Developments in the Built Environment 15 (2023) 100198

Contents lists available at ScienceDirect

Developments in the Built Environment

journal homepage: www.sciencedirect.com/journal/developments-in-the-built-environment

Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: Incorporating electroencephalography, electrodermal activity, and video signals

Imran Mehmood[a,*], Heng Li[a], Waleed Umer[b], Aamir Arsalan[c], Shahnawaz Anwer[a,**], Mohammed Aquil Mirza[a], Jie Ma[a], Maxwell Fordjour Antwi-Afari[d]

[a] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region
[b] Department of Architecture and Built Environment, Northumbria University, NE1 8ST, Newcastle upon Tyne, United Kingdom
[c] Department of Software Engineering, Fatima Jinnah Women University, Rawalpindi, 46000, Pakistan
[d] Department of Civil Engineering, College of Engineering and Physical Sciences, Aston University, Birmingham, B4 7ET, United Kingdom

ARTICLE INFO

Keywords:
Mental fatigue
Construction safety
Construction equipment operators
Machine learning
Multimodal data

ABSTRACT

Construction equipment operations that require high levels of attention can cause mental fatigue, which can lead to inefficiencies and accidents. Previous studies classified mental fatigue using single-modal data with acceptable accuracy. However, mental fatigue is a multimodal problem, and no single modality is superior. Moreover, none of the previous studies in construction industry have investigated multimodal data fusion for classifying mental fatigue and whether such an approach would improve mental fatigue detection. This study proposes a novel approach using three machine learning models and multimodal data fusion to classify mental fatigue states. Electroencephalography, electrodermal activity, and video signals were acquired during an excavation operation, and the decision tree model using multimodal sensor data fusion outperformed other models with 96.2% accuracy and 96.175%–98.231% F1 scores. Multimodal sensor data fusion can aid in the development of a real-time system to classify mental fatigue and improve safety management at construction sites.

1. Introduction

Over 350 million people are employed by the construction industry worldwide, which has made significant contributions to the economic growth of many countries (Birhane et al., 2022). Regardless of their significance in boosting the economy, the health and safety on construction sites should not be underestimated (Jaafar et al., 2018). Owing to its poor safety performance, the construction industry is considered to be of the most hazardous industries (Khalid et al., 2021; Kines et al., 2010). It should also be noted, however, that the safety of workers is in a perilous state as they are three to six times more vulnerable to accidents in the construction industry than in other industries (Choi et al., 2020). Accidents occur frequently because of the unique and dynamic environment of construction projects (Koc and Gurgun, 2022), causing injuries and fatalities at construction sites (Sarkar et al., 2020). Among these, the equipment-related accidents are unarguably one of the most prevalent types of construction accidents and constitute a significant proportion (Li et al., 2021). For instance, according to construction industry statistics in the United Kingdom, “struck by moving equipment” accidents were the fourth leading cause of worker injuries (HSE, 2020). Furthermore, according to Vahdatikhaki et al. (2019), equipment is also a major cause of work-related fatalities and injuries in the United States construction industry. For this reason, it is imperative to eliminate equipment-related events at construction sites by addressing the underlying causes. One of the major contributors to these events is mental fatigue (Yang et al., 2021a), which is human behavior (Ma et al., 2021). The reason is that the equipment operations are cognitively demanding and require operators to maintain a significant level of sustained attentiveness (Li et al., 2020b), thus leading the operators to mental fatigue (Wagstaff and Sigstad Lie, 2011). Consequently, the operators’

* Corresponding author.
** Corresponding author.
E-mail addresses: imran.mehmood@connect.polyu.hk (I. Mehmood), heng.li@polyu.edu.hk (H. Li), waleed.umer@northumbria.ac.uk (W. Umer), aamir.arsalan@fjwu.edu.pk (A. Arsalan), shah-nawaz.anwer@polyu.edu.hk (S. Anwer), aquilmirza.mohammed@polyu.edu.hk (M.A. Mirza), jie2021.ma@connect.polyu.hk (J. Ma), m.antwiafari@aston.ac.uk (M.F. Antwi-Afari).

https://doi.org/10.1016/j.dibe.2023.100198
Received 19 April 2023; Received in revised form 17 June 2023; Accepted 11 July 2023
Available online 13 July 2023

ability to concentrate and make sound decisions is compromised (Das et al., 2020). This increases the likelihood of accidents involving the equipment being operated, which can lead to serious injuries or even fatalities at construction sites. Because of this, it is crucial to monitor the operators of construction equipment to ensure that they pay appropriate attention. To do this, construction equipment operators' mental fatigue levels must be tracked automatically. As a result, monitoring construction equipment operators for signs of attention deficit will be an efficient means of preventing accidents that cause property loss.

1.1. Mental fatigue monitoring and challenges in construction industry

It is challenging to directly quantify mental fatigue because it depends on a wide variety of factors, including the characteristics of the task, operator, and setting in which the work is being carried out (Hancock and Matthews, 2019; Li et al., 2018; Young et al., 2015). Previously, many researchers have attempted to monitor mental fatigue, for example, using questionnaires, physiological signals, and facial features. Initially, operators' mental fatigue was conventionally evaluated subjectively using questionnaires, with NASA-TLX being the most extensively used evaluation tool. However, it lacks precision because it is susceptible to individual bias (Han et al., 2019). This prompted a search for a more quantitative method. Consequently, researchers have been encouraged to establish objective measures of mental fatigue. In recent years, wearable sensors have attracted considerable interest from researchers because of the technological advancements that enable more objective monitoring of mental fatigue at construction sites. Therefore, research has been conducted to assess mental fatigue by studying physiological signals of workers. Examples include electroencephalography (Wang et al., 2023; Mehmood et al., 2023; Jeon and Cai, 2022; Ke et al., 2021a), electrodermal activity (Lee et al., 2021; Choi et al., 2019), eye-tracking (Noghabaei et al., 2021; Li et al., 2019b), and electrocardiography (ECG) (Umer, 2022). Compared to questionnaires, physiological indicators have better performance in terms of sensitivity, diagnostic ability, and non-intrusiveness (Zhao et al., 2018). It has been shown that physiological signals can be used to reliably identify worker fatigue because of their strong correlation with workers' mental fatigue states. However, the use of physiological technologies requires workers to wear sensors on their bodies, which hinders their routine work. Therefore, researchers have been particularly interested in and motivated towards a non-invasive method for detecting mental fatigue in construction site workers and operators. Recently, Mehmood et al. (2022) validated the geometric measurements of facial features and stated that they can be effectively utilized to detect mental fatigue in construction equipment operators on real construction sites. The findings of this study indicate a significant correlation between the geometric measurement of facial features and changes in brain activity during prolonged excavation operation. Moreover, Liu et al. (2021a) identified mental fatigue in crane operators using facial expressions in a stimulated environment.

