Transportation Research Interdisciplinary Perspectives 24 (2024) 101047

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip

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

ARTICLE INFO

MSC:
62R10
62H30
68T10

Keywords:
Maritime navigation
Biosignal
Machine learning
Stress
Simulation-based training

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.

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 (Hanzu-Pazara et al., 2008; Islam et al., 2017; Akyuz and Celik, 2016).

Traditionally, risk assessment in the maritime industry has been hindered by a lack of standardized accident reporting systems (Fan et al., 2020; Hetherington et al., 2006). However, with the advent of alerting and reporting systems for maritime incidents (The CHIRP Charitable Trust, 2022), 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 (Basak, 2017). 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 (Orlandi et al., 2014; Ghosh et al., 2014). Additionally, performance assessment is often evaluated subjectively by instructors, which can be unreliable, invalid, and unfair (Demirel and Bayer, 2016). Furthermore, studies have indicated that psychophysiological states such as cognitive workload and stress levels are key factors affecting performance (Liu et al., 2020). Therefore, monitoring stress levels and workload during assessments is crucial.

* Corresponding author.
E-mail address: hxu001@uit.no (H. Xue).

https://doi.org/10.1016/j.trip.2024.101047
Received 20 February 2023; Received in revised form 18 August 2023; Accepted 14 February 2024
Available online 17 February 2024

Abbreivations

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
IMO	International Maritime Organization
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
STCW	International Convention on Standards of Training Certification and Watchkeeping for Seafarers
SVM	Support vector machine
VAS	Psychometric evaluation of a visual analogue scale
VHF	Very High Frequency
VR	Virtual reality

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 (Hancock, 1989). 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 (LeBlanc, 2009). 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 (Wang et al., 2018b).

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 (Basak, 2017). 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 (Makransky and Klingenberg, 2022). 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 (Jaeyong et al., 2016). 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 (Sanfilippo, 2017).

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 (Lazarus, 1990; Vlachos et al., 2022). In the past, research has predominantly relied on subjective measurements, such as surveys and self-reported measures, as stress is difficult to measure objectively (Jiang et al., 2021). 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 (Hagmüller et al., 2006), and eye movements measured with an Electrooculogram (EOG) to identify different emotional states (Wang et al., 2018a). 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 (cheol Jeong et al., 2007).

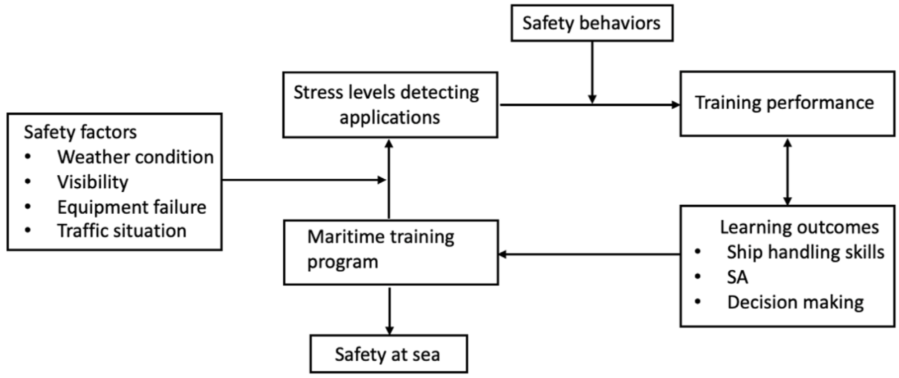

Fig. 1. Conceptual model of a stress-based maritime training program.

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 (Liu et al., 2020). 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 (Taelman et al., 2009; Gevirtz, 2013; Munla et al., 2015; Kim et al., 2018; Herbell and Zauszniewski, 2019).

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 (Xu and Saleh, 2021). 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 (Liapis et al., 2021). 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.

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) (Spitzer et al., 1999; Manea et al., 2012; Löwe et al., 2004) 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 determination of specific event designs in the comparative training scenarios and the selection of performance metrics were informed by the teaching program in UiT, which is the study plan based on the International Convention on Standards of

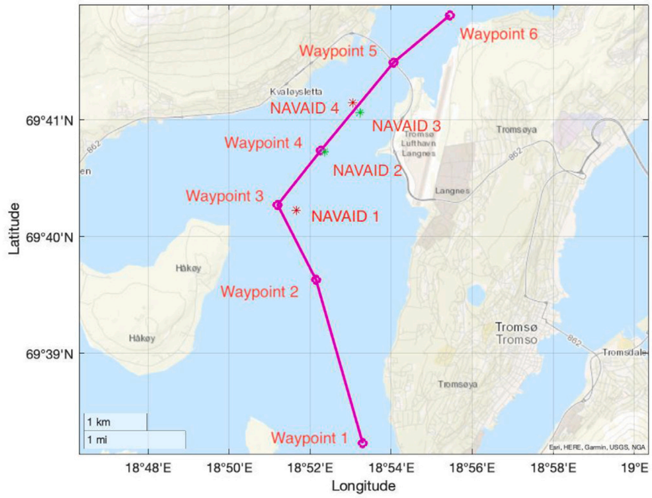

Fig. 2. The planned sailing route Sandnessundet consists of five straight legs. Waypoint 5 is located at the midpoint of the Sandnessund Bridge. The map of the route includes the placement of four navigational aids (NAVAIDs) which were distinguished by two distinct colors.

