993), are present in this exercise. For example, the participants can experience high stakes and a certain amount of time pressure during the experiment; meanwhile, the surrounding conditions change, caused or not as a result of the decision made by participants, as time passes. The decisions can also be considered a chain of events, as one decision will affect the next decision. An example could be the decision to cut the line between the functional tug and the object being towed. This decision will lead to a change in the number of options, increase the time pressures and certainly affect the participants' next decision.

From the results, it seems that the decisions made by the control group can be explained by the use of the RPD mode because the participants in the control group had undergone three towing operation exercises at different levels during their routine training and one more year of study with several simulator-based training exercises than the participants in the experiment group. Therefore, the scenario was somewhat familiar to them, and they used their experience from previous simulator-based exercises and lectures as input for making their decisions. Since all the control group participants had been through the same exercises and lectures, that could be the reason for them making homogenous decisions, as the RPD mode is applied. Moreover, the reason for the participants in the control group experiencing a higher level of stress could be their awareness that they should know this task and be able to handle this situation, based on their previous exercises and lectures. They might expect to perform better than before. However, the participants in the experiment group might think that their performance is not as important, as they only have been through a short and rapid training course, and therefore their perceived stress level is lower. In addition, the participants in the experiment group also have limited experience, so they might not realize the severity of the situation and what could go wrong.

When it comes to the decisions for the participants in the experiment group, it can be assumed that they also initially try to follow the RPD mode; however, their level of experience is not high enough to apply this mode effectively, which means that they are not able to find relevant input from the crash course or other simulator-based exercises to assist them in making their decision. This hypothesis is strengthened, for instance, by the data from the time elapsing before they decide to cut off the line. It is significantly longer for the experiment group, which means that they took more time to assess the situation, search for previous experience that could guide them to their decisions and/or wait for more information. As the time pressure increases, there is reason to believe that many of them would start exploring the field of creative decisions rather than recognition-primed decisions. This hypothesis is strengthened by the high number of different solutions from this group. As the different solutions are analysed, solutions can be found that are not in line with either the theory from the crash course or other simulator-based exercises. These solutions might by effective, as the example mentioned above where the participant decided to pass between the front tug and the object being towed. However, the solution is connected to high risk and could lead to increased complexity for the next set of decisions that must be made.

To summarize, it seems that the participants who have been through conventional teaching over a longer period are able to apply their knowledge and skills at a deeper level when exposed to unfamiliar and critical situations, compared to participants who have been through project-aimed rapid training.

6. Conclusions and future work

Consequently, this study answered the two questions: (1) Project-aimed rapid training can give enough knowledge for participants to some degree to help them make efficient decisions in stressful and critical situations and (2) Different training methods can affect the decision-making model applied by the participants. This research finds that the participants who have been through conventional teaching over a longer period are more able to apply their knowledge and skills at a deeper level when exposed to unfamiliar and critical situations, compared to participants who have been through project-aimed rapid training. Participants who had only rapid training did not have enough experience to apply the RPD mode; however, most of them managed to use creative decisions to solve the problems.

In this experiment, the participants had a very compressed training content, including a high number of participants simultaneously in the simulator-based practising. Based on the findings, we recommend that, in project-aimed rapid training, the training time should be increased appropriately and the number of simultaneous participants in simulator-based practising should be reduced, as this can effectively improve the learning outcomes. Good learning outcomes can improve decision-making skills in emergency situations. Well-designed project-aimed rapid training enables the trainees to accomplish the training task effectively.

Further, the results indicate that conventional teaching over a longer period is important for establishing a solid fundament. Solid fundamentals are the foundation for creative solutions to unfamiliar situations, for instance emergency situations. To be able to deal with emergencies safely and effectively, experience and knowledge are required. In particular, the number of simulator-based exercises can have an impact on the success rate of a specific task, even if the exercises are not necessarily task-aimed. The more simulator-based training undertaken, the better the decision-making skills.

For future work, we intend to implement project-aimed rapid training in other parts of training and education, to make the routine/regular education programme more efficient. Another option is to execute extended project-aimed rapid training with increased training time, to measure the effect of the length of the training.

Acknowledgements

The authors would like to thank instructor Gudmund Johansen, from the Nautical Science Group at UiT, The Arctic University of Norway, for providing support; Lars Daniel Bottolfs Håland, from UiT, The Arctic University of Norway, for translating the dialogue; and all the participants who gave us their time and approval to use the data in this study.

References

Allen, C. H. (2004). *Farwell's Rules of the Nautical Road*. 8th ed. Annapolis, Maryland: Naval Institute Press.

Bao, J., Li, Y., Duan, Z., Li, T. and Zhang, P. (2021). Key Factors Affecting the Quality of Maritime Education and Training: Empirical Evidence from China. *The Journal of Navigation,* **74**(2), 396–408. https://doi.org/10.1017/S0373463320000740.

Basak, S. K. (2017). A Framework on the Factors Affecting to Implement Maritime Education and Training System in Educational Institutions: A Review of the Literature. *Procedia Engineering*, 10th International Conference on Marine Technology, MARTEC 2016, **194** (January), 345–50. https://doi.org/10.1016/j.proeng.2017.08.155.

Bryman, A. (2012) *Social Research Methods.* 4th ed. Oxford University Press.

Čampara, L., Frančić, V. and Bupić, M. (2017). Quality of Maritime Higher Education from Seafarers' Perspective. *Pomorstvo,* **31**(2), 137–50. https://doi.org/10.31217/p.31.2.8.

Carotenuto, A., Fasanaro, A. M., Molino, I., Sibilio, F., Saturnino, A., Traini, E. and Amenta, F. (2013). The Psychological General Well-Being Index (PGWBI) for Assessing Stress of Seafarers on Board Merchant Ships. *International Maritime Health,* **64** (4), 215–20. https://doi.org/10.5603/IMH.2013.0007.

Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L. and Verdini, F. (2017). Real-Time Mental Stress Detection Based on Smartwatch. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 110–11. https://doi.org/10.1109/ICCE.2017.7889247.

Cicek, K. and Er, I. D. (2008). Economic Constraints on Maritime Training and Education in Turkey. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, **2**(2), 193–96.

Cohen, M., Freeman, J. and Thompson, B. (1998). “Training the Naturalistic Decision Maker.” In *Decision Making Under Stress: Emerging Themes and Applications*. Gower TEchnical. https://doi.org/10.4324/9781315095080-11.
DeKeyser, R. (2020). “Skill Acquisition Theory.” In *Theories in Second Language Acquisition*, 3rd ed. Routledge.
Fountoulakis, K. N., Papadopoulou, M., Kleanthous, S., Papadopoulou, A., Bizeli, V., Nimatoudis, I., Iacovides, A. and Kaprinis, G. S. (2006). Reliability and Psychometric Properties of the Greek Translation of the State-Trait Anxiety Inventory Form Y: Preliminary Data. *Annals of General Psychiatry,* (1), 2. https://doi.org/10.1186/1744-859X-5-2.
Grewen, K. M., Girdler, S. S., Amico, J. and Light, K. C. (2005). Effects of Partner Support on Resting Oxytocin, Cortisol, Norepinephrine, and Blood Pressure Before and After Warm Partner Contact. *Psychosomatic Medicine,* **67**(4), 531–38. https://doi.org/10.1097/01.psy.0000170341.88395.47.
Gudmestad, O. T., Rettedal, W. K., Sand, S. S., Brabazon, P., Trbojevic, V., and Helsøe, E. (1995). *Use of simulator training to reduce risk in offshore marine operations* (No. CONF-950695-). American Society of Mechanical Engineers, New York, NY (United States).
Gug, S.-G., Yun, J-H., Harshapriya, D. and Han, J-J. (2022). A Prefatory Study on the Effects of Alcohol on Ship Manoeuvring, Navigational and Decision-Making Abilities of Navigators. *The Journal of Navigation*, April, 1–13. https://doi.org/10.1017/S0373463322000133.
Hanzu-Pazara, R., Barsan, E., Arsenie, P., Chiotoroiu, L., & Raicu, G. (2008). Reducing of maritime accidents caused by human factors using simulators in training process. *Journal of Maritime Research,* **5**(1), 3-18.
Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P. A. and Meshkati, N. (ed.), *Advances in Psychology*, **52**, 139–83. Human Mental Workload. North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9.
Hetherington, C., Flin, R. and Mearns, K. (2006). Safety in Shipping: The Human Element. *Journal of Safety Research,* **37**(4), 401–11. https://doi.org/10.1016/j.jsr.2006.04.007.
Hjelmervik, K., Nazir, S., & Myhrvold, A. (2018). Simulator training for maritime complex tasks: an experimental study. *WMU Journal of Maritime Affairs, 17*(1), 17-30.
Hystad, S. W. and Eid, J. (2016). Sleep and Fatigue Among Seafarers: The Role of Environmental Stressors, Duration at Sea and Psychological Capital. *Safety and Health at Work,* **7**(4), 363–71. https://doi.org/10.1016/j.shaw.2016.05.006.
Inoue, K. (2000). Evaluation Method of Ship-Handling Difficulty for Navigation in Restricted and Congested Waterways. Journal of Navigation, 53(1), 167–80. https://doi.org/10.1017/S0373463399008541.
Jaeyong, O. H., Park, S., & Kwon, O. S. (2016). Advanced navigation aids system based on augmented reality. *International Journal of e-Navigation and Maritime Economy*, **5**, 21-31.https://doi.org/10.1016/j.enavi.2016.12.002.
Jensen, H.-J. and Oldenburg, M. (2021). Objective and Subjective Measures to Assess Stress among Seafarers. *International Maritime Health,* **72**(1), 49–54. https://doi.org/10.5603/IMH.2021.0007.
Kim, T-e., Sharma, A., Bustgaard, M., Gyldensten, W. C., Nymoen, O. K., Tusher, H. M. and Nazir, S. (2021). The Continuum of Simulator-Based Maritime Training and Education. *WMU Journal of Maritime Affairs,* **20**(2), 135–50. https://doi.org/10.1007/s13437-021-00242-2.
Klein, G. (1997). Developing Expertise in Decision Making. *Thinking & Reasoning,* **3**(4), 337–52. https://doi.org/10.1080/135467897394329.
Klein, G. A. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In: *Decision Making in Action: Models and Methods*, 138–47. Westport, CT, US: Ablex Publishing.
Klein, G. A., Orasanu, J., Calderwood, R. and Zsambok, C. E. (eds.) (1993). *Decision Making in Action: Models and Methods*. Norwood, N.J: Ablex Pub.
Markopoulos, E., Lauronen, J., Luimula, M., Lehto, P. and Laukkanen, S. (2019). Maritime Safety Education with VR Technology (MarSEVR). In *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 283–88. https://doi.org/10.1109/CogInfoCom47531.2019.9089997.
Nazir, S., Jungefeldt, S., and Sharma, A. (2019). Maritime simulator training across Europe: a comparative study. *WMU Journal of Maritime Affairs, 18*(1), 197-224.
Norros, L. and Hukki, K. (2003). Utilization of information technology in navigational decision-making. *Cooperative Process Management: Cognition And Information Technology: Cognition And Information Technology, 77.*
Orasanu, J. and Connolly, T. (1993). The Reinvention of Decision Making. In: *Decision Making in Action: Models and Methods*. Norwood, N.J: Ablex Pub.