1.2. Study aims and objective

Addressing mental fatigue in construction workers is a multifaceted challenge (Ding et al., 2020). This is due to the fact that the unregulated nature of the labor-intensive construction industry poses a significant threat to workers' well-being (Ojha et al., 2023). While previous studies in the construction sector have attempted to address this problem, recent advancements in wearable sensing technology have opened up new possibilities for the continuous and accurate monitoring of mental fatigue. However, determining the physiological indicator that yields the most reliable assessment of mental fatigue for workers at construction sites remains an important question to be answered by safety experts. Moreover, previous studies have assessed workers' mental states by individually investigating various physiological indicators. In contrast, Tao et al. (2019), Charles and Nixon (2019) and Young et al. (2015), have reported, no single approach has proven to be superior to others when it comes to assessing mental fatigue using physiological indicators. The same uncertainty applies to the geometric measurements of the facial features. Consequently, it remains unclear whether one physiological indicator is superior to another, or whether geometric measurements of facial features are more dependable than physiological indicators in determining a construction worker's level of mental fatigue. In light of this uncertainty, the objective of the current research is to investigate the feasibility of a multimodal data fusion approach to recognize mental fatigue in equipment operators during prolonged excavation operations, which has two major contributions.

First, the present study is the first to attempt to investigate a novel approach of integrating data from multiple sensors, such as electroencephalography, electrodermal activity, and geometric measurements of facial features, and machine learning techniques to classify various levels of mental fatigue. Although the concept of merging data streams from multiple sources may seem straightforward, combining data from different sensors has proven to be more accurate in predicting outcomes (Walambe et al., 2021). Each of the aforementioned unimodal measures has its strengths and limitations in terms of accuracy and suitability for detecting worker fatigue. Hence, the integration of multiple sensor data is intended to enhance mental fatigue recognition accuracy and reduce false warnings, facilitating comprehensive and holistic monitoring of mental fatigue. The literature also supports the effectiveness of combining data from multiple sensors to assess outcomes (Zhao et al., 2022). While research on multimodal approaches is ongoing in other industrial domains, studies investigating the classification of equipment operators' mental fatigue through the integration of multimodal sensor data, such as physiological indicators and facial features' geometric measurement, are scarce within the construction industry (Hu et al., 2023).

Second, the current study acquired multimodal data in a natural settings, which provided a more realistic and authentic perspective for research. This aspect is crucial, as it enhances the study's external validity, which refers to the extent to which the findings can be generalized. Previous investigations of mental fatigue have primarily relied on controlled laboratory settings, for instance by Liu et al. (2021a), Li et al. (2020b), and Li et al. (2019b). However, conducting experiments in laboratory environments presents challenges in terms of generalization and validity, mainly because they lack the dynamic nature and complexity of construction sites (Xing et al., 2020). To address this limitation, this study collected multimodal sensor data directly from construction equipment operators during on-site excavation operations. By capturing data in a realistic environment, the study's outcomes are more likely to reflect the complexities and nuances associated with mental fatigue in construction settings, and also hold practical relevance for understanding and managing mental fatigue among construction workers. Therefore, the current research is motivated by the need to learn more about how to recognize the mental fatigue of equipment operators holistically. This paper is organized as follows: Section 2 will provide a literature review on mental fatigue detection in the construction industry. Section 3 will discuss the methodologies adopted in this research. Section 4 discusses the findings and results for mental fatigue classification along with performance of different machine learning algorithms. Section 5, 6 and 7, provide discussion of findings, implication of current research, limitations and future research, respectively. Lastly, section 8 will present conclusions.

2. Related work

In this section, we provide a literature review of general fatigue detection in the construction industry, which can serve as a foundation for the implementation of a fatigue detection system based on multimodal sensor data. Considering that mental fatigue has such excruciating consequences, several researchers have previously attempted to detect its existence using multiple techniques, including (a) subjective assessment through questionnaires, e.g., NASA-TLX score, FAS, etc.; (b) physiological measures, e.g., electroencephalogram (EEG), electrodermal activity, electrocardiography, electrooculograms, electromyography, etc.; and (c) video signal-based facial feature detection, e.g., eye aspect ratio, mouth aspect ratio, and head motion.

2.1. Conventional assessment method

Historically, operators' mental fatigue has been measured using self-report questionnaires, with the NASA-TLX being the most widely utilized assessment tool (Li et al., 2019b). Such an assessment involves measuring mental fatigue using assessment scales that rely on subjective responses to a set of questions relating to mental states (Hart, 2006). These assessments usually lack accuracy and are prone to biased information (Han et al., 2019). Moreover, the time and effort required to answer questions carefully and precisely can disrupt ongoing work when a survey-based approach is used for continuous assessment (Hwang et al., 2018). Due to their subjective nature, fatigue questionnaires have scientific limitations in terms of reliability and construct validity (Techera et al., 2018). Consequently, this underlined the necessity for objective technologies that can continuously monitor and detect mental weariness without interfering with construction operations.