Training, Certification and Watchkeeping for Seafarers (STCW Convention) set by the International Maritime Organization (IMO) (IMO, 2018). The choice of a stressful scenario for maritime navigators was made based on training expert recommendations within the teaching program. 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 m, 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 min 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 km in length and is traversed by the Sandnessund bridge, connecting the Kvaløysletta district to the Tromsø city center, as described in Norgeskart (Norwegian Mapping and Cadastre Authority, 2022). This route is commonly used for navigational training for nautical students at UiT The Arctic University of Norway. The route, as depicted in Fig. 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 Fig. 3.

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) :

$$D = arccos[sin(LatA) * sin(LatB) + cos(LatA) * cos(LatB) * cos(LongA - LongB)] * 3440.1 * 1852 \tag{1}$$

  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 m.
  The distance between the sailing point and the planned route between two waypoints can be derived using Heron's formula (Nelsen, 2001).
- 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. Sang et al. (2016).

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 Fig. 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

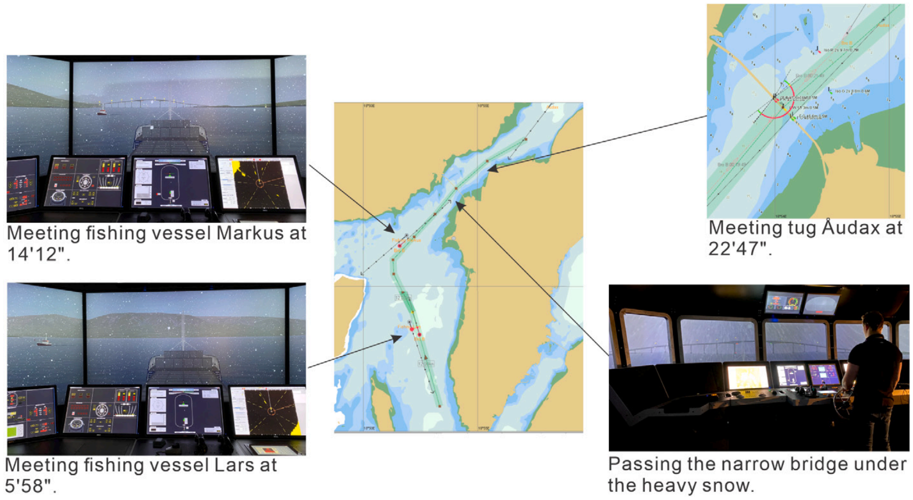

Fig. 3. An illustration of the sailing route of one of the participants, highlighting the geographical locations of the traffic situations.

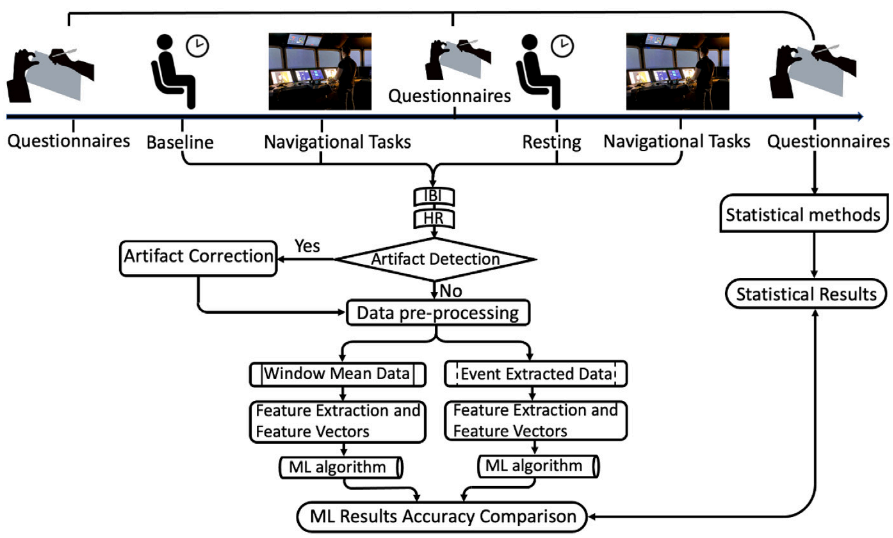

Fig. 4. Mixed-methods approach for stress level analysis in maritime training.

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

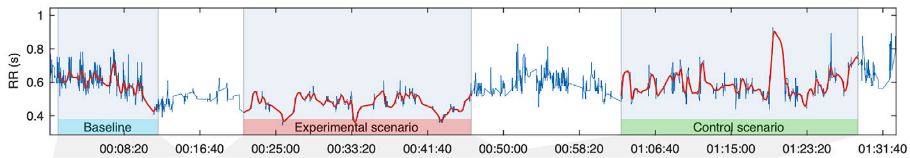

Fig. 5. Sample RR intervals with activity windows recorded from a participant during the sailing task are shown. Note that the gaps between windows represent the time between activities.

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 (Lesage et al., 2012).