Sampson, H. (2004). Romantic Rhetoric, Revisionist Reality: The Effectiveness of Regulation in Maritime Education and Training. *Journal of Vocational Education & Training,* **56**(2), 245–67. https://doi.org/10.1080/13636820400200256.
Sampson, H. and Thomas, M. (2003). The Social Isolation of Seafarers: Causes, Effects, and Remedies. *International Maritime Health,* **54**(1–4), 58–67.
Schein, E. H. and Sloan School of Management (1992). How Can Organizations Learn Faster?: The Problem of Entering the Green Room. https://dspace.mit.edu/handle/1721.1/2399.
Sharek, D. (2011). A Useable, Online NASA-TLX Tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* **55**(1), 1375–79. https://doi.org/10.1177/1071181311551286.
Spielberger, C. D. (1983). State-Trait Anxiety Inventory for Adults. https://doi.org/10.1037/t06496-000.
Starcke, K. and Brand, M. (2012). Decision Making under Stress: A Selective Review. *Neuroscience & Biobehavioral Reviews,* **36**(4), 1228–48. https://doi.org/10.1016/j.neubiorev.2012.02.003.
Størkersen, K. V., Laiou, A., Nævestad, T. O. and Yannis, G. (2018). Production and Protection. Seafarers' Handling of Pressure in Gemeinschaft and Gesellshaft. In: *Safety and Reliability – Safe Societies in a Changing World*. CRC Press.
Vrijkotte, T. G. M., van Doornen, L. J. P. and de Geus, E. J. C. (2000). Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. *Hypertension,* **35**(4), 880–86. https://doi.org/10.1161/01.HYP.35.4.880.
Waern, Y. (2003). *Cooperative Process Management: Cognition and Information Technology: Cognition and Information Technology*. CRC Press.
Wróbel, K., Montewka, J. and Kujala, P. (2017). Towards the Assessment of Potential Impact of Unmanned Vessels on Maritime Transportation Safety. *Reliability Engineering & System Safety,* **165** (September), 155–69. https://doi.org/10.1016/j.ress.2017.03.029.
Wu, B., Cheng, T., Yip, T. L. and Wang, Y. (2020). Fuzzy Logic Based Dynamic Decision-Making System for Intelligent Navigation Strategy within Inland Traffic Separation Schemes. *Ocean Engineering,* **197** (February): 106909. https://doi.org/10.1016/j.oceaneng.2019.106909.
Xiufeng, Z., Biguang, H., Yicheng, J. and Yong, Y. (2005). Simulating Test of Ship Navigation Safety Evaluation Using Ship Handling Simulator. *Proceedings of OCEANS 2005 MTS/IEEE*, 1902-1905 Vol. 2. https://doi.org/10.1109/OCEANS.2005.1640037.

Declarations

Ethics approval
The project was approved by NSD Norsk Senter for Forskningsdata (Norwegian Centre for Research Data).

Conflict of interest
The authors declare no competing interests.

Appendix A. Assessment of participants' skill levels

Groups	Number of participants	Skills on average	Simulator-based training time
Control group	8	Good navigation skills, good ship handling skills, reasonably good communication skills, some emergency handling skills.	32 exercises or 96 hours of training
Experiment group	14	Good navigation skills, moderate ship handling skills, less efficient communication skills, no emergency handling skills.	16 exercises or 48 hours of training

Note that the project-aimed rapid training was aiming to minimize the skills gap between the two groups.

Appendix B View from simulator

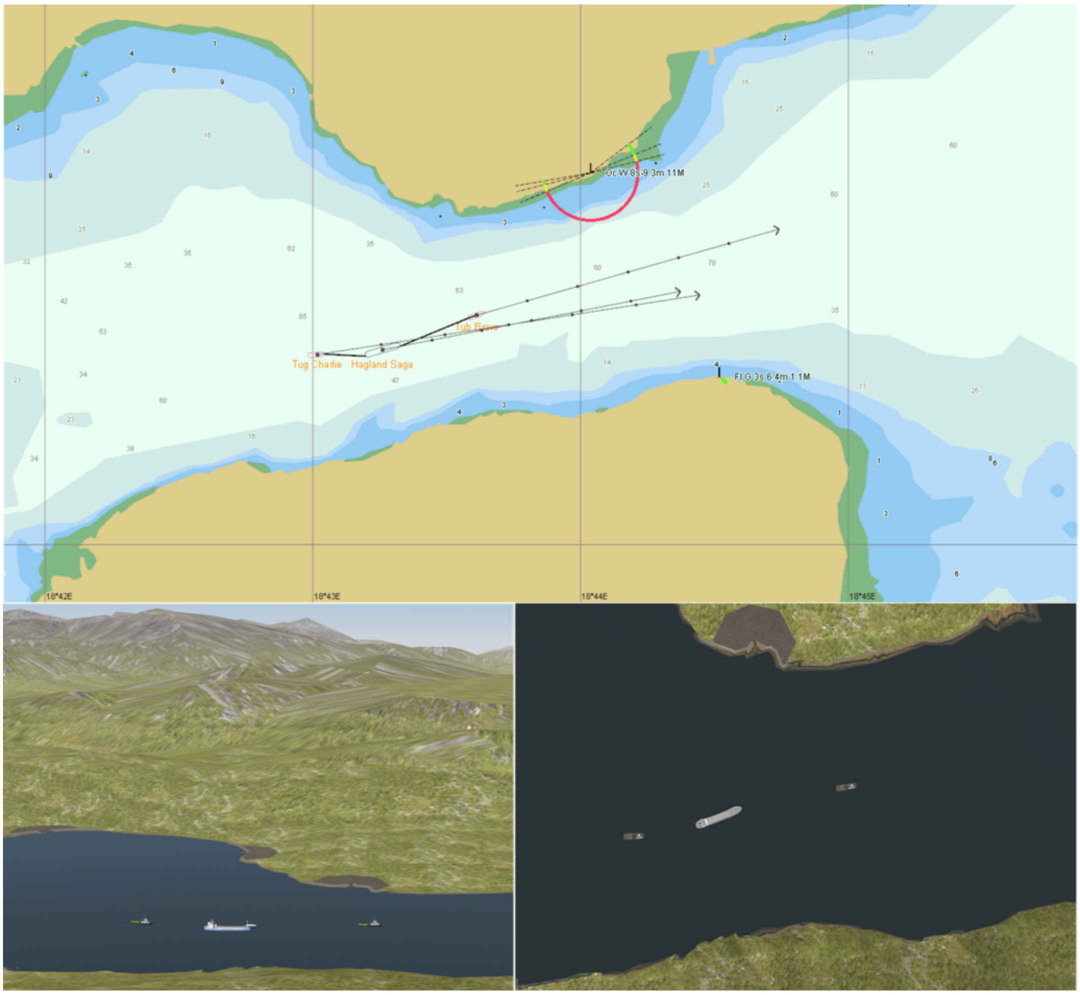


Fig. A1 View from the simulator at UiT, The Arctic University of Norway. The location where the critical situation took place shown on the map and its corresponding 3D views in two different directions of vision.

Paper V

Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing performance

H. Xue, Ø. Haugseggen, J.-F. Røds, B.-M. Batalden, D. K. Prasad

Submitted to *Transportation Research Interdisciplinary Perspectives*

Highlights

Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing route reliability

Hui Xue, Øyvind Haugseggen, Johan-Fredrik Røds, Bjørn-Morten Batalden, Dilip K. Prasad

- Construction and comparison of various machine learning methods for classifying biosignal data.
- Examine the correlation between self-assessed stress levels and objective stress levels.
- Evaluate the impact of stress on performance in maritime navigation and sailing route reliability.
- A model of a stress-based maritime training program is proposed.
- Experimental results verify the proposed training model is achievable.

Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing route reliability

Hui Xue[a,*], Øyvind Haugseggen[a], Johan-Fredrik Røds[a], Bjørn-Morten Batalden[a], Dilip K. Prasad[b]

[a]*Department of Technology and Safety, UiT The Arctic University of Norway, Postboks 6050 Langnes, Tromsø, 9037, Norway*
[b]*Department of Computer Science, UiT The Arctic University of Norway, Postboks 6050 Langnes, Tromsø, 9037, Norway*

Abstract

Maritime training can improve safety by equipping seafarers with the knowledge and skills to manage risk. However, designing a quality training program can be challenging and stress can negatively impact performance and safety. To address this, the present study aims to investigate the relationship between stress and training outcomes, with the goal of developing more effective stress-based training systems. Two stressful scenarios were designed with varying safety factors involved during navigation tasks. The study examines the impact of stress levels on training outcomes and performance based on safety factors and the correlation between self-assessed stress levels and objective stress levels obtained from biosignal data. The study was conducted in a simulated bridge environment in Tromsø, Norway, and analyzed using statistical tests and machine learning models. The findings of this study indicate that training scenarios can be classified by stress levels, which were found to be associated with reduced visibility, equipment failures, and severe weather conditions. Additionally, the study revealed that stress levels can negatively impact performance in maritime navigation and sailing route reliability. These findings provide insights into how to improve the quality and effectiveness of maritime training programs and ultimately enhance safety at sea.

Keywords: maritime navigation, biosignal, machine learning, stress, simulation-based training
PACS: 01.40.-d, 01.50.-i, 01.50.-H, 07.05.Kf, 28.41.Te
2000 MSC: 62R10, 62H30, 68T10

*Corresponding author
Email address: `hui.xue@uit.no` (Hui Xue)

Abbreviations

The next list describes several symbols that are used within the body of the document:

AR	Augmented Reality
BVP	Blood volume pulse
CPA	Closest point of approach
DWT	discrete wavelet transforms
ECG	Electrocardiogram
EEG	Electroencephalography
EOG	Electrooculogram
HOC	Higher-Order Crossings
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-beat interval
KNN	K-nearest neighbors
LDA	Linear discriminant analysis
ML	Machine learning
NASA-TXL	NASA Task Load Index
NAVAID	Navigational aid
NM	Nautical miles
PPG	Photoplethysmography
SA	Situation awareness
STAI	State-Trait Anxiety Inventory
SVM	Support vector machine
VAS	Psychometric evaluation of a visual analogue scale
VHF	Very High Frequency
VR	Virtual reality

1. Introduction

The growth of advanced technology on board ships in the maritime industry has led to an improvement in safety measures over the years. Despite these efforts, however, the rate of accidents has not seen a significant decrease. A plethora of studies have indicated that human factors are a major contributing factor to this phenomenon, with estimates of contributing to 75-96% of the accidents [1, 2, 3].

Traditionally, risk assessment in the maritime industry has been hindered by a lack of standardized accident reporting systems [4, 5]. However, with the advent of alerting and reporting systems for maritime incidents [6], analysis of accident trends through statistical methods has become more prevalent. Additionally, the use of virtual maritime simulators to study human factors has gained popularity, as they provide a comprehensive means of collecting information on board.

Studies have also revealed that incompetent officers are often a significant contributor to shipping accidents. This highlights the importance of high-quality maritime training in order for seafarers to acquire the knowledge and skills necessary to effectively manage risk and ensure safety at sea [7]. Simulator-based maritime training is a widely used method, due to its ability to provide a controlled environment, adjustable task difficulty levels, cost-effectiveness, and a risk-free practice environment. Furthermore, virtual maritime simulators are useful in designing exercises that allow for the comparison of student performance and learning outcomes.

However, developing and evaluating a quality maritime training program is challenging, as it involves a variety of factors such as student skill levels, exercise design, and assessment of learning outcomes, etc. In particular, assessing learning outcomes can be difficult, as traditional methods such as written and oral exams may not accurately reflect a student's capacity to process information during a sea voyage [8, 9]. Additionally, performance assessment is often evaluated subjectively by instructors, which can be unreliable, invalid, and unfair [10]. Furthermore, studies have indicated that psychophysiological states such as cognitive workload and stress levels are key factors affecting performance [11]. Therefore, monitoring stress levels and workload during assessments is crucial.