2.2. Physiological measures

In recent years, researchers have paid more attention to wearable sensors because of technical advancements that enable more objective monitoring of mental fatigue at construction sites. As a result, efforts have been made to detect mental fatigue by analyzing the physiological signals of workers. Using data from the worker's brain, eyes, muscles, and heart, physiological measurements monitor the worker's attention and can detect indications of fatigue before it negatively affects performance (Zhao et al., 2022; Doudou et al., 2020). In construction industry, several researchers have attempted to evaluate mental condition based on physiological information gathered through wearable sensors e.g., electrooculogram (Zhang and Etemad, 2021), electroencephalogram (Wang et al., 2019b, 2022; Lee and Lee, 2022; Jeon and Cai, 2022; Tehrani et al., 2021; Ke et al., 2021a, 2021b; Xing et al., 2019, 2020; Li et al., 2019a; Jebelli et al., 2018b, 2019a, 2019b; Hwang et al., 2018), electrocardiograph (Umer, 2022), eye-tracking (Noghabaei et al., 2021; Bitkina et al., 2021; Li et al., 2019b, 2020b; Han et al., 2020; Das et al., 2020; Jeelani et al., 2019; Hasanzadeh et al., 2018), and electrodermal activity (Lee et al., 2021; Choi et al., 2019; Jebelli et al., 2018a). It has been shown that physiological indicators can be employed to detect construction worker fatigue at construction sites because of their strong association with workers' mental states. However, Zhang et al. (2019) concluded that EEG is one of the fastest-growing technologies that researchers use to assess workers' cognitive and mental states under dynamic construction site conditions. Furthermore, Saedi et al. (2022) described it as a potent approach in the field of construction research because it measures brain activity quickly, cost-effectively, with high temporal resolution, and in a portable manner. The researchers assessed the mental state of construction workers by analyzing their captured brainwaves using statistical techniques and machine learning. For instance, Aryal et al. (2017) predicted the fatigue of construction workers with 82% accuracy using a boosted tree classifier. Previous studies have achieved sufficient results, for example, Chae et al. (2021) and Jeon and Cai (2022). In addition, physiological indicators demonstrate good individual performance. However, it is still an unsolved problem as to which physiological signal should be employed for a preeminent assessment of workers' mental fatigue states during construction operations.

2.3. Video based indicators

In this research area, workers' activities on construction sites are tracked through their facial videos, and their mental fatigue levels are subsequently detected by extracting useful features from their facial videos or images (Mehmood et al., 2022). When an individual feels fatigued, observable indications of fatigue can be identified by measuring their atypical behaviors (Zhao et al., 2022). Similarly, Cheng et al. (2019) concluded that studying a person's facial indications could provide information about their fatigue levels. In addition, Dziuda et al. (2021) observed that continuous analysis of face cues captured while performing activities allowed for effective and contactless detection of fatigue. Although this method has been thoroughly researched in other sectors of the industry, its application to the construction industry is still in its adolescence, with very few studies. Recently, Li et al. (2022) presented a decentralized deep learning solution to monitor operator fatigue without privacy exposure risks and reached an accuracy of approximately 86%. Furthermore, Liu et al. (2021a) proposed a combined deep-learning architecture and achieved an average accuracy of approximately 79%. Both researchers achieved acceptable accuracy. Mehmood et al. (2022) also conducted a construction site procedure on excavator operators and acquired facial videos as well as brain waves. Facial features were then extracted in the form of Euclidean distances, and temporal variations in the facial features were compared with the corresponding changes in brain activity. This study opens opportunities for future research in this field because the data were acquired from real construction. Despite these studies, it is still unclear whether video-based signals are better, worse or preferable than physiological measures. This is due to the fact that the quality of video-based measures is subject to or dependent on ambient and operator factors, as well as constraints including unbalanced lighting, camera angle, face pose, head movement, and personality factors e.g., size of an eye (Zhao et al., 2022; Zhu et al., 2021; Maior et al., 2020; Gromer et al., 2019; Jabbar et al., 2018). For instance, in a simulated setting, utilizing blink detection metrics, such as the % of eyelid closure over the pupil at a particular time, has been reported to have a high detection rate (You et al., 2019; Soares et al., 2019). However, as reported by Ji et al. (2019), the detection rate was significantly reduced when experiments were conducted in controlled settings. This was the case in the study by Liu et al. (2021a), in which data from crane operators were acquired in a simulated environment. Therefore, according to Doudou et al. (2020), video-based measurements are less precise and more unstable than physiological measurements are. However, researchers in the case of physiological measures are also not sure which physiological measure is preferred (Ding et al., 2020) to assess mental fatigue in construction workers.

2.4. Integration of physiological measures and facial indicators

Extant research on mental fatigue assessment in the construction industry primarily consists of studies that employ a single assessment modality, such as the studies by Mehmood et al. (2022), Tehrani et al. (2021) and Li et al. (2020b). Nonetheless, the literature suggests that identifying mental fatigue is a difficult task because of the multiple implicated factors (Ding et al., 2020). Therefore, researchers have paid increasing attention to the integration of data gathered from various sensors in recent years. Consequently, it has been extensively used in numerous fields, including medical applications for computer-assisted patient diagnosis based on a combination of different types of data (Cai et al., 2019). In comparison to traditional unimodal data analysis, it seems to perform better, according to previous studies (Zhu et al., 2020; Vidya et al., 2015). In addition, the integration of physiological indicators and facial features to complement each other's strengths for mental fatigue monitoring has been developed in recent years. For instance, a multimodal emotion recognition system based on a combination of facial movements and physiological measures was proposed by Wang et al. (2020). In addition, Utomo et al. (2019) introduced a fatigue prediction system that integrates heart rate variations and PERCLOS characteristics to efficiently detect fatigue. Although its usage is widespread in other fields (Yang et al., 2021b; Doudou et al., 2020; Wang et al., 2019c), its performance is better to that of single-modality analysis, and no preferential of one measure over others, scant prior research has focused on employing multimodal sensor data for monitoring mental fatigue in construction equipment operators. Therefore, this study proposes a machine learning-based multimodal analysis employing electroencephalography (EEG), electrodermal activity (EDA), and geometric measurement of facial features (FF) to recognize mental fatigue in construction equipment operators. Electroencephalography (EEG), electrodermal activity (EDA), and geometric measurements of facial features reflect fluctuations in mental fatigue in a distinct way and are affected by constant attention (Mehmood et al., 2022; Wan et al., 2021; Posada-Quintero et al., 2018; Giannakakis et al., 2017). Considering that mental fatigue monitoring is a complicated problem, integrating the data acquired from multiple modalities will assist in recognizing mental fatigue from multiple perspectives. Furthermore, we have proposed a geometrical measurement of facial features to be integrated with physiological features, rather than video measures such as blinking rate and PERCLOS. This is because the blink rate increases when a person is in a stressful situation (Giannakakis et al., 2017), whereas it tends to decrease with prolonged attention (Zhao et al., 2022), which is the case for construction equipment operators during construction operations. This could lead to an incorrect fatigue detection result. In summary, the contribution of this research is the implementation of multimodal integration using machine learning to recognize the mental fatigue levels of equipment operators at real construction sites, given three specific modalities: electroencephalography (EEG), electrodermal activity (EDA), and geometric measurement of facial features.