State-trait anxiety inventory (stai) form y-1 (Spielberger, 1983). 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 (Fountoulakis et al., 2006). 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 (Sharek, 2011; Hart and Staveland, 1988). 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 (Xue et al., 2021).

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 Fig. 5. The instantaneous heart rate, measured in beats per minute (bpm), was derived from the IBI values using the following formula (Eq. (2)) :

$$HR[bpm] = 60/IBI \tag{2}$$

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 s and a step size of every 15 s.
- **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.

$$F_S = [\mu_{X_i}, \sigma_{X_i}], \qquad (i = 1, 2, \dots, l) \tag{3}$$

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) (Daubechies, 1992; Akansu et al., 2001). The wavelet coefficients were computed for specified scales (Mallat, 1999), 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 (Dickstein et al., 1991). 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 kth order difference is (see Eq. (5) (Petrantonakis and Hadjileontiadis, 2009)):
- From $\nabla^{k-1} Z_t$, a binary process $X_t^{(k)}$ was defined in Eq. (6) (Kedem, 1987; Petrantonakis and Hadjileontiadis, 2009; Kedem and Yakowitz, 1994) :

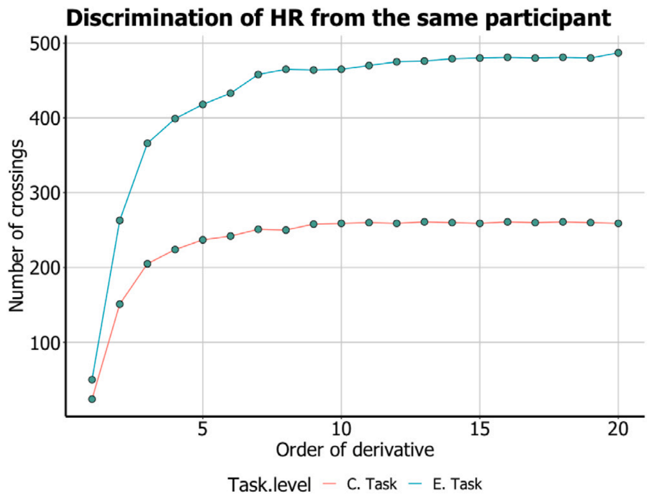

Fig. 6. Graphical comparison of HOC features from the same participant doing a different level of the task.

- The count of the symbol changes from $X_t^{(k)}$, D_k, was calculated in Eq. (7) (Petrantonakis and Hadjileontiadis, 2009; Kedem, 1987; Xue et al., 2021):

$$\nabla^{k-1} Z_t = \sum_{i=1}^{k} C_{i-1}^{k-1} (-1)^{i-1} Z_{t+1-i} \quad with \quad C_{i-1}^{k-1} = \frac{(k-1)!}{(i-1)!(k-i)!} \tag{5}$$

where $k = 1, 2, \ldots$, 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 $k = 1, 2, \ldots$.

$$D_k = \sum_{t=2}^{N} [X_t^{(k)} - X_{t-1}^{(k)}]^2 \tag{7}$$

where D_k is the count of symbol changes in kth 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 Fig. 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.

$$F_{HOC} = [D_1, D_2, \ldots, D_L], \qquad (1 < L < J) \tag{8}$$

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.

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 neighbor(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 (Fernández et al., 2018). Recall is typically used to measure the coverage of the minority class (He and Ma, 2013). The F-Score weights precision and recall equally (Fernández et al., 2018). The following equations, (9), (10), and (11), provide the definitions for these measures.

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$F\text{-}Score = \frac{2 * P * R}{P + R} \tag{11}$$

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 (Anon, 1953; Glen, 2022). 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 Fig. 7 and are presented in Table 3. The test statistic was calculated using Eq. (12) (Hollander and Wolfe, 1973) and the degrees of

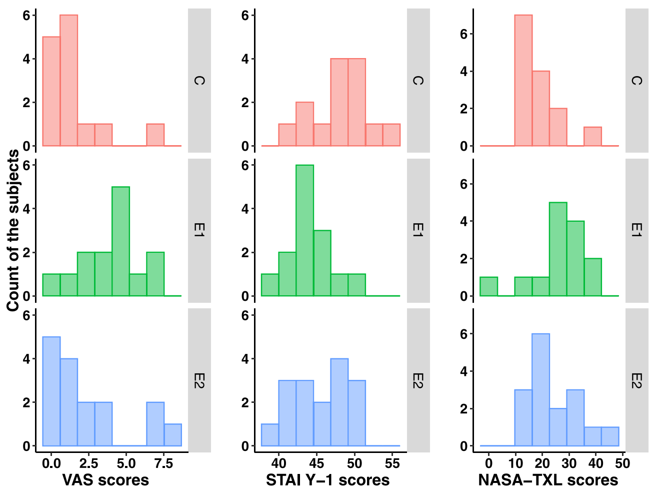

Fig. 7. Visualization of the results of the questionnaires from each group.

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.

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

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.