Furthermore, human behavior and physiology adapt to stress in such a way that performance remains stable within a certain range of stress levels, this is called the "comfort zone" where the level of learning and response is optimal [12]. In the maritime domain, stress not only affects the health and well-being of seafarers but also negatively impacts maritime navigation safety by distracting attention, memory retrieval, and decision-making [13]. In other words, high safety-related stress can impair safety performance by allocating limited cognitive resources to different aspects of performance, such as work requirements or emergency tasks, leading to compromised compliance and participation in safety performance [14].

The above description illustrates the critical nature of stress as it affects safety and training outcomes in the maritime industry. However, stress-based training systems have not been thoroughly studied, and research on objective stress analysis in the maritime field is limited, particularly in the measurement of biosignal-based stress levels. There is a need for further research in this area to effectively assess and address the impact of stress on maritime training and safety. In light of this, the current study aims to examine the relationship between stress and training outcomes and establish a foundation of data for stress-based training systems. The study is designed to accomplish this by: (1)Creating two different scenarios involving stress with varying numbers of stressful events in the same navigation tasks; (2)Measuring the relationship between self-reported stress levels and objective stress levels measured from biosignal data; (3)Determining whether stress levels are associated with safety factors in navigation tasks such as visibility, equipment failure, and traffic situations; (4)Assessing learning outcomes and performance to determine the impact of stress on training programs. Overall, this study aims to contribute to a better understanding of the relationship between stress and training outcomes in the maritime industry in order to improve safety and optimize training programs.

The research paper is organized in the following manner: Section 2, the methodology for evaluating the workload and stress levels in maritime navigation is presented. Section 3, the details of the study's experiment are outlined, including both self-assessment and biosignal-based assessment of the stress level and workload during maritime tasks. The process of applying machine learning algorithms for biosignal data analysis is also illustrated in this section. Section 4, the results of the data analysis are presented and discussed. Section 5, the findings from the biosignal data and simulation data are discussed in terms of stress levels and training performance. Finally, the conclusions of the study are presented and suggestions for future work are provided in section 6.

1.1. Related Work

Research in the field of maritime safety and training has shown that marine accidents are closely related to the untimely, negligent, and incorrect decision-making of seafarers' situation awareness (SA) forecasting. Quality maritime training is essential in equipping seafarers with the knowledge and skills to manage risks, solve problems, and conduct operations safely and efficiently, thus ensuring the safety of life at sea [7]. With the advancement of technology, maritime training has evolved from traditional simulator-based training to training methods that incorporate the use of various advanced technologies, such as augmented reality (AR) and virtual reality (VR) technology, and multi-sensor frameworks as auxiliary equipment. These technologies have been found to enhance the training of seafarers' SA and decision-making skills. For example, the use of VR glasses provides a fully immersive virtual environment for training and makes the experience more engaging and enjoyable, like playing a game [15]. Due to their portability and ease of use, VR technologies allow students to train at their convenience, increasing opportunities for training and enhancing their SA and other skills.

Other training methods are used in conjunction with the simulator, including the use of AR glasses in simulators, providing a semi-immersive experience. Students can learn and practice related knowledge by the application set up in the AR glasses, reducing the repetitive work of the instructor [16]. Another pilot study in maritime training employed a multi-sensor fusion framework, using the training method of briefing/debriefing in the simulator, collecting audio, video, eye-tracking data, etc., visualizing operational procedures, thereby achieving the goal of improving the SA of seafarers [17].

In addition to training, the assessment of stress and workload states is also a crucial indicator of maritime safety. Research has shown that working at sea can be stressful and is a risk factor for maritime safety. Assessing the stress and workload of seafarers and improving the working environment at sea is vital for ensuring safety [18, 19]. In the past, research has predominantly relied on subjective measurements, such as surveys and self-reported measures, as stress is difficult to measure objectively [20]. However, with the advancement of sensor and system technology, researchers have begun to use wearable sensors and biosignal data to analyze stress levels in various fields. For example, the use of the human voice to detect pilot stress and workload [21], and eye movements measured with an Electrooculogram (EOG) to identify different emotional states [22]. The use of an Electrocardiogram (ECG) to monitor stress while driving has been found to prevent safety risks and traffic accidents caused by driving fatigue [23].

In the field of maritime navigation, for the advantage of wearable sensors that can continuously monitor the psychophysiological state of the human body without interfering with the subject's activities, biosignal-based tools are increasingly being used. Pilot studies have been conducted using Electroencephalography (EEG) data to identify seafarers' cognitive stress and workload during simulator exercises and to recommend performance improvements [11]. These technologies, however, are complex, costly, and may not be practical for use with large numbers of students. These shortcomings make various training methods still in the pilot study stage and have not been widely popularized. In light of these limitations, heart rate (HR) and heart rate variability (HRV) have been identified as the most convenient, simple, and accurate indicators of stress emotion assessment when compared to other methods. This is because the collection of HRV and HR data only requires the subject to wear a device with a photoplethysmography (PPG) sensor on the wrist, which is commonly available in smartwatches and wristbands. Additionally, it is well established in the literature that stress is correlated with high heart rate levels, hence HRV can be utilized to estimate stress levels with a high level of accuracy. This has been demonstrated in various studies that have focused on using HRV as the primary feature for stress assessment [24, 25, 26, 27, 28].

Additionally, the use of machine learning (ML) algorithms in biosignal data analysis have been found to significantly improve the accuracy of stress level assessment. ML, which comprises a set of methods for learning from data and uncovering patterns within it, can be used to extract meaningful insights from physiological data [29]. However, it is essential to note that the accuracy of using publicly available physiological datasets in maritime settings, which are typically emotionally annotated in environments where users are exposed to intense stressors, remains uncertain [30]. This is due to the subjective nature of stress, which can vary greatly across different settings. Therefore, the use of appropriate data and proper methodology is crucial for ML-based stress assessment studies. To the best of our knowledge, there have been few studies on the use of biosignal data, specifically HR/HRV data, to assess stress levels and evaluate performance in maritime training.

To establish a stress-related maritime training system, we aim to investigate the following research hypotheses: (1) determine if biosignal data is sufficient to be an objective tool to assess stress levels in maritime training, (2) examine if the complexity of scenarios can be classified based on biosignal data, and (3) evaluate how stress levels affect training performance. The conceptual model illustrated in Fig. 1 demonstrates the relationship among safety factors and highlights the connection between stress and maritime training programs. The results of these analyses will be studied in the subsequent sections.

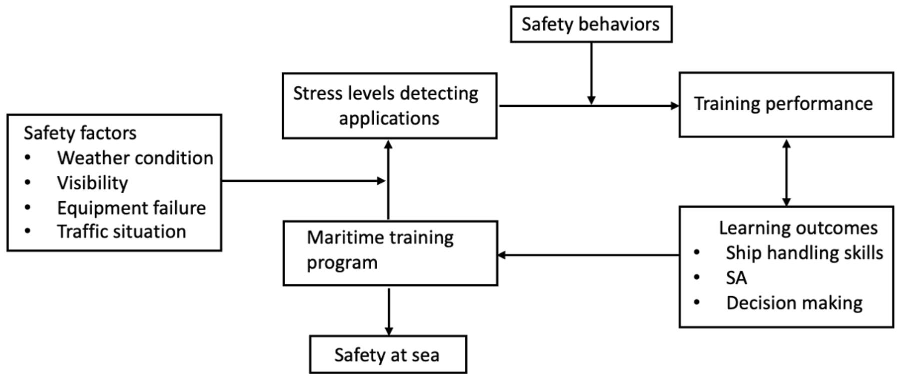


Figure 1: Conceptual model of a stress-based maritime training program.

1.2. Objective and Contributions

The objective of the study is to evaluate the effectiveness of using biosignal data, specifically heart rate and heart rate variability, as an objective tool for assessing stress levels in maritime training. The main contributions of the study include the following:

1. Investigating the relationship between stress levels and performance during maritime training through a systematic evaluation of stress level analysis in simulator-based training.

2. Demonstrating the reliability of analyzing stress levels using biosignals obtained from wearable sensors, providing a new tool for assessing the reliability of maritime training, and laying the foundation for a proposed stress-based training system.

3. Introducing a novel method for analyzing biosignal data, including the use of preprocessing techniques and feature selection methods, specifically the use of Higher-Order Crossings (HOC)-Based Features extraction, which provides a good classification result on the biosignal data.

4. Proposed a conceptual model that illustrates the relationship among the safety factors and shows the connection between stress and the maritime training program. This model can serve as a guide for future research in the field of stress analysis and maritime training.

2. Methodology

2.1. Participants

A total of 23 nautical science students from UiT The Arctic University of Norway (UiT) voluntarily participated in the study. The demographic characteristics of the participants include a mean age of 22.43 years (standard deviation = 2.35 years) and a gender distribution of 7 females and 16 males. Prior to the study, all participants were administered the Patient Health Questionnaire (PHQ-9) [31, 32, 33] for a screening of depression. The participants were randomly divided into three groups for the sailing tasks, with 22 (mean age = 22.36 years, standard deviation = 2.38 years) valid data samples analyzed and included in the study. All participants provided informed consent for their participation in the trial.

2.2. Materials and Apparatus

In order to investigate the relationship between the complexity of maritime navigation training scenarios and the stress levels of participants, two distinct levels of complexity were evaluated using a simulated environment. The experiment was conducted on three different simulator bridges, all of which were equipped with the K-sim Navigation software from Kongsberg Digital and featured a 240° and 360° view. Each simulator bridge was equipped with an independent instructor station, enabling the simultaneous execution of three exercises. The vessel model utilized in the study was the BULKC11 Hagland Saga, a small bulk carrier with a length between perpendiculars of 85 meters, and was deemed appropriate for the tasks being evaluated. Additionally, all participants were familiar with the vessel model as a result of their prior navigational training.

The participants were randomly divided into three groups: a control group (Group C), an experiment group 1 (Group E1), and an experiment group 2 (Group E2). Group C performed the easy scenario twice, while the experiment groups completed either an easy scenario followed by a complex scenario (Group E1) or a complex scenario followed by an easy scenario (Group E2), with a 10-minute break between the two sections.

Each participant wore a medical-grade wearable device, the Empatica E4 Wristband, to collect biosignal data. The E4 wristband is equipped with a PPG sensor that measures blood volume pulse (BVP) from which HR can be derived. Before the trial, participants were asked to spend 10 minutes in a seated and relaxed position, and the resulting biosignal data were collected as the baseline.

2.3. Scenario Design

Sailing route. The experiment utilized the sailing route of Sandnessundet as the location for navigational training. Sandnessundet is a strait located between Tromsøya and Kvaløya in the Tromsø municipality of Troms in Norway, which spans approximately 14 kilometers in length and is traversed by the Sandnessund bridge, connecting the Kvaløysletta district to the Tromsø city center, as described in Norgeskart [34]. This route is commonly used for navigational training for nautical students at UiT The Arctic University of Norway. The route, as depicted in Figure 2, starts in the southern region of the strait and proceeds north, making a sharp turn towards the northeast. It then passes under a tall, narrow bridge before opening up until it reaches the end of Tromsøya. The participants will encounter two fishing vessels and a tug during their navigation on this route, as shown in Figure 3.