3. Methodology

Fig. 1 presents an outline of the research process, which details the proposed approach for detecting mental fatigue in construction equipment operators through the integration of physiological and facial feature data obtained from EEG, EDA sensors, and video cameras. The research methodology comprises of four distinct steps. The initial step entailed conducting an excavation operation at the construction site to gather pertinent data. This involved mounting a headband on the heads of construction equipment operators to capture EEG data, positioning an E4 watch on the wrists of operators to collect EDA data, mounting a video camera on the inside of the front screen of the excavator to capture facial feature data, and administering a questionnaire to elicit data related to subjective feelings of mental fatigue. In the second stage, data acquired from multiple sensors were analyzed, and mental fatigue levels were designated using subjective scores. The data was then subjected to artifact removal, and relevant features were extracted. The third stage involved the use of supervised machine-learning techniques to detect multiple levels of mental fatigue in construction equipment operators. Each machine learning technique was trained using features extracted from multiple sensors as input data. Finally, in the last step, the performance of each supervised machine learning technique was evaluated using metrics.

3.1. Experiment procedure and data collection

The experiment was conducted at a construction site to gather data on the mental fatigue of construction equipment operators, as shown in Fig. 2. The study was conducted at a construction site where a time-on-task approach was employed to induce mental fatigue in operators. Li et al. (2020b) and Morales et al. (2017) indicated that time-on-task is a common approach to induce mental fatigue. The experiment was conducted on multiple days, at the same time in the morning, in consistent weather, with clear skies on each day of data collection. It involves repetitive and time-consuming excavation and discharge tasks carried out by excavator operators over the course of an hour. The repetitive task was an excavation operation that involved excavating the ground and transporting excavated material from pits to vehicles. All excavator operators were subjected to the same conditions, which involved the continuous operation of the equipment in a cyclical manner. As this was a time-on-task experiment, the amount of earth excavated or moved, and the number of vehicles filled were not predetermined. In addition, no practice session was arranged for the operators because they already had prior experience with excavation operations. During the experiment, the excavator operators wore an E4 watch on their wrist and a headband-based wearable EEG device to record electrodermal activity and brain waves, respectively. Moreover, a video camera was attached to the excavator's windscreen to capture the operators' facial expressions while operating the equipment. The video footage was later converted into frames and analyzed to extract geometric measurements of the facial features. To evaluate the operators' mental fatigue levels, the NASA-TLX score was used, which was recorded every 20 min during the 1-h experiment. The collected data was then transferred to a desktop computer, where noise removal techniques were applied to eliminate any artifacts. The electrodermal activity, EEG data, and geometric measurements of facial features were labeled according to subjective measurements into three mental fatigue states: alert, mild fatigue, and fatigue (Prabaswari et al., 2019; Grier, 2015). The duration of the experiment was not disclosed to the operators to avoid reactivation of the end-spurt effect that could occur when participants realized the experiment is approaching towards the end.

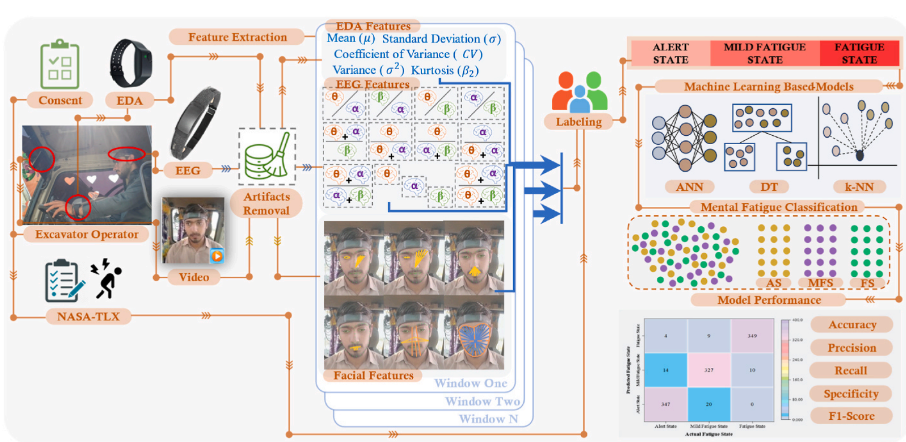

Fig. 1. Outline of research process.

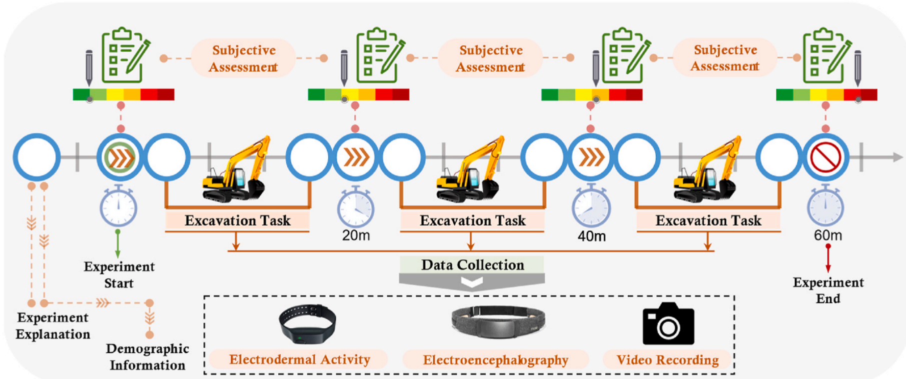

Fig. 2. Experimental design and procedure.