(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 Figs. 8 and 9. Fig. 8 illustrates the results from the three questionnaires as grouped by participant groups (C, E1, and E2) respectively. The results, presented in Figs. 8, indicate

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

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 experimental scenarios were analyzed. Fig. 9 compares the results of the questionnaires, as grouped by E1 and E2, respectively. The results, presented in Fig. 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.

$$S_{norm} = \sqrt{S} \tag{14}$$

where S is the data sample (scores of VAS of doing control task), S_{norm} is the normally distributed data sample.

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. Fig. 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 Fig. 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.

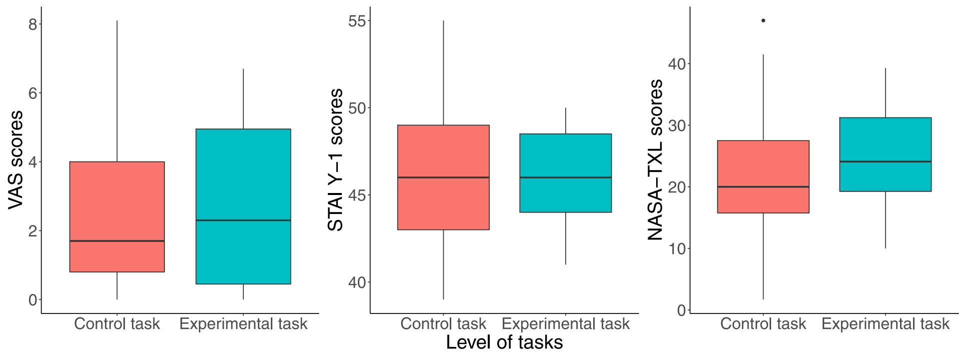

Fig. 8. Visualization of the results of the questionnaires from two different levels of scenarios.

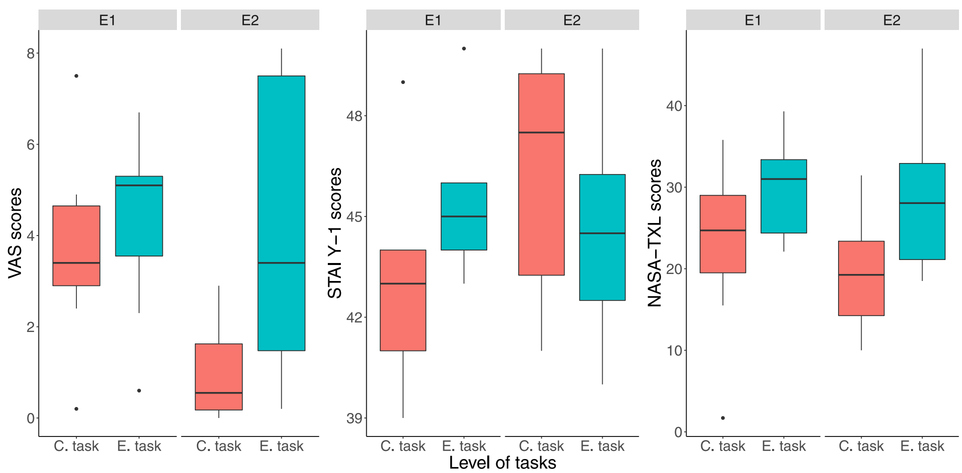

Fig. 9. Comparison of the questionnaire results from two different levels of the scenario by groups E1 and E2 (E1 and E2 have different orders of sailing the control scenario and experimental scenario).

Table 4
Welch t-test results of the questionnaires.

Questionnaires		Shapiro test normality (*p*-value)	Welch Two Sample t-test: Mean	SD	df	*t*-statistic	p-value	95% CI		Effect size Cohen's ds
VAS	C.Scenario	0.4548*	1.9069	1.9840	15.13	−4.2886	0.00063	−4.4993	−1.5133	1.0179
	E.Scenario	0.3376	4.1800	2.6622						(large)
STAI Form Y-1	C.Scenario	0.2674	46.3793	4.0037	38.101	1.3969	0.1705	−0.6493	3.5413	0.3968
	E.Scenario	0.5404	44.9333	2.7894						(small)
NASA-TLX	C.Scenario	0.4342	19.9862	8.5724	29.945	−3.5714	0.0012	−14.7398	−4.0144	1.1145
	E.Scenario	0.5140	29.3633	8.0871						(large)

* The star means the value was calculated after transforming the data to normal distribution.

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

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

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

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 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 m and a minimum score of 0 for deviations greater than 1000 m. 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. A Permutation Test was utilized to examine the mean disparity in total scores between the control and experimental scenarios. The observed mean difference was 1.11. Upon conducting 1000 permutations, the calculated *p*-value was 0.07 at a significance level of 0.1. This signifies a notable divergence in overall performance across the scenarios, prompting the rejection of the null hypothesis. These results suggest an evident performance advantage for participants in the control scenario. Due to the small sample size in our experiment, wider statistical tests may not be relevant or useful.

In addition to the performance criteria, an examination of the deviation from the planned route was conducted. The result, as illustrated in Fig. 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 Fig. 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 m away. Conversely, deviation in the experimental scenario was primarily greater than 200 m. 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.