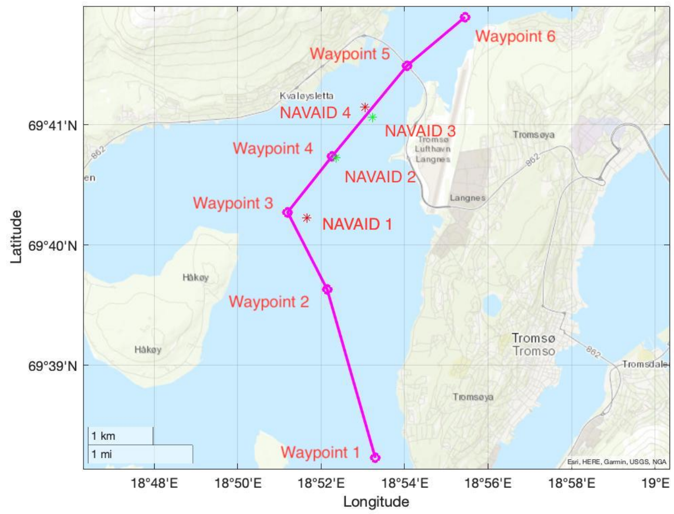


Figure 2: The planned sailing route Sandnessundet consists of five straight legs. Waypoint 5 is located at the midpoint of the Sandnessund Bridge. The map of the route includes the placement of four navigational aids (NAVAIDs) which were distinguished by two distinct colors.

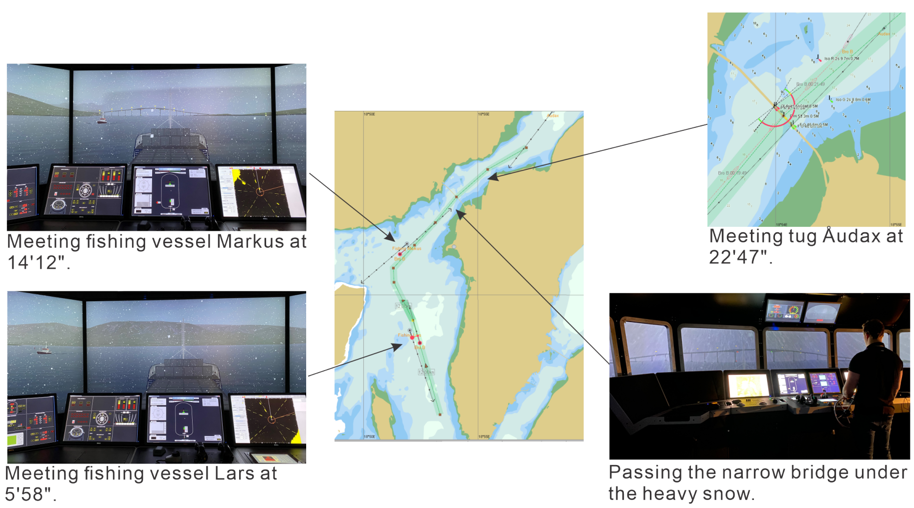


Figure 3: An illustration of the sailing route of one of the participants, highlighting the geographical locations of the traffic situations.

Table 1: Events design.

Event time	Event in control task scenario	Event in experiment task scenario
0.5 min	None.	Steering pump failure.
2 min	Weather forecast VHF.	Weather forecast VHF.
4 min	None.	Echo sounder failure.
After 6.5 min	Meeting fishing vessel.	Meeting fishing vessel.
9 min	None.	Gyro failure.
After 13 min	Meeting fishing vessel.	Meeting fishing vessel.
16 min to 18.5 min	None.	Add the snow intensive 100% at 16 min, and then change the snow intensive back to 50% at 18 min. Stop the snow after 18.5 min.
After 20 min	Passing narrow bridge.	Passing narrow bridge.
22 min	GPS failure.	GPS failure.
After 22 min	Meeting tug.	Reduce visibility (fog intensive 100%), and meeting tug.
Total:	6	10

Events in the sailing task. In this study, the maritime navigation training scenarios were designed to have no current, tidal stream, or wind. Two different levels of complexity were used, based on the number of events that occurred during the sailing tasks. The control task scenario was conducted under fair weather conditions with six events, while the experimental task scenario was performed under snowy weather conditions with an additional four events compared to the control task scenario. Table 1 presents a comparison of the events in the two different scenarios at the same time point. Other simulated variables, such as location and traffic situation, were kept constant across the two trials.

2.4. Learning Objectives and Performance Criteria

The learning objectives of the control task and the experimental task are identical, which include::

- Learning when and where to fix the position in the chart during the sailing.
- Adhering to the planned route.
- Managing and maintaining a safe distance from other vessels while navigating.
- Handling equipment malfunctions.

In order to evaluate the achievement of the learning objectives, performance was evaluated using the following metrics:

- Number of position fixes in the chart.
- Deviation of the actual route from the planned route, with the deviation score being calculated based on the distance from the planned course using the assessment tool within the simulator. Deviation also can be calculated

mathematically as follow: The distance between two points in geographic coordinates can be calculated using a mathematical formula, Eq. (1) :

$$D = arccos[sin(LatA) * sin(LatB) + cos(LatA) * cos(LatB) * cos(LongA - LongB)] * 3440.1 * 1852 \quad (1)$$

where D is the distance in meters, $LatA$ is the latitude of point A expressed in radians, $LatB$ is the latitude of point B expressed in radians, $LongA$ is the longitude of point A expressed in radians, $LongB$ is the longitude of point B expressed in radians, 3440.1 is the radius of the earth in nautical miles (NM), and 1 NM is 1852 meters.

The distance between the sailing point and the planned route between two waypoints can be derived using Heron's formula [35].

- Score graded based on the closest point of approach (CPA). CPA was calculated based on the speed and direction of the approaching ship, as CPA is an essential factor of ship safety, particularly in situations where the ship must avoid a collision. [36].

3. Experiment

In this study, a comprehensive analysis of both questionnaire data related to stress and workload assessment, as well as biosignal data, is conducted to investigate the classification of complexity of maritime navigation training scenarios and the associated stress levels. As illustrated in Figure 4, the analysis includes data pre-processing and the application of machine learning (ML) algorithms. To assess the subjective stress levels of the participants, several validated questionnaires were utilized. The results of these questionnaires were analyzed using statistical tests to determine the significance of the differences in stress levels between the control and experimental scenarios. The results indicate a significant difference in stress levels between the two scenarios. Based on these findings, it is hypothesized that the biosignal data collected during the control and experimental scenarios can be classified. To verify this hypothesis, features were extracted from the biosignal data and analyzed using various ML algorithms.

3.1. Self-assessment of the Stress Level and Workload

Psychometric evaluation of a visual analogue scale (VAS) for the assessment of stress . VAS was administered to each participant following the completion of each scenario. The VAS scale ranges from 0 to 10, with 10 indicating the highest level of stress. Participants were instructed to mark their perceived stress level on the scale immediately after completing each sailing scenario. The use of a VAS for the assessment of stress has been previously validated in clinical research [37].

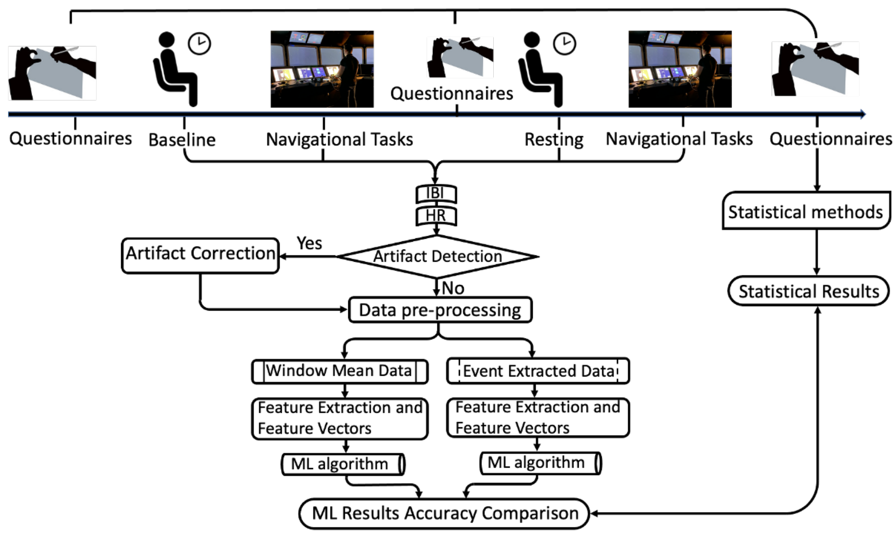


Figure 4: Mixed-methods approach for stress level analysis in maritime training.

State-Trait Anxiety Inventory (STAI) Form Y-1 [38]. STAI Y-1 form is a widely used self-assessment tool for evaluating state and trait anxiety in individuals. The questionnaire, which consists of 20 questions, is designed to measure the participant's current feelings and emotions [39]. The scores obtained from the STAI Y-1 form are commonly classified into three categories: "no or low anxiety" (20-37), "moderate anxiety" (38-44), and "high anxiety" (45-80). These ranges are used as a benchmark to classify the level of anxiety experienced by the participants.

NASA Task Load Index (NASA-TLX) . NASA-TLX is a widely recognized assessment tool that is used to evaluate the perceived workload of participants in a given task[40, 41]. NASA-TLX consists of six categories that are rated by participants following the completion of each sailing scenario. These categories include Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. The ratings are then converted to a ten-point scale score, with 0 representing low levels of workload and 10 representing high levels of workload [42].

3.2. Biosignal Data Pre-Processing

In this study, data on the inter-beat interval (IBI) was extracted from a photoplethysmogram (PPG) sensor embedded in an Empatica E4 wristband. The IBI also referred to as the RR interval, is the time interval between individual heartbeats. Data on incorrect peaks were removed prior to analysis. A sample of RR intervals for a participant is illustrated in Figure 5. The instantaneous heart rate, measured in beats per minute (bpm), was derived from the IBI values using the following formula (Eq.(2)) :

$$HR[bpm] = 60/IBI \tag{2}$$

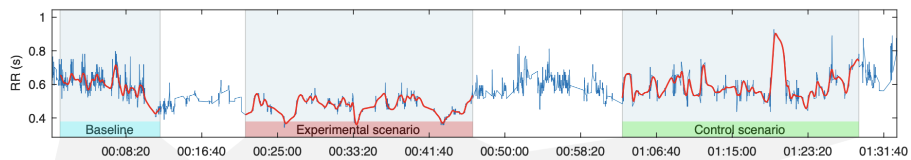


Figure 5: Sample RR intervals with activity windows recorded from a participant during the sailing task are shown. Note that the gaps between windows represent the time between activities.

In the analysis, HR data were analyzed from the collected IBI data. The frequency of HR data is 1 Hz. The average HR during the relaxation period was calculated for each group as the baseline. The cleaned HR data of each participant were subtracted from the group's baseline, resulting in the HR difference (HRD) data. Additionally, two data preparation methods were employed:

- **Window mean data (D_W):**

  The mean of the window data was calculated for each HRD data of each participant using a window size of every 30 seconds and a step size of every 15 seconds.

- **Event extracted data (D_E):**

  The HRD data were extracted after one minute of every event.

3.3. Classification Features Extraction

Three types of features are extracted:

(1) Statistical-Based Features.
In this study, statistical-based features were created in two types (Eq.(3)). The first one was the mean of the HRD of each participant in each task. The second one was the standard deviation.

$$F_S = [\mu_{X_i}, \sigma_{X_i}], \qquad (i = 1, 2, ..., l) \tag{3}$$

where F_S is the statistical-based feature vector, μ_{X_i} is the mean of the data series, σ_{X_i} is the standard deviation of the data series, X_i is the HRD of each participant in each task, l is the length of the X_i.