3.2. Participants

Sixteen male construction equipment operators with a mean age of 32.65 years (SD = 3.02) were recruited voluntarily to participate in this study. This study focused on excavator operators because excavation operation tasks, such as ground excavation and material transport, are repetitive, cognitively demanding, and often involve prolonged working hours that require operators to maintain sustained attention (Li et al., 2020b). All participants were experienced excavator operators with prior experience in excavator operations at construction sites, as shown in Table 1. The operators were required to report directly to the experiment on their designated days and were not involved in any other tasks or activities before the start of the experiment. Furthermore, we ensured that each operator remained fully engaged during the length of the task. They had slept for at least 8 h the previous night and abstained from alcoholic drinks for at least 24 h before the experiment. The experimental protocol was reviewed and approved by the ethics sub-committee of Hong Kong Polytechnic University (Reference Number: HSEARS20210927008) and conducted in accordance with the Declaration of Helsinki. Written consent was obtained from each participant after verbal explanation of the experimental procedures. Table 1 provides demographic information on the construction equipment operators who participated in this study.

3.3. Apparatus and measurement

3.3.1. Subjective assessment

The NASA-TLX score was used to evaluate the construction equipment operators' subjective feelings of mental fatigue and to provide a ground truth for their mental fatigue levels. Since its inception, the NASA-TLX has been widely utilized in numerous research studies, and its reliability and sensitivity have been established through a significant number of independent assessments. Moreover, a growing body of research has demonstrated that an increase in the NASA-TLX score over time during the same task can reliably indicate mental fatigue, as reported by Kaduk et al. (2021), Bitkina et al. (2021), Das et al. (2020), Li et al. (2019b), and Chen et al. (2017). Additionally, Mehmood et al. (2022), Li et al. (2020b), and Li et al. (2019b) utilized the increase in the NASA-TLX score for the same task as a subjective indicator of mental fatigue in construction equipment operators. In line with these findings, the current study considered a temporal increase in the NASA-TLX score to be a reliable indicator of an increase in mental fatigue.

Table 1
Demographic information of operators.

	Mean (Standard Deviation)	Range (Minimum-Maximum)
Height (cm)	171.47 (5.32)	15 (165–180)
Age (Years)	32.65 (3.02)	15 (26–41)
Weight (kg)	76.41 (7.66)	27 (65–92)
Job Experience (Years)	6.24 (3.49)	10 (2–12)
Body Mass Index (kg/m2)	25.96 (2.05)	7.61 (21.80–29.41)

3.3.2. Electroencephalogram (EEG) recording

In this study, EEG signals were acquired using the Muse headband, which is a flexible and user-friendly recording system. The Muse headband has four channels with dry electrodes positioned at the AF7, AF8, TP9, and TP10 sites, while the reference electrode FPz is located at the forehead position. The electrodes were made of silver, and the sampling rate of the Muse headband for the EEG signal acquisition was 256 Hz. The Muse headband was worn by all excavator operators during the excavation operation for an hour. The EEG data was transmitted in real time from the Muse headband to a smartphone via Bluetooth, where the "Mind Monitor" app was used to record the EEG signals. After recording, the data in the form of a comma-separated value file was transferred to a PC for further processing, as described by Mehmood et al. (2022) and Arsalan et al. (2019) and shown in Fig. 3.

3.3.3. Electrodermal activity (EDA) recording

The study utilized a photoplethysmography (PPG) wristwatch, specifically Empatica E4, to measure the electrodermal activity (EDA) in excavator operators to assess their mental fatigue. The Empatica E4 wristwatch includes four light-emitting diodes and four photoreceptors that automatically monitor the changes in the electrical properties of the skin to derive the EDA. The Empatica E4 watch was worn by all operators for an hour during the excavation operation. EDA data was collected in real-time and transmitted from the Empatica E4 to a smartphone via Bluetooth, where the "E4 Realtime" app was utilized to record the EDA signals. The recorded data was subsequently downloaded and transferred to a PC for further processing. The EDA datasheet includes a single column that indicates the EDA data in MicroSiemens sampled at 4 Hz. These methods are consistent with the approaches adopted by Milstein and Gordon (2020). Fig. 3 shows an example of an Empatica E4 PPG wristwatch.

3.3.4. Camera-based video signals

This study recorded the operators' facial behavior using a color video camera placed inside the equipment cabin. The camera was positioned on the interior side and was approximately 0.6 m away from the operator. The placement of the camera was carefully chosen so that it would not interfere with the operator's routine work. It was mounted on the windscreen of the equipment, with no chance of visual obstruction. The color video camera had a sampling frequency of 30 fps, capturing 24-bit RGB with three channels or 8-bit RGB per channel. It had a resolution of 1440 x 1440 pixels, providing an intricate view of the operator's facial behavior for the study.

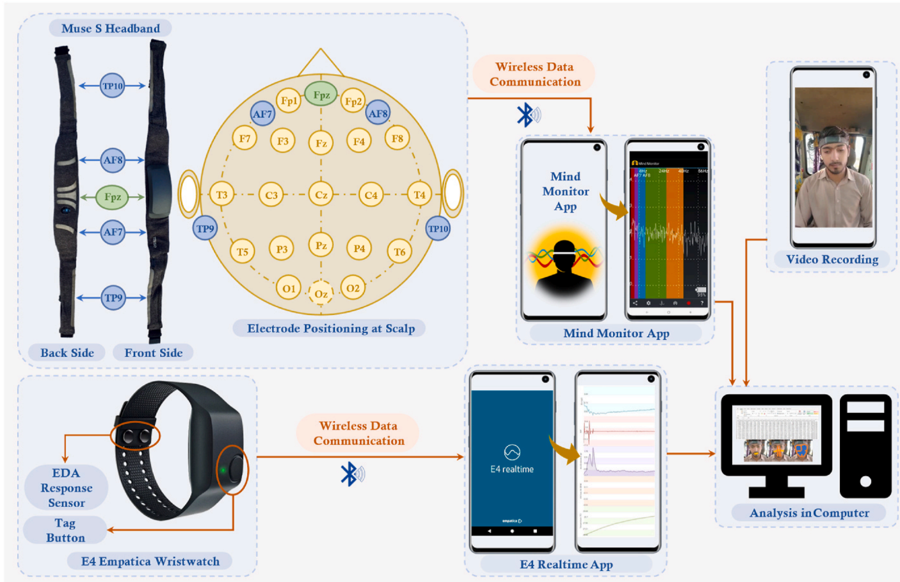

Fig. 3. Overview of apparatus utilized to collect and transfer the acquired data.