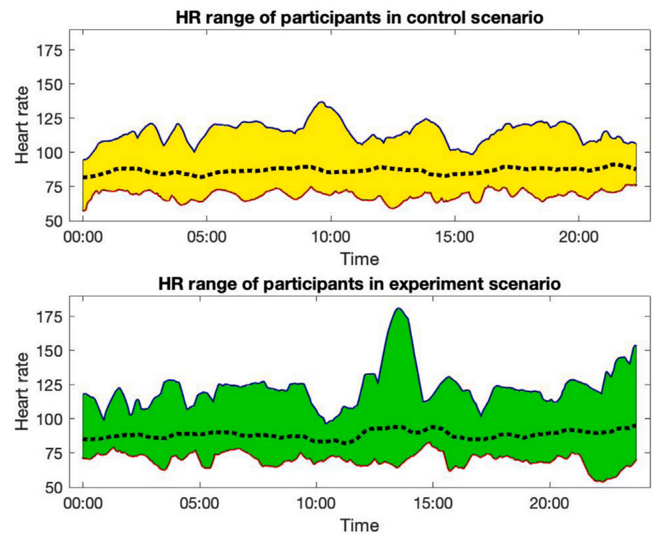

Fig. 10. Maximum and minimum HR of participants in the control scenario and experimental scenario. The dashed lines represent the mean HR.

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

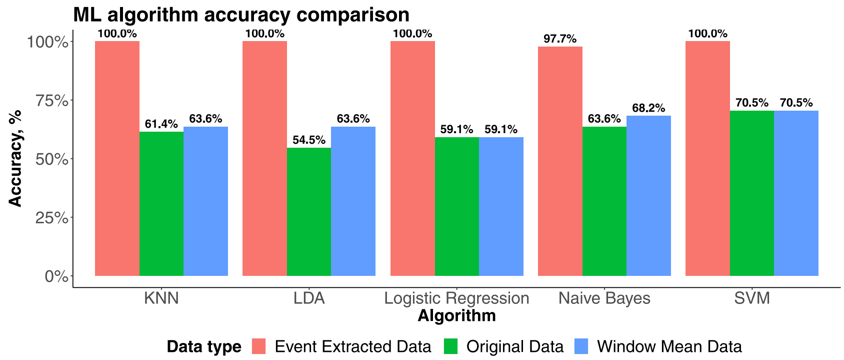

Fig. 11. Comparison of the machine learning results from five different algorithms in a different way of pre-processing data.

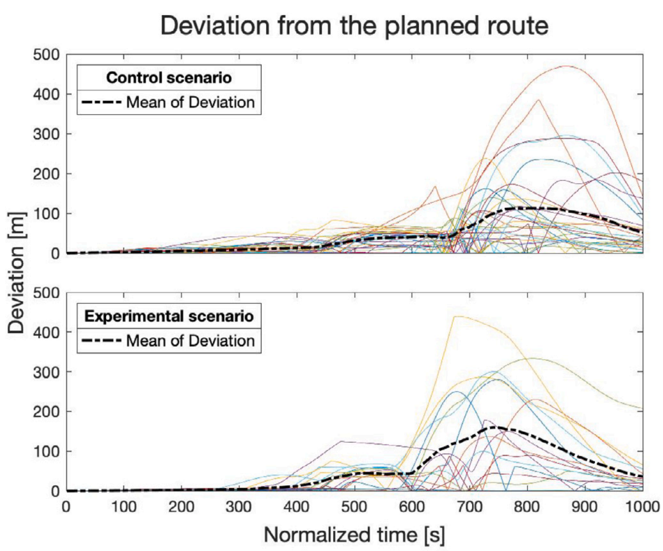

Fig. 12. Deviation from the designed route for the control scenario (upper) and the experimental scenario (lower). The black dashed lines represent the mean of the deviation.

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 towards 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.

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 (Weng and Li, 2019) and fewer happened during night-time periods (Weng and Yang, 2015). 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 m, 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 (Rothblum et al., 2002) 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.