(2) Wavelet-Based Features.
In this study, wavelet-based features were extracted based on the coefficients of the discrete wavelet transforms(DWT), specifically the Daubechies wavelets (with a number of vanishing moments of 4) [43, 44]. The wavelet coefficients were computed for specified scales [45], in this case, 2, 4, and 8, in order to obtain three levels of scales. The resulting matrix of the

wavelet coefficients had three rows and columns equal to the length of the HRD data for each participant in each task. Subsequently, wavelet-based features were computed using two different methods, as outlined in Eq.(4). The first method was the sum of the square of the wavelet coefficients, while the second method was the sum of the product of the square of the wavelet coefficient and the natural logarithm of the square of the wavelet coefficient.

$$\begin{aligned} F_W &= [F_{W_1}, F_{W_2}] \\ F_{W_1} &= [\sum_{i=1}^{l} Y_{L_1}^2, \sum_{i=1}^{l} Y_{L_2}^2, \sum_{i=1}^{l} Y_{L_3}^2] \\ F_{W_2} &= [\sum_{i=1}^{l}(Y_{L_1}^2 * ln(Y_{L_1}^2)), \sum_{i=1}^{l}(Y_{L_2}^2 * ln(Y_{L_2}^2)), \sum_{i=1}^{l}(Y_{L_3}^2 * ln(Y_{L_3}^2))] \end{aligned} \tag{4}$$

where F_W is the wavelet-based feature, F_{W_1} and F_{W_2} are the two different ways of computing, Y is the Daubechies wavelet coefficient in three levels L_1, L_2, and L_3, and l is the length of the prepared data.

(3) Higher-Order Crossings (HOC)-Based Features.
Higher-Order Crossings (HOC)-based features, also known as zero-crossing-based features, are a set of features that are extracted from the analysis of the patterns of zero-crossings in a signal. Zero-crossing, a commonly used concept in signal processing, refers to the point at which the signal changes from positive to negative or vice versa [46]. In this study, the HOC features were extracted in the following steps:

- Computing the difference between adjacent elements in data series in different orders. The k^{th} order difference is (see Eq. (5) [47]):
- From $\nabla^{k-1} Z_t$, a binary process $X_t^{(k)}$ was defined in Eq. (6) [48, 47, 49]):
- The count of the symbol changes from $X_t^{(k)}$, D_k, was calculated in Eq. (7) [47, 48, 42]:

$$\begin{aligned} \nabla^{k-1} Z_t &= \sum_{i=1}^{k} C_{i-1}^{k-1} (-1)^{i-1} Z_{t+1-i} \\ with \quad C_{i-1}^{k-1} &= \frac{(k-1)!}{(i-1)!(k-i)!} \end{aligned} \tag{5}$$

where $k = 1, 2, ...,$ and ∇^0 is the zero-mean data series we computed before.

$$X_t^{(k)} = \begin{cases} 1, & \nabla^{k-1} Z_t \geq 0 \\ 0, & \nabla^{k-1} Z_t < 0 \end{cases} \tag{6}$$

where $k = 1, 2, ...$

$$D_k = \sum_{t=2}^{N} [X_t^{(k)} - X_{t-1}^{(k)}]^2 \tag{7}$$

where D_k is the count of symbol changes in k^{th} order.

Above all, the extraction of HOC-based features from the biosignal data was represented by a vector consisting of the number of axis crossings in a zero-mean data series outlined in Eq. (8). The resulting HOC-based features were found to be beneficial in improving the performance of the machine learning (ML) models used in the study, providing useful insights and better accuracy in identifying and classifying biosignals. As illustrated in Figure 6, the number of crossing with the order of derivative varies for the two HR signals from the same participant performing tasks of different levels.

$$F_{HOC} = [D_1, D_2, .., D_L], \qquad (1 < L < J) \tag{8}$$

where F_{HOC} is the HOC features, J denotes the maximum order of the estimated HOC and L is the HOC order used in this study. D_1 denotes the number of axis crossing in the zero-mean data series, D_2 denotes the number of axis crossing in the first difference of the series, D_3 denotes the number of axis crossing in the second series, and so on.

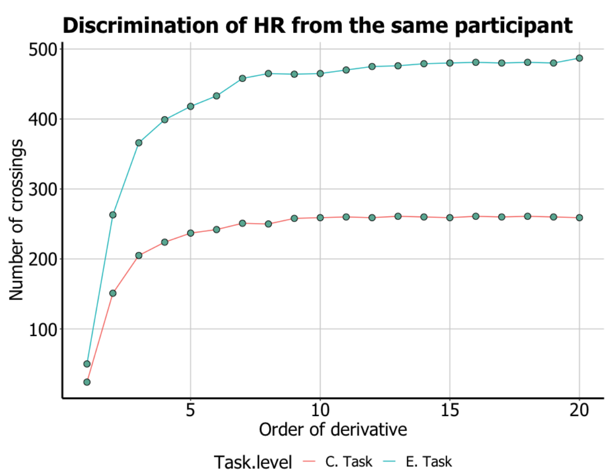


Figure 6: Graphical comparison of HOC features from the same participant doing a different level of the task.

3.4. Machine learning (ML) algorithms

Following ML algorithms are used to be compared in the study (see Table 2). The classification models and their main parameters are resented in the table.

Table 2: A summary of classification models' parameters.

Classification model	**Main parameters**
SVM	Kernel function: Linear.
KNN	Using 6 nearest neighbour(s) for classification.
Naive Bayes	Use a kernel estimator for numeric attributes.
LDA	Multivariate Gaussian for each class, ridge 10^{-6}.
Logistic Regression	With ridge parameter of 10^{-8} coefficients.

3.5. K Folds Cross-validation and ML Performance Measure

In this study, in order to ensure that every sample is included in both the training and testing sets, a commonly used machine learning validation method, K-folds cross-validation, was employed. Ten folds were selected as a standard utilization.

In the context of ML classification problems, precision and recall metrics were employed as performance measures in addition to classification accuracy. This is because when the class of samples is imbalanced, the large number of examples from the majority class can overwhelm the number of examples in the minority class, resulting in unskilled models achieving high accuracy scores. Precision and recall metrics include precision, recall, and F-Score. Precision evaluates the fraction of correctly classified instances among those classified as positive [50]. Recall is typically used to measure the coverage of the minority class [51]. The F-Score weights precision and recall equally [50]. The following equations, (9), (10), and (11), provide the definitions for these measures.

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$F - Score = \frac{2 * P * R}{P + R} \tag{11}$$

where P denotes precision, R to recall, TP to True Positives, FP to False Positives, and FN to False Negatives. TP and FP belong to Positive Prediction, and FN belongs to Negative Prediction.

4. Results

4.1. Self-Assessment of the Stress Level Results

The present study aimed to investigate the relationship between self-assessment stress levels and training performance in the context of maritime navigation. To do so, several questionnaires were used to measure the stress levels and workload of participants during training sessions in both control and experimental scenarios. Results were analyzed using a combination of statistical methods, including the Kruskal-Wallis H test, Spearman rank correlation coefficient, and Welch Two Sample t-test.

(1) Kruskal-Wallis H test.
The present study utilized the Kruskal-Wallis H test to determine whether the medians of ratings from the three groups (C, E1, and E2) were different. The Kruskal-Wallis H test (also as known as "one-way ANOVA on ranks") is a rank-based non-parametric statistical test that can be used to determine if there are statistically significant differences between two or more independent groups on a continuous or ordinal dependent variable [52, 53]. This test was applied to the data obtained from the three questionnaires that were used to assess the stress levels and workload of the participants in each group.
The results of the Kruskal-Wallis H test were visualized in Figure 7 and are presented in Table 3. The test statistic was calculated using Eq. (12) [54] and the degrees of freedom were determined using Eq. (13). The corresponding p-value was calculated using the chi-square distribution with 2 degrees of freedom.
The results of the Kruskal-Wallis H test showed that there was a statistically significant difference in stress levels and workload between the three groups in the three questionnaires. These findings indicate that the experimental scenarios had a relatively strong effect on the stress levels and workload of the participants and support the use of the Kruskal-Wallis H test as a tool for analyzing the data obtained from the questionnaires in this study.

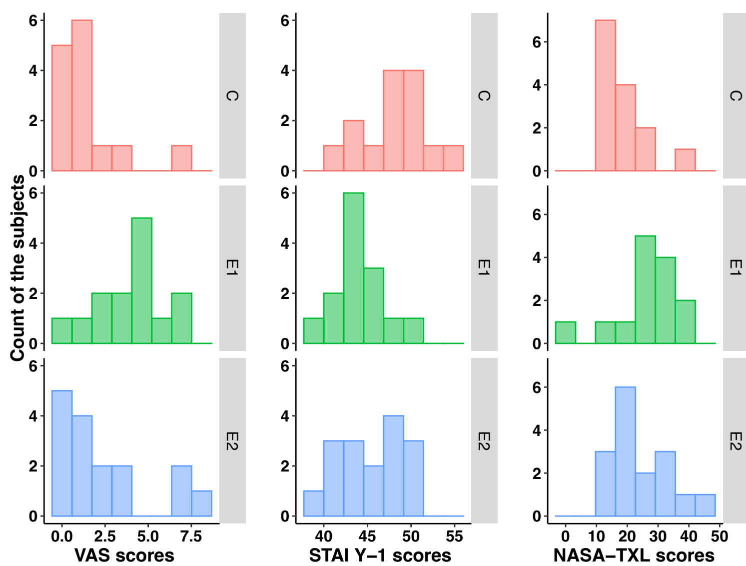


Figure 7: Visualization of the results of the questionnaires from each group.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i(\overline{R_i} - \overline{R})^2 \tag{12}$$

where H is the test statistic, $N = 44$ is the total data sample size (three groups and rated for two different level tasks) for each questionnaire, $k = 3$ is the number of groups we are comparing, n_i is the sample size for group i ($n_C = 14, n_{E1} = 14, n_{E2} = 16$), $\overline{R_i}$ is the average of the ranks in a group i, $\overline{R}$ is the average of all the ranks among all samples.

$$df = k - 1 = 2 \tag{13}$$

where df is degrees of freedom, and $k = 3$ is the number of groups we are comparing.

Table 3: Questionnaire statistical results, Kruskal-Wallis H test results.

Questionnaire	**H statistic**	**p-value**	**Effect size**	**Conclusion**
VAS	8.0353	0.01800	0.1869 (Relatively strong)	Statistically significant
STAI Form Y-1	8.0894	0.01752	0.1881 (Relatively strong)	Statistically significant
NASA-TLX	7.3748	0.02504	0.1715 (Relatively strong)	Statistically significant

(2) Spearman rank correlation coefficient.

The relationship between the performance of participants and their perceived workload was also of interest in this study. The Spearman rank correlation coefficient (also known as the Spearman rho) was employed to assess the association between the two variables. The results indicated that there was a moderate and statistically significant association between the workload rating given by the participants and their scores on the performance assessment. Specifically, the Spearman correlation coefficient (rho) was $\rho = -0.3171226$, with a p-value of 0.03595. This suggests that as the perceived workload of the participants increased, their performance scores decreased, and vice versa.

(3) Welch Two Sample t-test.
A question of interest in this study was the self-evaluated stress levels of participants during both the sailing control scenario and the experimental scenario. The results of this investigation are presented in Figures 8 and 9. Figure 8 illustrates the results from the three questionnaires as grouped by participant groups (C, E1, and E2) respectively. The results, presented in figures 8, indicate that participants reported higher levels of stress as measured by the Visual Analog Scale (VAS) and NASA-Task Load Index (NASA-TLX) during the experimental scenario compared to the control scenario. Scores on the State-Trait Anxiety Inventory Form Y-1 (STAI-Y1) were found to be similar across both scenarios. Subsequently, the questionnaire results from participants who sailed in both the control and ex-

perimental scenarios were analyzed. Figure 9 compares the results of the questionnaires, as grouped by E1 and E2, respectively. The results, presented in Figure 9, indicate that participants in both groups E1 and E2 reported higher levels of stress in the VAS and NASA-TLX questionnaires during the experimental scenario compared to the control scenario. However, the results for the STAI-Y1 questionnaire revealed a different pattern, with group E1 reporting higher scores during the experimental scenario and group E2 reporting lower scores.