3.4. *Feature extraction*

3.4.1. *Electroencephalogram (EEG)*

In the current study, ten distinct EEG metrics from each channel, including theta, alpha, and beta, were computed and analyzed to evaluate and classify mental fatigue in construction equipment operators. The investigation did not include Delta and Gamma activities because they were not expected to exhibit any activity during the mental fatigue assessment. Previous studies, such as that of Eoh et al. (2005), have reported that delta activity corresponds to a person's sleeping state. Therefore, the current study concentrated on generating EEG metrics for the other three EEG bands, as an indication of mental fatigue. The process involved generating band ratios from EEG channels over time following the methodology used by Dasari et al. (2013) and Borghini et al. (2012). For example, the θ/α EEG metric was computed as the ratio of the average power spectral density value from the theta band with the average power spectral density value from the alpha band. Table 2 outlines all the computed EEG metrics used in this study.

3.4.2. *Geometric measurement of facial features*

When performing excavation operations at the construction site, all operators were video recorded for 1 h on the camera. OpenCV, a freely available open-source computer vision toolkit developed in Python, was initially utilized to convert the video footage of each operator into frames. Subsequently, face recognition was performed on each frame of the video recording using a locally constrained neural field model (Baltrušaitis et al., 2016). The operator's face was detected in each frame using this model, and the results were expressed as a vector $M = [l_1, l_2, l_3, \ldots\ldots\ldots, l_i]^F$, representing 68 landmarks identified on the operator's face in each frame via Dlib (King, 2009). In this case, l represents a detected facial landmark at position (x_i, y_i) in any frame F, F is the number of any frame, and i is the index of the detected landmarks at any frame, with values ranging from one to 68. Then, Eq. (1) was used to compute the Euclidean distance between any two desirable points. This Euclidean distance was used to compute the geometric measurements of eleven facial features investigated in this study (Mehmood et al., 2022). The proposed eleven facial features were retrieved separately from each individual frame and are described in Table 3 and presented in Fig. 4.

Table 2
Description of extracted EEG features.

EEG Metric	Previous Research
(i) θ, (ii) α, (iii) β,	Liu et al. (2021b), Li et al. (2020a), Jap et al. (2009)
(iv) θ/α, (v) $\beta/\ \alpha$, (vi) $\theta/\ \beta$, (vii) α/β	Raufi and Longo (2022), Dissanayake et al. (2022), Stancin et al. (2021), Fan et al. (2015), Jap et al. (2009), Eoh et al. (2005)
(viii) $(\theta+\alpha)/\ \beta$, (ix) $\theta/\ (\theta+\alpha)$, (x) $\alpha/(\theta+\alpha)$, (xi) $\theta/\ (\alpha+\beta)$	Dissanayake et al. (2022), Wu et al. (2021), Wang et al. (2019a), Fan et al. (2015), Eoh et al. (2005)
(xii) $(\theta+\alpha)/\ (\alpha+\beta)$, (xiii) $(\theta+\alpha)/(\theta+\beta)$	Mehmood et al. (2022), Stancin et al. (2021), Tyas et al. (2020),

3.4.3. *Electrodermal activity (EDA)*

Initially, EDA was separated into two components: tonic (EDL) and phasic (EDR). The former signifies differences in sympathetic arousal among individuals, while the latter represents the dynamic component of EDA, which reflects rapid changes in response to external stimuli (Greco et al., 2015; Braithwaite, 2013). In this research, we utilized the electrodermal response as a reliable indicator of mental fatigue. According to Poh et al. (2010), attention-demanding tasks can elicit electrodermal responses. Moreover, Collet et al. (2014) found that electrodermal response is a useful tool for detecting mental fatigue. Subsequently, five distinct features were extracted from the phasic component of electrodermal activity of each construction equipment operator. These features are mean (μ), standard deviation (σ), coefficient of variance (CV), variance (σ^2) and kurtosis (β_2). Kurtosis is a statistical measure describing the shape or peakedness of a probability distribution. It is typically measured using the standardized fourth moment of a distribution, which is the fourth central moment divided by the variance of the distribution. Similarly, variance is a statistical measure used to quantify the degree of variability or dispersion in a data sample, such as the electrodermal response of the operators.

3.5. *Artifacts removal*

Artifacts and unwanted fluctuations in data due to external sources are present in experimental data (Sweeney et al., 2012). Because of their potential for misinterpretation and skewness in analysis, these artifacts need to be cleaned from the data (Jebelli et al., 2018b; Hwang et al.,

Table 3
Description of extracted facial features.