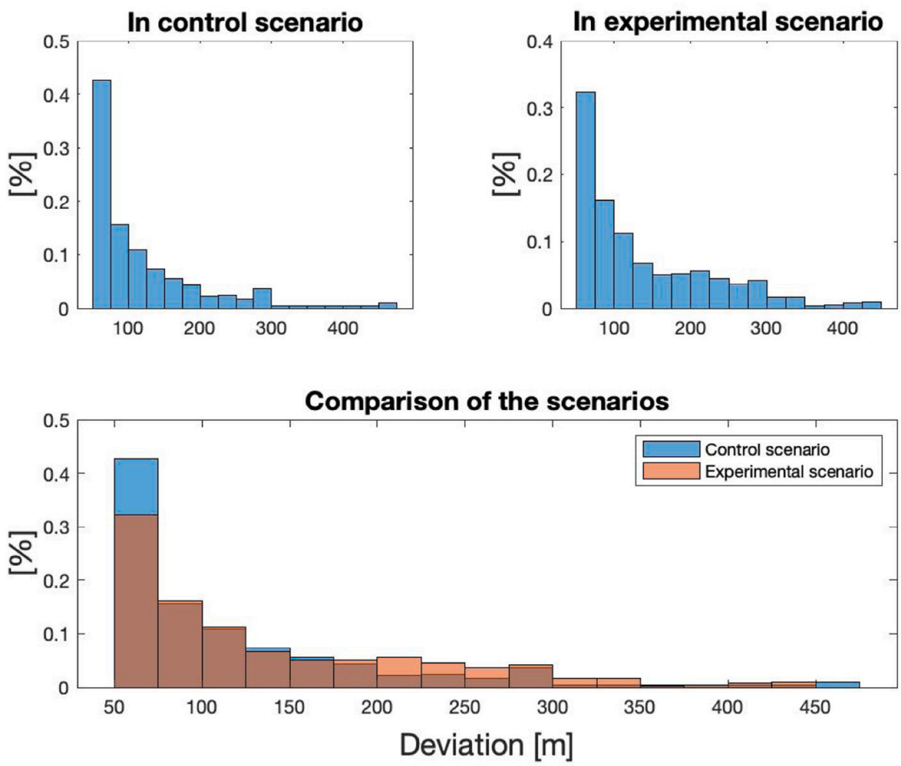

Fig. 13. Deviation from the planned route for the experimental scenario (orange) and the control scenario (blue) in the comparison graph (experts suggest that the deviation of shorter than 100 meters is negligible). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

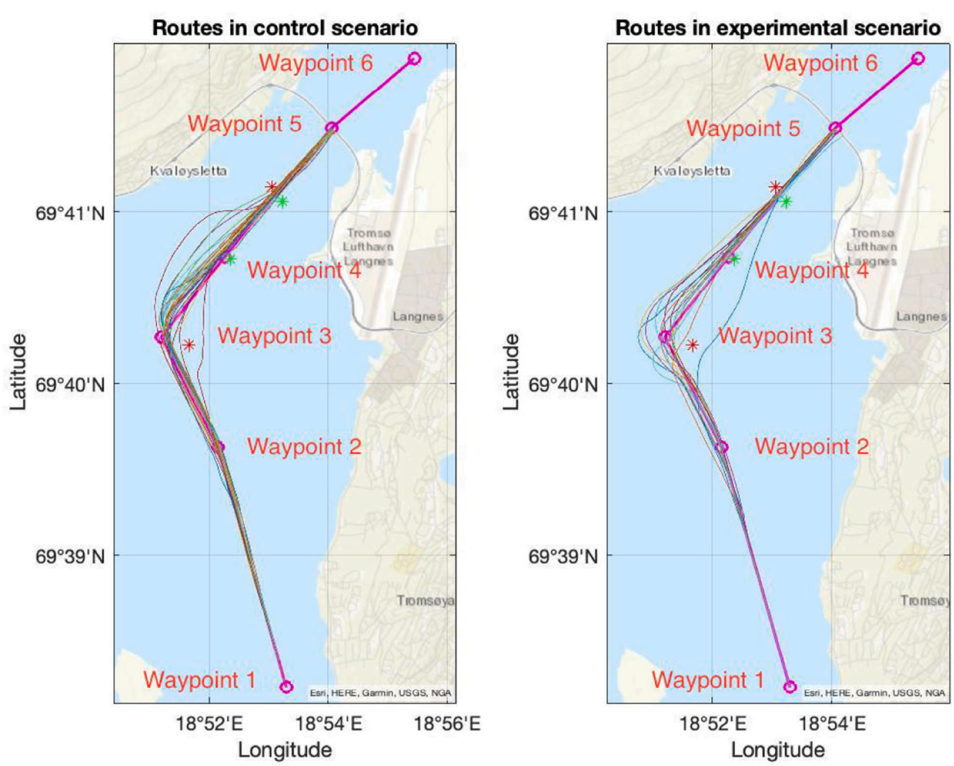

Fig. 14. Participants sailed routes from the control scenario (left) and the experimental scenario (right). The magenta lines represent the planned route. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

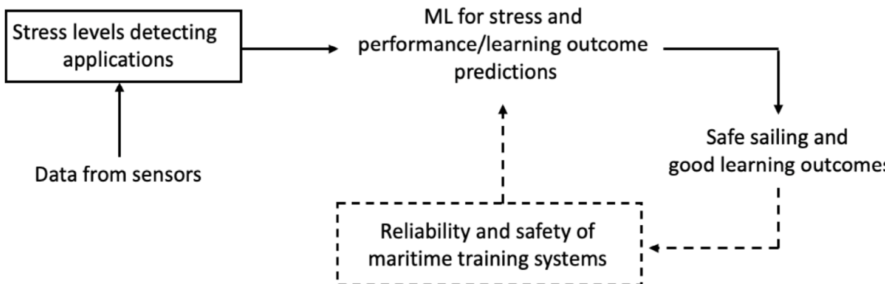

Fig. 15. A conception of a reliable and safe maritime training system.