To further investigate these findings, a Welch Two Sample t-test was conducted on the data, with a 95% confidence interval (CI) for the mean difference. The Welch t-test is a parametric test that assumes a normal distribution of data, and thus, a normality test (Shapiro-Wilk) was performed to ensure that the assumptions of the test were met. In this study, the transformation method of the square root was used for moderate positive skew (see Eq. (14)). The results of the t-test, presented in Table 4, indicate that there was a statistically significant difference in stress levels as measured by the VAS between the control and experimental scenarios, with participants reporting higher levels of stress in the experimental scenario. No significant differences were found for STAI-Y1, and there was a statistically significant difference in perceived workload as measured by NASA-TLX between the control and experimental scenarios, with participants reporting a higher workload in the experimental scenario. Cohen's d was also calculated to measure the effect size, and it was found to be a large effect on VAS and NASA-TLX while small on STAI-Y1.

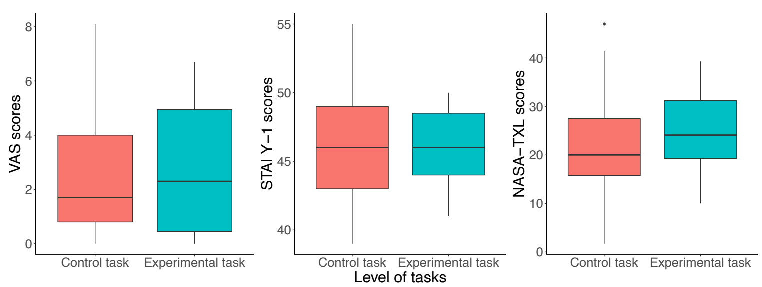


Figure 8: Visualization of the results of the questionnaires from two different levels of scenarios.

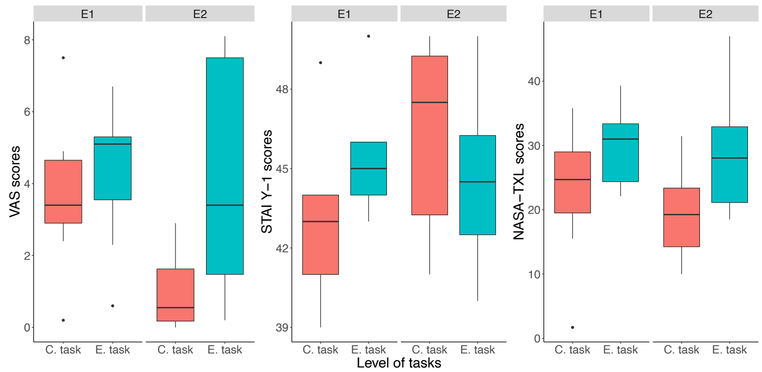


Figure 9: Comparison of the questionnaire results from two different levels of the scenario by groups E1 and E2 (E1 and E2 have different orders of sailing the control scenario and experimental scenario).

$$S_{norm} = \sqrt{S} \tag{14}$$

where S is the data sample (scores of VAS of doing control task), S_{norm} is the normally distributed data sample.

Table 4: Welch t-test results of the questionnaires. The star * means the value was calculated after transforming the data to normal distribution.

Questionnaires		Shapiro test normality (p-value)	Mean	SD	Welch Two Sample t-test df	t-statistic	p-value	95% CI		Effect size Cohen's ds
VAS	C.Scenario E.Scenario	0.4548* 0.3376	1.9069 4.1800	1.9840 2.6622	15.13	-4.2886	0.00063	-4.4993	-1.5133	1.0179 (large)
STAI Form Y-1	C.Scenario E.Scenario	0.2674 0.5404	46.3793 44.9333	4.0037 2.7894	38.101	1.3969	0.1705	-0.6493	3.5413	0.3968 (small)
NASA-TLX	C.Scenario E.Scenario	0.4342 0.5140	19.9862 29.3633	8.5724 8.0871	29.945	-3.5714	0.0012	-14.7398	-4.0144	1.1145 (large)

4.2. Results of the Objective Assessment

In this study, the stress level of the participants was objectively assessed by analyzing HR data obtained from IBI data collected via wearable sensors. Figure 10 illustrates that the range of HR values for participants in the control scenario is generally smaller than that in the experimental scenario. However, it is difficult to discern a significant difference in the average HR between the two scenarios. To address this, ML algorithms were employed to classify HR data from the two different scenarios. Five different ML algorithms were selected and their results were compared using three different methods of pre-processing the HR data. The results, as shown in Figure 11, indicated that when using the event extraction method, all five ML algorithms achieved high accuracy. Conversely, when using the window mean data or raw data directly, the accuracy was found to be relatively low, as detailed in Table 5.

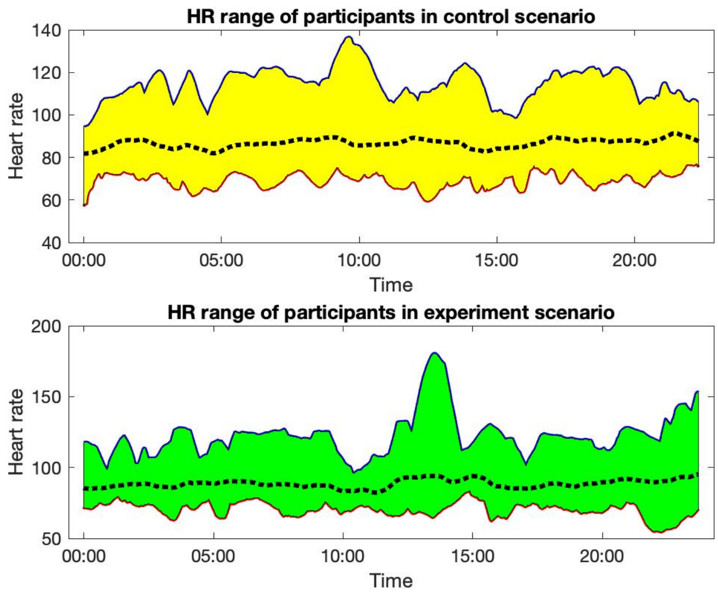


Figure 10: Maximum and minimum HR of participants in the control scenario and experimental scenario. The dashed lines represent the mean HR.

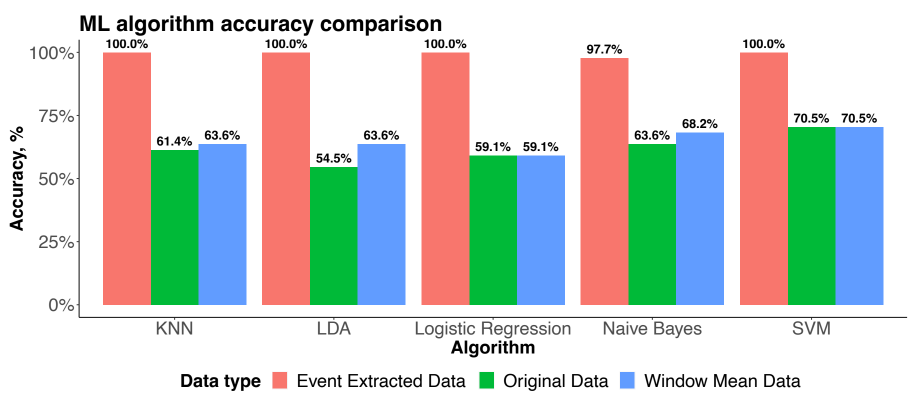


Figure 11: Comparison of the machine learning results from five different algorithms in a different way of pre-processing data.

4.3. Results of the Performance

In this study, the performance of the participants was evaluated based on a set of established criteria. The criteria used to assess performance included the number of times participants fixed their position during the voyage, the deviation from the planned route, and the participant's ability to maintain safe

Table 5: Detailed accuracy by the recall, precision, and F-Score for biosignal data classification in different types of data pre-processing.

Classification model	Recall			Precision			F-Score		
	D_R	D_W	D_E	D_R	D_W	D_E	D_R	D_W	D_E
SVM	0.705	0.705	1	0.796	0.796	1	0.619	0.619	1
KNN	0.614	0.636	1	0.424	0,429	1	0.501	0.513	1
Naive Bayes	0.636	0.682	0.977	0.576	0,677	0.979	0.572	0.603	0.977
LDA	0.545	0.636	1	0.588	0.662	1	0.557	0.644	1
Logistic Regression	0.591	0.591	1	0.591	0.578	1	0.591	0.583	1

clearance when encountering two fishing vessels. Each criterion was scored on a scale from 0 to 3, with higher scores indicating better performance. For example, a score of 3 was awarded for fixing a position more than 3 times, while a score of 0 was given for fixing a position less than 3 times. Similarly, a maximum score of 3 was awarded for deviations from the planned route less than 180 meters and a minimum score of 0 for deviations greater than 1000 meters. When encountering fishing vessels, a maximum score of 3 was given for CPA greater than one nautical mile (nm), and a minimum score of 0 for CPA less than 0.5 nm. The total maximum score was 12. The results, presented in Table 6, indicate that participants tended to fix their position more frequently in the control scenario and maintained closer proximity to the planned route and better traffic clearance when encountering fishing vessels. Overall, the participants performed better in the control scenario.

Table 6: The comparison of the scores from sailing in different scenarios based on the proposed criteria.

Criteria	Positioning (3)	Deviation (3)	TC 1 (3)	TC 2 (3)	Scores in total (12)
Control scenario	0.86	2.71	2.50	1.79	7.86
Experimental scenario	0.25	2.43	2.03	2.03	6.75

In addition to the performance criteria, an examination of the deviation from the planned route was conducted. The result, as illustrated in Figure 12, indicates that while participants generally adhered to the planned route in the initial stages, deviation increased as they approached the midpoint of the route. A two-sample t-test was performed to compare the deviation in the control scenario and experimental scenario. The results reveal a statistically significant difference in a deviation between the control scenario (geometric mean M = 39.34, SD = 60.76) and experimental scenario (geometric mean M = 43.21, SD = 70.25); $t(41431) = 8.2681$, $p < .001$. Furthermore, as shown in Figure 13, the majority of deviation in the control scenario remained within 300 meters from the planned route, with the majority of instances between 0 and 200 meters away. Conversely, deviation in the experimental scenario was primarily greater than 200 meters. Additionally, the results suggest that participants in the experimental scenario were able to return to the planned route more swiftly, whereas participants in the control scenario took more time to do so.

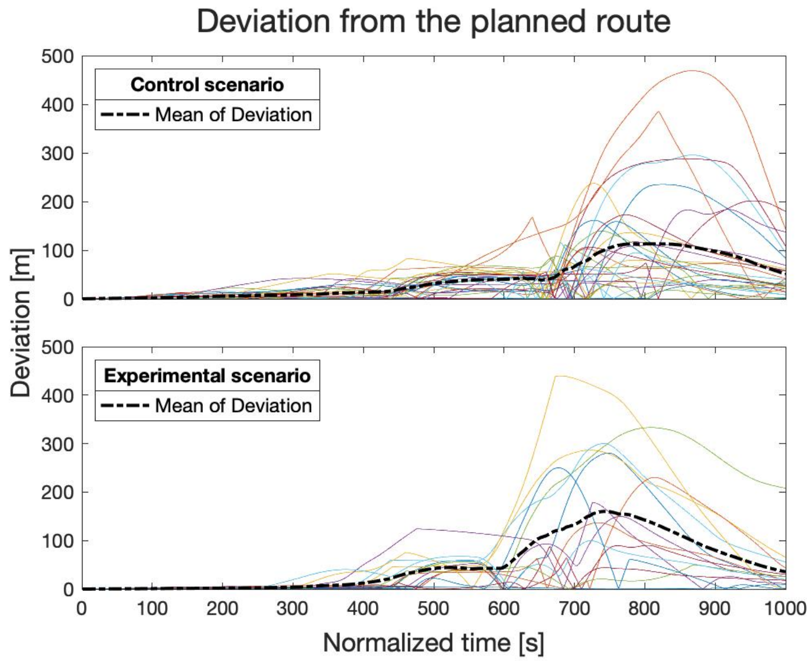


Figure 12: Deviation from the designed route for the control scenario (upper) and the experimental scenario (lower). The black dashed lines represent the mean of the deviation.