Facial Feature	Description and Computation
Eye Area Average (EAA)	The average area of a closed polygon formed by joining the external landmarks on the eyes. $EAA = \sqrt{S_a[S_a - d(l_{37}, l_{38})][S_a - d(l_{37}, l_{42})][S_a - d(l_{38}, l_{42})]} + [(l_{38}, l_{39} + l_{42}, l_{41})/2][(l_{38}, l_{42} + l_{39}, l_{41})/2] + \sqrt{S_b[S_b - d(l_{39}, l_{41})][S_b - d(l_{39}, l_{40})][S_b - d(l_{41}, l_{40})]}$ $\therefore S_a = [d(l_{37}, l_{38}) + d(l_{38}, l_{42}) + d(l_{37}, l_{42})]/2$ $\therefore S_b = [d(l_{39}, l_{41}) + d(l_{39}, l_{40}) + d(l_{41}, l_{40})]/2$
Eye Distance Sum (SED)	The distance between the anchor and eye landmarks summed together. $SED = \|l_{31} - l_{43}\| + \|l_{31} - l_{44}\| + \|l_{31} - l_{45}\| + \|l_{31} - l_{46}\| + \|l_{31} - l_{47}\| + \|l_{31} - l_{48}\|$
Head Motion (HMO)	The computation of total distance between the anchor point and external landmarks of the face, per frame. $HMO = \frac{1}{Q}\sum_{i=1}^{F} \lvert l_{F1} - l_{F2} \rvert$
Eyebrow Sum (SEB)	The total distance between the anchor and eyebrow landmarks, computed as the sum of the Euclidean distances between corresponding points. $SEB = \|l_{31} - l_{18}\| + \|l_{31} - l_{19}\| + \|l_{31} - l_{20}\| + \|l_{31} - l_{21}\| + \|l_{31} - l_{22}\|$
Nose to Chin Ratio (NTC)	The distance from the anchor landmark to the chin. $NTC = \frac{2\|l_9 - l_{31}\|}{\|l_8 - l_{22}\| - \|l_{10} - l_{23}\|}$
Face Area (FAA)	The facial area enclosed by connecting the outermost landmarks on the face to form a closed polygon. $FAA = \frac{1}{Q}\sum_{i=1}^{N=27}(S(S - d(l_{31}, l_{12}))^2(S - d(l_{31}, l_{13}))^2(S - d(l_{12}, l_{13}))^2), \therefore S = \frac{d(l_{31}, l_{12}) + d(l_{31}, l_{13}) + d(l_{12}, l_{13})}{2}$
Eye Aspect Ratio (EAR)	The ratio of the height to the width of an operators' eye. $EAR = \frac{\|l_{44} - l_{48}\| + \|l_{45} - l_{47}\|}{2\|l_{43} - l_{46}\|}$
Mouth Corner (MCR)	The sum of distance between the anchor and mouth corner landmarks. $MCR = (\|l_{31} - l_{49}\| + \|l_{31} - l_{55}\|)$
Mouth Outer (MOR)	The total distance between the anchor landmark and the external landmarks, located around the mouth. $MOR = (\|l_{31} - l_{50}\| + \|l_{31} - l_{51}\| + \|l_{31} - l_{52}\| + \|l_{31} - l_{53}\| + \|l_{31} - l_{54}\| + \|l_{31} - l_{55}\| + \|l_{31} - l_{56}\| + \|l_{31} - l_{57}\| + \|l_{31} - l_{58}\| + \|l_{31} - l_{59}\| + \|l_{31} - l_{60}\| + \|l_{31} - l_{49}\|)$
Mouth Aspect Ratio (MAR)	The ratio of the height to the width of an operators' mouth. $MAR = \frac{\|l_{64} - l_{66}\| + \|l_{62} - l_{68}\| + \|l_{63} - l_{67}\|}{3\|l_{49} - l_{55}\|}$
Nose to Jaw Ratio (NTJ)	The distance from the anchor landmark to the jaws. $NTJ = \frac{\|l_3 - l_{31}\|}{\|l_3 - l_{15}\|}$

2018). In the construction industry, excavator operators are subjected to persistent and strenuous movements during excavation operations. These movements are caused by the vibrations of the equipment and the operator's movements as they track the bucket to excavate and deposit material (Mehmood et al., 2022). Unfortunately, these movements generate artifacts that must be eliminated from the collected data.

The study employed a Muse headband to acquire EEG data from construction-equipment operators. This device has its own on-board noise-cancellation mechanism, which is based on the statistical properties of the data, such as amplitude, variance, and kurtosis, to filter out the noise. If the statistical properties of an EEG signal exceed a predetermined threshold, the signal is deemed noisy and discarded, whereas if it falls below the threshold, the signal is considered clean (Cannard et al., 2021; Arsalan et al., 2019). Considering the constant movement of operators during excavation operations, the third-order one-dimensional median filter and the Savitzky-Golay (SG) filter (Orfanidis, 1995; Krauss et al., 1994) were further applied to the acquired EEG data for artifact removal. The principle of least-squares polynomial approximation is the foundation of the SG filter, making it a good choice for data smoothing (Savitzky and Golay, 1964). In the construction industry, Mehmood et al. (2022) and Aryal et al. (2017) used this noise cancellation method to smooth data while preserving the quality of EEG data.

In the current study, freely available MATLAB-based software, Ledalab, was used to obtain clean, scaled, and meaningful EDA data. EDA recording is susceptible to various forms of noise such as electrode noise and operator movement. To minimize the most common artifacts in EDA signals, a low-pass filter was applied (Taylor et al., 2015). A high-pass filter with a cut frequency of 0.5 Hz was also used to smooth the EDA signals (Braithwaite, 2013). However, large-magnitude artifacts, such as excessive electrode pressure and body motion, have not been adequately filtered by these methods (Taylor et al., 2015). To address this, a rolling filter was applied to the EDA signals with a rolling filter of 500 data points (Posada-Quintero and Chon, 2020), and the EDA was estimated every 500 ms in Micro Siemens.

Facial feature data of the construction equipment operators was carefully analyzed to eliminate artifacts. The process involves identifying stable facial regions during the extraction of features from every frame. Geometric measurements of facial features were then divided by the Euclidean distance of these stable regions to remove artifacts. A previous study in the construction industry by Mehmood et al. (2022) revealed that the length of the nose line, formed by connecting nose landmarks represented by the vector $D = [\|l_{32} - l_{28}\|]^F$, was effective in eliminating artifacts, as shown in Fig. 4(m). Specifically, the landmarks indicated by vector D were used to calculate the Euclidean distance of the nose line, as stated by the equation $d(l_{32},\ l_{28}) = \sqrt{(x_{32} - x_{28})^2 + (y_{32} - y_{28})^2}$. Subsequently, all facial features were normalized by dividing them by D, resulting in normalized facial features for each frame.

3.6. Machine learning-based mental fatigue classification

In this study, multiple sensor data points were integrated to classify mental fatigue in construction equipment operators using machine learning. This study utilized three types of input data: EEG, EDA, and geometric measurements of facial features. A wearable Muse headband at 256 Hz per second provided the EEG data, and a wearable E4 watch at 4 Hz per second acquired the EDA data. Similarly, geometric measurements of facial features were extracted from video recordings of equipment operators at a frequency of 30 fps. A sliding window approach was utilized with a window size segmentation of 16 s to split the multimodal data, and the overlapping of consecutive windows was then employed to ensure that no relevant data was missing. A 50% overlap of adjacent data segment lengths was used in this study (Liu et al., 2021c). Consequently, a dataset of 3,600 samples for 16 construction equipment operators was generated. In addition, this dataset was split into two parts, with 70% (2520 samples) designated for training and 30% (1080 samples) designated for testing. Subsequently, to accurately classify mental fatigue using data acquired from multiple sensors, we utilized three supervised machine learning classifiers: k-nearest neighbor (KNN), decision tree (DT), and artificial neural network (ANN). Although we cannot provide an in-depth introduction to these algorithms because of the length of this paper, the relevant machine learning literature can be consulted for more information (Umer et al., 2020; Aryal et al., 2017; Murphy, 2012; Witten and Frank, 2002). We chose these algorithms because prior research has demonstrated their efficacy in classifying mental fatigue. For instance, Ding et al. (2020) and Hu and Min (2018) compared various machine learning classifiers, including decision tree, k-nearest neighbor, support vector machine, and artificial neural network, for detecting fatigue in drivers. Considering these studies, three supervised machine-learning algorithms were trained on the acquired multimodal sensor data to classify mental fatigue in construction equipment operators.