In safety-critical domains, instructors often face the dilemma of balancing the need to train students to handle stressful situations, helping them cope with stress while avoiding overwhelming novices who are already grappling with complex simulator scenarios. Simulation training holds immense significance within the maritime training program, as it closely mirrors the challenges they will encounter in their future roles. During the simulator training tasks, instructors should be able to adjust the level of difficulty of the task for each student to get the best learning outcome. For instance, if a student's stress level becomes exceedingly high and they struggle to navigate the ongoing simulation scenario, the instructor can intervene to provide assistance or pause the training to modify the scenario accordingly. Therefore, it will be helpful to involve a reliable and safe maritime training system that utilizes biosignal data to measure trainee stress levels and provide real-time feedback, as shown in Fig. 15. This system will help to improve the performance and safety of maritime training by providing a more objective measure of stress levels. This will enable instructors to customize the training program to better suit the individual needs of each trainee, ultimately enhancing the effectiveness of the training by tailoring it to their unique capabilities and requirements. Furthermore, by providing real-time feedback, this system might help trainees to develop better stress management strategies and improve their overall performance, ultimately enhancing safety at sea. Additionally, it would be valuable to use the application to evaluate the assessment of SA and the training of decision-making in maritime contexts.

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.

For future work, the implementation process of the maritime training system remains a subject of further research. This includes the development of a real-time stress level-detecting application and conducting field tests across a range of scenarios with a sufficient amount of biosignal data. While the potential to attain real-time stress feedback based on our findings is evident, it is worth noting that achieving a comprehensive strategy for effectively managing trainees' stress is a goal that may require additional developments beyond the scope of this study.

CRediT authorship contribution statement

Hui Xue: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Øyvind Haugseggen:** Methodology, Investigation, Resources, Writing – review & editing. **Johan-Fredrik Røds:** Methodology, Investigation, Resources, Writing – review & editing. **Bjørn-Morten Batalden:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Dilip K. Prasad:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing, Supervision.

Declaration of competing interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangement), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Data availability

Data will be made available on request.

Acknowledgments

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.

References

Akansu, A.N., Haddad, P.A., Haddad, R.A., Haddad, P.R., 2001. Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets. Academic Press.
Akyuz, E., Celik, E., 2016. A modified human reliability analysis for cargo operation in single point mooring (SPM) off-shore units. Appl. Ocean Res. 58, 11–20.
Anon, 1953. Errata: Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 48 (264), 907.