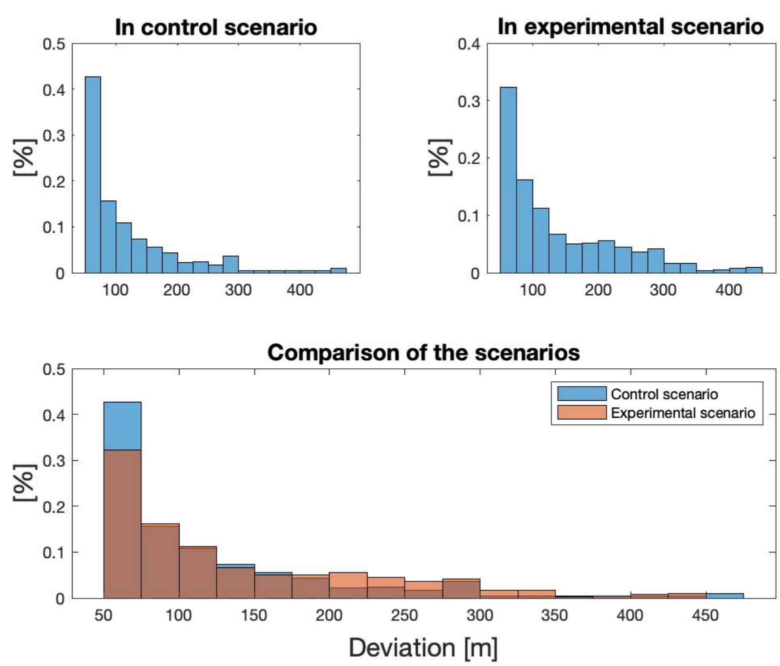


Figure 13: Deviation from the planned route for the experimental scenario (orange) and the control scenario (blue) in the comparison graph (experts suggest that the deviation of shorter than 100 meters is negligible).

Table 7: Summary of findings.

Hypothesis Number	Description	Accepted/ Rejected
H1	The biosignal data is sufficient to be an objective tool to assess stress levels in maritime training.	Accepted
H2	The complexity of the scenarios can be classified based on the biosignal data.	Accepted
H2.1	The more events in the scenario, the more stress the seafarers will get, especially an event on top of another event at the same time.	Accepted
H2.2	Abysmal visibility and complex traffic situations cause high-stress levels.	Accepted
H3	Stress levels affect training performance.	Accepted

5. Discussion

In this study, the impact of stress levels on simulator-based maritime training was investigated through the analysis of biosignal data. The performance of participants was measured objectively through the number of position fixes made during the simulated voyage and data collected from the simulator tools. Additionally, machine learning (ML) algorithms were employed to identify the most effective methods for pre-processing biosignal data, extracting relevant features, and classifying stress levels. It was determined that the control and experimental scenarios resulted in different levels of stress for participants, which affected their performance. A summary of the findings is presented in Table 7.

Results obtained from the proposed performance criteria (presented in Table 6) revealed that overall scores between the two groups were similar, but there were notable differences in each individual term. For example, participants in the control scenario exhibited better positioning and maintained a greater distance from the first encountered vessel, while those in the experimental scenario demonstrated better overall deviation and a greater distance from the second encountered vessel.

This similarity in overall performance despite differing levels of stress can be explained by the concept of maximal adaptability, which states that human behavior has the ability to adapt within a certain range of stress such that performance remains stable. However, this approach is not sufficient in distinguishing specific differences in performance if a more precise assessment is desired. For instance, when participants were under a higher stress level, they may have found it difficult to take multiple positionings, leading to uncertainty in their location and a larger deviation from the planned route. Additionally, when under high stress, participants may have been more focused on the situation, resulting in a greater distance from encountered vessels.

In this study, the stress levels of participants were analyzed and their impact on training performance was evaluated through the classification of biosignal data and examination of deviation from the planned route. Results from the deviation measurements (shown in Fig 14) indicate that participants in the control scenario deviated towards the port side (left) of the route after Waypoint 3. This deviation may be attributed to the participants starting their turns too late or not turning back toward the planned route quickly enough. Additionally, the time taken for participants to return to the planned route after deviation was substantial, with some passing Waypoint 4 before returning to or nearing the planned route. This deviation also coincided with the point at which participants encountered the second fishing vessel, which resulted in a closer passing distance. Furthermore, the analysis of performance measures revealed that the participants were not intense.

On the other hand, in the experimental scenario, deviation from the planned route was larger in comparison to the control scenario. The time and distance taken to return to the planned route after the turn at Waypoint 3 were shorter than in the control scenario, which may be an indication that participants had more difficulty following the planned route due to uncertainties such as lower visibility in heavy fog or snow, and therefore practiced safer sailing. Overall, this study highlights the importance of analyzing deviation from planned routes in order to understand the impact of stress levels on training performance.

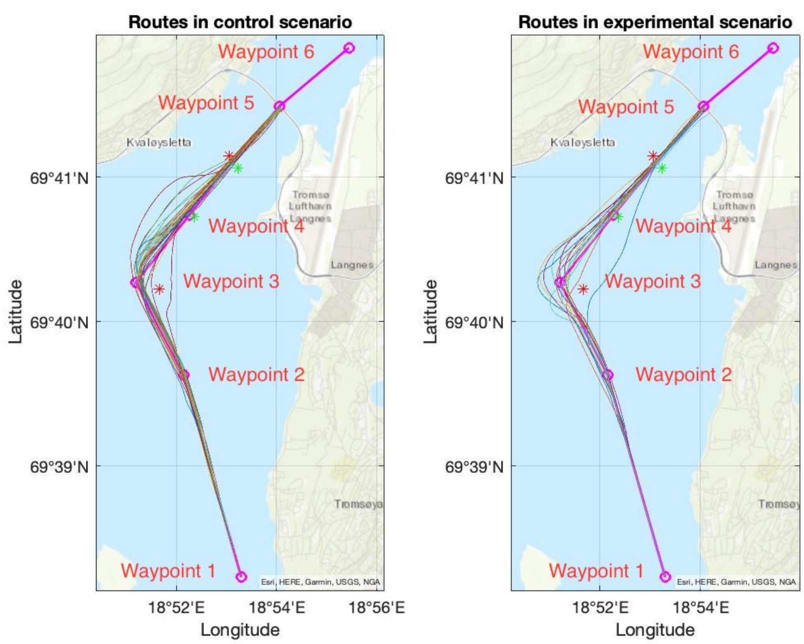


Figure 14: Participants sailed routes from the control scenario (left) and the experimental scenario (right). The magenta lines represent the planned route.

Results from the experimental scenario revealed that eight out of the fifteen participants deviated to the port side before making a significant course change (Waypoint 3) to starboard. This deviation may have been caused by heading loss (the event created in the experiment scenario), or the switching off of the autopilot to manual steering mode while the rudder was set at an angle. Hence, the switch happened to make the rudder turn the vessel to the wrong side. However, the proximity of this deviation to the loss of the gyro (also an event added in the experiment scenario) suggests that the participants may have made this decision based on the planned route. All the participants who turned to port first managed to get back to the planned route as fast or faster as those who did not take the wrong turn to port before turning to starboard. In the experimental scenario, there is more phenomenon to consider. Before the turn to starboard, the gyroscope error is induced and the participants experience the alarm. In the distance, the fog is also visible and may give uncertainty at the time. This may be a factor that focuses the participants' attention on following the planned route better in order to handle something unexpected later.

The experimental scenario was designed to be more intense cognitively than the control scenario. This is in line with real-world incidents, where most ship traffic accidents occur under fair weather conditions with good visibility [55] and fewer happened during night-time periods [56]. The present study's results suggest that under these conditions, seafarers may become more relaxed and less focused on their tasks, thus increasing the risk of accidents. In contrast, the experimental scenario in this study appears to have increased the participants' level of focus and attention to the task at hand.

The current study has contributed to the understanding of the relationship between stress levels and training performance in the maritime industry. Through the analysis of biosignal data and examination of deviation from planned routes, a correlation was observed between sailors' stress levels and route complexity. It was found that in many cases, deviation from the planned route exceeded 100 meters, which is not considered an unsafe level in a narrow water sailing task.

Given these findings, interventions can be made to improve the maritime training system by considering the impact of stress on performance. Instructors should be aware that low-stress levels may lead to overconfidence and delayed decision-making among students. Conversely, increasing stress levels may lead to heightened alertness and improved adherence to the planned route. However, it is important to note that under high stress, students may prioritize tasks differently and may be more prone to human errors as a wrong decision, erroneous action, missing action, or lack of action [57] might be a factor in the threat to maritime safety. Therefore, it is crucial for instructors to pay attention to the safety behaviors of students under different stress levels, rather than solely focusing on overall deviation from the planned route. It is always helpful for the instructors to have good control of the students' stress levels before and during the training, because the students who have less sea experience and are at the beginning of their education, may find it difficult to understand the situation and make the correct decision.

Additionally, it is essential to note that individuals may have varying perceptions of stress levels under the same training scenario. Thus, a flexible and adjustable training program, guided by objective stress level data, such as biosignal data, may be beneficial in achieving consistent learning outcomes while accounting for individual differences.

6. Conclusion and future work

Biosignal data-based training systems represent a novel approach to enhancing the performance and safety of maritime training by utilizing biosignals to measure trainees' stress levels during training sessions. These signals are then used to provide real-time feedback to trainees and instructors, enabling them to adjust the training program according to the trainee's stress level. One of the key benefits of biosignal data-based training systems is that they offer a more objective measure of stress levels compared to traditional self-report methods, thereby allowing for a more precise assessment of stress levels and enabling instructors to adapt the training program to better suit the needs of each trainee.

The present study analyzed questionnaire data using statistical methods and biosignal data using ML methods to investigate the impact of stress on training and performance in maritime navigation. The results of the study suggest that the stress levels of trainees are different under various training scenarios and that the complexity of the training scenarios can be classified based on the students' biosignal data. Additionally, different stress levels have specific effects on trainees' training performance, particularly in terms of safety behaviors. These findings provide a deeper understanding of the impact of stress on maritime training and performance, which can be used to improve the quality and effectiveness of maritime training programs and ultimately enhance safety at sea.

As shown in Figure 15, a reliable and safe maritime training system that utilizes biosignal data to measure trainee stress levels and provide real-time feedback is proposed in this study. This system aims to improve the performance and safety of maritime training by providing a more objective measure of stress levels, allowing instructors to adapt the training program to better suit the needs of each trainee. Furthermore, by providing real-time feedback, this system can help trainees to develop better stress management strategies and improve their overall performance, ultimately enhancing safety at sea. The process of implementing such a system, including the development of a real-time stress-level-detecting application and field testing in various scenarios with a sufficient amount of biosignal data, is left as future work. Additionally, it would be valuable to use the application to evaluate the assessment of SA and the training of decision-making in maritime contexts.