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 Eng. 194, 345–350.
cheol Jeong, I., hwan Jun, S., hee Lee, D., ro Yoon, H., 2007. Development of bio signal measurement system for vehicles. In: 2007 International Conference on Convergence Information Technology. ICCIT 2007, IEEE, pp. 1091–1096.
Daubechies, I., 1992. CBMS-NSF regional conference series in applied mathematics. Ten Lect. Wavelets 61.
Demirel, E., Bayer, D., 2016. 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), 12.
Dickstein, P., Spelt, J., Sinclair, A., 1991. 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), 355–365.
Fan, C., Wróbel, K., Montewka, J., Gil, M., Wan, C., Zhang, D., 2020. A framework to identify factors influencing navigational risk for maritime autonomous surface ships. Ocean Eng. 202, 107188.
Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from Imbalanced Data Sets, vol. 10, Springer.
Fountoulakis, K.N., Papadopoulou, M., Kleanthous, S., Papadopoulou, A., Bizeli, V., Nimatoudis, I., Iacovides, A., Kaprinis, G.S., 2006. Reliability and psychometric properties of the Greek translation of the state-trait anxiety inventory form Y: Preliminary data. Ann. Gener. Psychiatry 5 (1), 1–10.
Gevirtz, R., 2013. The promise of heart rate variability biofeedback: evidence-based applications. Biofeedback 41 (3).
Ghosh, S., Bowles, M., Ranmuthugala, D., Brooks, B., 2014. Reviewing seafarer assessment methods to determine the need for authentic assessment. Aust. J. Marit. Ocean Affairs 6 (1), 49–63.
Glen, S., 2022. "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 16 March 2022).
Hagmüller, M., Rank, E., Kubin, G., 2006. Evaluation of the human voice for indications of workload-induced stress in the aviation environment. EEC Note 18 (06).
Hancock, P.A., 1989. A dynamic model of stress and sustained attention. Hum. Factors 31 (5), 519–537.
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. J. Marit. Res. 5 (1), 3–18.
Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: Advances in Psychology, vol. 52, Elsevier, pp. 139–183.
He, H., Ma, Y., 2013. Imbalanced Learning: Foundations, Algorithms, and Applications. John Wiley & Sons.
Herbell, K., Zauszniewski, J.A., 2019. Reducing psychological stress in peripartum women with heart rate variability biofeedback: A systematic review. J. Holist. Nurs. 37 (3), 273–285.
Hetherington, C., Flin, R., Mearns, K., 2006. Safety in shipping: The human element. J. Saf. Res. 37 (4), 401–411.
Hollander, M., Wolfe, D., 1973. Nonparametric Statistical Methods 203-292. John Wiley & Sons, Inc., Toronto.
IMO, 2018. STCW (Standards of Training, Certification, & Watchkeeping for Seafarers), fourth ed. International Maritime Organization, London.
Islam, R., Yu, H., Abbassi, R., Garaniya, V., Khan, F., 2017. Development of a monograph for human error likelihood assessment in marine operations. Saf. Sci. 91, 33–39.
Jaeyong, O., Park, S., Kwon, O.-S., 2016. Advanced navigation aids system based on augmented reality. Int. J. e-Navig. Marit. Econ. 5, 21–31.
Jiang, Y., Wan, Z., Chen, J., Wang, Z., 2021. Knowledge mapping of seafarers' health research: A bibliometric analysis. Marit. Policy Manag. 1–14.
Kedem, B., 1987. Higher-order crossings in time series model identification. Technometrics 29 (2), 193–204.
Kedem, B., Yakowitz, S., 1994. Time Series Analysis By Higher Order Crossings. IEEE press New York.
Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y.H., Koo, B.-H., 2018. Stress and heart rate variability: A meta-analysis and review of the literature. Psychiatry Investig. 15 (3), 235.
Lazarus, R.S., 1990. Theory-based stress measurement. Psychol. Inquiry 1 (1), 3–13.
LeBlanc, V.R., 2009. The effects of acute stress on performance: Implications for health professions education. Acad. Med. 84 (10), S25–S33.
Lesage, F.-X., Berjot, S., Deschamps, F., 2012. Clinical stress assessment using a visual analogue scale. Occup. Med. 62 (8), 600–605.
Liapis, A., Faliagka, E., Katsanos, C., Antonopoulos, C., Voros, N., 2021. 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, pp. 238–247.
Liu, Y., Lan, Z., Cui, J., Krishnan, G., Sourina, O., Konovessis, D., Ang, H.E., Mueller-Wittig, W., 2020. Psychophysiological evaluation of seafarers to improve training in maritime virtual simulator. Adv. Eng. Inform. 44, 101048.
Löwe, B., Kroenke, K., Herzog, W., Gräfe, K., 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the patient health questionnaire (PHQ-9). J. Affect. Disord. 81 (1), 61–66.
Makransky, G., Klingenberg, S., 2022. Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-WEIRD sample. J. Comput. Assist. Learn..
Mallat, S., 1999. A Wavelet Tour of Signal Processing. Elsevier.
Manea, L., Gilbody, S., McMillan, D., 2012. Optimal cut-off score for diagnosing depression with the patient health questionnaire (PHQ-9): A meta-analysis. Cmaj 184 (3), E191–E196.
Munla, N., Khalil, M., Shahin, A., Mourad, A., 2015. Driver stress level detection using HRV analysis. In: 2015 International Conference on Advances in Biomedical Engineering. ICABME, IEEE, pp. 61–64.
Nelsen, R.B., 2001. Heron's formula via proofs without words. College Math. J. 32 (4), 290.
Norwegian Mapping and Cadastre Authority, 2022. 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 14 February 2022).
Orlandi, L., Brooks, B., Bowles, M., 2014. 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, pp. 131–140.
Petrantonakis, P.C., Hadjileontiadis, L.J., 2009. Emotion recognition from EEG using higher order crossings. IEEE Trans. Inf. Technol. Biomed. 14 (2), 186–197.
Rothblum, A.M., Wheal, D., Withington, S., Shappell, S.A., Wiegmann, D.A., Boehm, W., Chaderjian, M., 2002. Human Factors in Incident Investigation and Analysis. Tech. Rep., COAST GUARD RESEARCH AND DEVELOPMENT CENTER GROTON CT.
Sanfilippo, F., 2017. A multi-sensor fusion framework for improving situational awareness in demanding maritime training. Reliab. Eng. Syst. Saf. 161, 12–24.
Sang, L.-z., Yan, X.-p., Wall, A., Wang, J., Mao, Z., 2016. CPA calculation method based on AIS position prediction. J. Navig. 69 (6), 1409–1426.
Sharek, D., 2011. A useable, online NASA-TLX tool. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, (no. 1), SAGE Publications Sage CA: Los Angeles, CA, pp. 1375–1379.
Spielberger, C.D., 1983. State-trait anxiety inventory for adults.
Spitzer, R.L., Kroenke, K., Williams, J.B., Patient Health Questionnaire Primary Care Study Group, Patient Health Questionnaire Primary Care Study Group, a.o., 1999. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. JAMA 282 (18), 1737–1744.
Taelman, J., Vandeput, S., Spaepen, A., Huffel, S.V., 2009. 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, pp. 1366–1369.
The CHIRP Charitable Trust, 2022. 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 18 November 2022).
Vlachos, I., Pantouvakis, A., Karakasnaki, M., 2022. Determinants and stressors of seafarers' job satisfaction: Evidence from a global survey. Marit. Policy Manag. 1–21.
Wang, Y., Lv, Z., Zheng, Y., 2018a. Automatic emotion perception using eye movement information for E-healthcare systems. Sensors 18 (9), 2826.
Wang, D., Wang, X., Xia, N., 2018b. How safety-related stress affects workers' safety behavior: The moderating role of psychological capital. Saf. Sci. 103, 247–259.
Weng, J., Li, G., 2019. Exploring shipping accident contributory factors using association rules. J. Transp. Saf. Secur. 11 (1), 36–57.
Weng, J., Yang, D., 2015. Investigation of shipping accident injury severity and mortality. Accid. Anal. Prev. 76, 92–101.
Xu, Z., Saleh, J.H., 2021. Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. Reliab. Eng. Syst. Saf. 211, 107530.
Xue, H., Batalden, B.-M., Sharma, P., Johansen, J.A., Prasad, D.K., 2021. Biosignal-based driving skill classification using machine learning: A case study of maritime navigation. Appl. Sci. 11 (20), 9765.