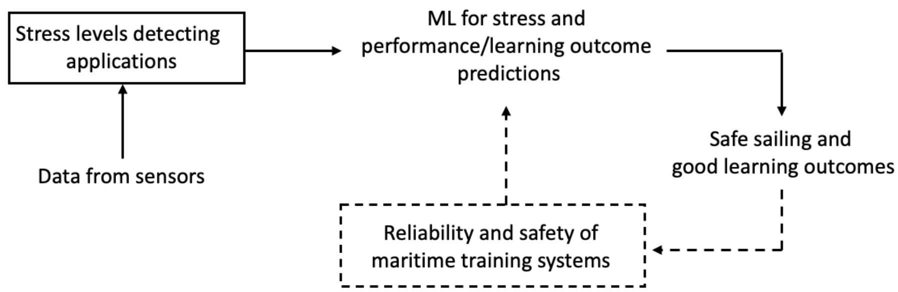


Figure 15: Future work for a reliable and safe maritime training system.

Acknowledgement

The authors extend their appreciation to Instructor Anders Johan Christensen, from the Nautical Science Group at UiT, The Arctic University of Norway, for his invaluable support and guidance throughout the study. Additionally, the authors acknowledge the participants for their participation and consent to use their data in this research. Their contributions were essential to the successful completion of this study.

Conflict of Interest

All authors declare that they have no conflicts of interest.

References

[1] R. Hanzu-Pazara, E. Barsan, P. Arsenie, L. Chiotoroiu, G. Raicu, Reducing of maritime accidents caused by human factors using simulators in training process, Journal of Maritime Research 5 (1) (2008) 3–18.

[2] R. Islam, H. Yu, R. Abbassi, V. Garaniya, F. Khan, Development of a monograph for human error likelihood assessment in marine operations, Safety science 91 (2017) 33–39.

[3] E. Akyuz, E. Celik, A modified human reliability analysis for cargo operation in single point mooring (spm) off-shore units, Applied Ocean Research 58 (2016) 11–20.

[4] C. Fan, K. Wróbel, J. Montewka, M. Gil, C. Wan, D. Zhang, A framework to identify factors influencing navigational risk for maritime autonomous surface ships, Ocean Engineering 202 (2020) 107188.

[5] C. Hetherington, R. Flin, K. Mearns, Safety in shipping: The human element, Journal of safety research 37 (4) (2006) 401–411.

[6] T. C. C. Trust, Chirp's independent, confidential incident and near-miss reporting programme improves safety at sea for mariners worldwide. we investigate every report and publish our anonymised findings to raise awareness of safety issues., https://chirp.srgry.com/maritime/what-we-do/, Last accessed on 2022-11-18 (2022).

[7] S. K. Basak, A framework on the factors affecting to implement maritime education and training system in educational institutions: A review of the literature, Procedia engineering 194 (2017) 345–350.

[8] L. Orlandi, B. Brooks, M. Bowles, The development of a shiphandling assessment tool (sat): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator, in: 15th Annual General Assembly of the International Association of Maritime Universities, IAMU AGA 2014-Looking Ahead: Innovation in Maritime Education, Training and Research, Australian Maritime College, 2014, pp. 131–140.

[9] S. Ghosh, M. Bowles, D. Ranmuthugala, B. Brooks, Reviewing seafarer assessment methods to determine the need for authentic assessment, Australian Journal of Maritime & Ocean Affairs 6 (1) (2014) 49–63.

[10] E. DEMİREL, D. BAYER, A study on the assessment of sea training as an integral part of maritime education and training, The Online Journal of Quality in Higher Education 3 (3) (2016) 12.

[11] Y. Liu, Z. Lan, J. Cui, G. Krishnan, O. Sourina, D. Konovessis, H. E. Ang, W. Mueller-Wittig, Psychophysiological evaluation of seafarers to improve training in maritime virtual simulator, Advanced Engineering Informatics 44 (2020) 101048.

[12] P. A. Hancock, A dynamic model of stress and sustained attention, Human factors 31 (5) (1989) 519–537.

[13] V. R. LeBlanc, The effects of acute stress on performance: implications for health professions education, Academic Medicine 84 (10) (2009) S25–S33.

[14] D. Wang, X. Wang, N. Xia, How safety-related stress affects workers' safety behavior: The moderating role of psychological capital, Safety science 103 (2018) 247–259.

[15] G. Makransky, S. Klingenberg, Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-weird sample, Journal of Computer Assisted Learning (2022).

[16] O. Jaeyong, S. Park, O.-S. Kwon, Advanced navigation aids system based on augmented reality, International Journal of e-Navigation and Maritime Economy 5 (2016) 21–31.

[17] F. Sanfilippo, A multi-sensor fusion framework for improving situational awareness in demanding maritime training, Reliability Engineering & System Safety 161 (2017) 12–24.

[18] R. S. Lazarus, Theory-based stress measurement, Psychological inquiry 1 (1) (1990) 3–13.

[19] I. Vlachos, A. Pantouvakis, M. Karakasnaki, Determinants and stressors of seafarers' job satisfaction: evidence from a global survey, Maritime Policy & Management (2022) 1–21.

[20] Y. Jiang, Z. Wan, J. Chen, Z. Wang, Knowledge mapping of seafarers' health research: a bibliometric analysis, Maritime Policy & Management (2021) 1–14.

[21] M. Hagmüller, E. Rank, G. Kubin, Evaluation of the human voice for indications of workload-induced stress in the aviation environment, EEC Note 18 (06) (2006).

[22] Y. Wang, Z. Lv, Y. Zheng, Automatic emotion perception using eye movement information for e-healthcare systems, Sensors 18 (9) (2018) 2826.

[23] I. cheol Jeong, S. hwan Jun, D. hee Lee, H. ro Yoon, Development of bio signal measurement system for vehicles, in: 2007 International Conference on Convergence Information Technology (ICCIT 2007), IEEE, 2007, pp. 1091–1096.

[24] J. Taelman, S. Vandeput, A. Spaepen, S. V. Huffel, Influence of mental stress on heart rate and heart rate variability, in: 4th European conference of the international federation for medical and biological engineering, Springer, 2009, pp. 1366–1369.

[25] R. Gevirtz, The promise of heart rate variability biofeedback: evidence-based applications., Biofeedback 41 (3) (2013).

[26] N. Munla, M. Khalil, A. Shahin, A. Mourad, Driver stress level detection using hrv analysis, in: 2015 international conference on advances in biomedical engineering (ICABME), IEEE, 2015, pp. 61–64.

[27] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, B.-H. Koo, Stress and heart rate variability: A meta-analysis and review of the literature, Psychiatry investigation 15 (3) (2018) 235.

[28] K. Herbell, J. A. Zauszniewski, Reducing psychological stress in peripartum women with heart rate variability biofeedback: a systematic review, Journal of Holistic Nursing 37 (3) (2019) 273–285.

[29] Z. Xu, J. H. Saleh, Machine learning for reliability engineering and safety applications: Review of current status and future opportunities, Reliability Engineering & System Safety 211 (2021) 107530.

[30] A. Liapis, E. Faliagka, C. Katsanos, C. Antonopoulos, N. Voros, Detection of subtle stress episodes during ux evaluation: Assessing the performance of the wesad bio-signals dataset, in: IFIP Conference on Human-Computer Interaction, Springer, 2021, pp. 238–247.

[31] R. L. Spitzer, K. Kroenke, J. B. Williams, P. H. Q. P. C. S. Group, P. H. Q. P. C. S. Group, et al., Validation and utility of a self-report version of prime-md: the phq primary care study, Jama 282 (18) (1999) 1737–1744.

[32] L. Manea, S. Gilbody, D. McMillan, Optimal cut-off score for diagnosing depression with the patient health questionnaire (phq-9): a meta-analysis, Cmaj 184 (3) (2012) E191–E196.

[33] B. Löwe, K. Kroenke, W. Herzog, K. Gräfe, Measuring depression outcome with a brief self-report instrument: sensitivity to change of the patient health questionnaire (phq-9), Journal of affective disorders 81 (1) (2004) 61–66.

[34] Norwegian Mapping and Cadastre Authority, Norwegian map from norwegian mapping and cadastre authority, https://www.norgeskart.no/#!/?zoom=10&lon=650694.32&lat=7736559.19&project=norgeskart&layers=1002, Last accessed on 2022-02-24 (2022).

[35] R. B. Nelsen, Heron's formula via proofs without words, The College Mathematics Journal 32 (4) (2001) 290.

[36] L.-z. Sang, X.-p. Yan, A. Wall, J. Wang, Z. Mao, Cpa calculation method based on ais position prediction, The journal of navigation 69 (6) (2016) 1409–1426.

[37] F.-X. Lesage, S. Berjot, F. Deschamps, Clinical stress assessment using a visual analogue scale, Occupational medicine 62 (8) (2012) 600–605.

[38] C. D. Spielberger, State-trait anxiety inventory for adults (1983).

[39] K. N. Fountoulakis, M. Papadopoulou, S. Kleanthous, A. Papadopoulou, V. Bizeli, I. Nimatoudis, A. Iacovides, G. S. Kaprinis, Reliability and psychometric properties of the greek translation of the state-trait anxiety inventory form y: preliminary data, Annals of General Psychiatry 5 (1) (2006) 1–10.

[40] D. Sharek, A useable, online nasa-tlx tool, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 55, SAGE Publications Sage CA: Los Angeles, CA, 2011, pp. 1375–1379.

[41] S. G. Hart, L. E. Staveland, Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Advances in psychology, Vol. 52, Elsevier, 1988, pp. 139–183.

[42] H. Xue, B.-M. Batalden, P. Sharma, J. A. Johansen, D. K. Prasad, Biosignal-based driving skill classification using machine learning: A case study of maritime navigation, Applied Sciences 11 (20) (2021) 9765.

[43] I. Daubechies, Cbms-nsf regional conference series in applied mathematics, Ten lectures on wavelets 61 (1992).

[44] A. N. Akansu, P. A. Haddad, R. A. Haddad, P. R. Haddad, Multiresolution signal decomposition: transforms, subbands, and wavelets, Academic press, 2001.

[45] S. Mallat, A wavelet tour of signal processing, Elsevier, 1999.

[46] P. Dickstein, J. Spelt, A. Sinclair, Application of a higher order crossing feature to non-destructive evaluation: A sample demonstration of sensitivity to the condition of adhesive joints, Ultrasonics 29 (5) (1991) 355–365.

[47] P. C. Petrantonakis, L. J. Hadjileontiadis, Emotion recognition from eeg using higher order crossings, IEEE Transactions on information Technology in Biomedicine 14 (2) (2009) 186–197.

[48] B. Kedem, Higher-order crossings in time series model identification, Technometrics 29 (2) (1987) 193–204.

[49] B. Kedem, S. Yakowitz, Time series analysis by higher order crossings, IEEE press New York, 1994.

[50] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera, Learning from imbalanced data sets, Vol. 10, Springer, 2018.

[51] H. He, Y. Ma, Imbalanced learning: foundations, algorithms, and applications (2013).

[52] Errata: Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association. 48 (264) (1953) 907.

[53] S. Glen, ”kruskal wallis h test: Definition, examples, assumptions, spss” from statisticshowto.com: Elementary statistics for the rest of us!, https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/kruskal-wallis/, Last accessed on 2022-03-16 (2022).

[54] M. Hollander, D. Wolfe, Nonparametric statistical methods 203-292, john wiley & sons, Inc. Toronto (1973).

[55] J. Weng, G. Li, Exploring shipping accident contributory factors using association rules, Journal of Transportation Safety & Security 11 (1) (2019) 36–57.

[56] J. Weng, D. Yang, Investigation of shipping accident injury severity and mortality, Accident Analysis & Prevention 76 (2015) 92–101.

[57] A. M. Rothblum, D. Wheal, S. Withington, S. A. Shappell, D. A. Wiegmann, W. Boehm, M. Chaderjian, Human factors in incident investigation and analysis, Tech. rep., COAST GUARD RESEARCH AND DEVELOPMENT CENTER GROTON CT (2002).