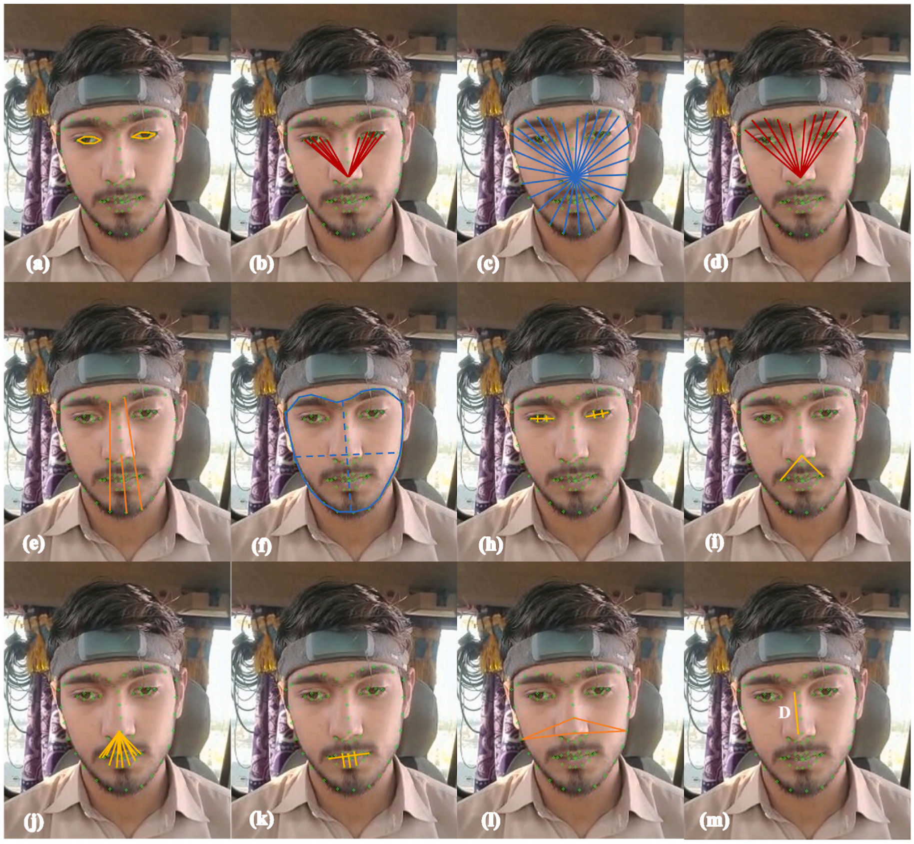

Fig. 4. Extraction of facial features; (a) eye area, (b) eye distance, (c) head motion, (d) eyebrow, (e) nose-to-chin ratio, (f) face area, (h) eye aspect ratio, (i) mouth corner (j) mouth outer, (k) mouth aspect ratio, (l) nose-to-jaw ratio, and (m) 68 landmarks detection.

3.7. Training and performance evaluation of machine learning models

To evaluate the accuracy of the models, we employed k-fold cross-validation, which involved dividing the original training set into k subsets. The value of k was set to 10 and each subset was approximately equal in size. The models were trained using k-1 subsets and validated using the remaining subset. By repeating this process for each subset, each sample was used to train and validate the models, allowing for a comprehensive assessment of their performance. This method ensures that the models are tested on a diverse range of data and minimizes the risk of overfitting (Antwi-Afari et al., 2023; Özdemir and Barshan, 2014). To evaluate the performance of the three machine learning models, we used accuracy, precision, recall, specificity, and the F1-score (Attal et al., 2015). Table 4 presents a detailed breakdown of each metric. Accuracy is the most commonly used metric for assessing classification performance across all classes. It is calculated as the ratio of instances that are correctly labeled to the total number of instances. Precision measures the rate at which positive cases are correctly identified, which is the ratio of positive instances correctly labeled to the total number of positive instances classified. Recall (sensitivity) is a measure of how accurately positive examples are identified and is defined as the percentage of all positive instances that were correctly classified. Specificity, on the other hand, measures the rate at which negative examples are correctly identified as negative and is calculated as the ratio of correctly identified false negatives to the total number of false negatives. Precision and recall are combined into the F1-score, which is used to evaluate the effectiveness of the classification model without introducing any systematic bias (Antwi-Afari et al., 2023). Additionally, we plotted the confusion matrix to evaluate each model's performance in specific classes, and accuracy and loss curves were used to determine the best-performing model. The confusion matrix displays the differences between the true labels of the data and model-generated labels. The elements on the diagonal represent correctly classified fatigue states, whereas those on the diagonal represent incorrectly classified fatigue states.

Table 4
Performance assessment metrics for machine learning models.

Performance metric	Equation
Accuracy	$\left((TN+TP)/_{(TN+TP+FN+FP)}\right)X\,100$
Precision	$\left((TP)/_{(FP+TP)}\right)X\,100$
Recall	$\left((TP)/_{(FN+TP)}\right)X\,100$
Specificity	$\left((TN)/_{(FP+TN)}\right)X\,100$
F1-Score	$\left(2\,x\,\frac{Recall\;x\;Precision}{Recall+Precision}\right)$

In this study, an orange data mining tool, Python-based open-source software (Version 3.33.0, Bioinformatics Lab, the University of Ljubljana, Slovenia), was used to compare and assess various classification algorithms (Demšar et al., 2013). The canvas interface of the orange software enables users to design data analysis workflows by dragging and dropping widgets, which perform various functions such as reading data, displaying tables, selecting features, training predictors, contrasting learning methods, and visualizing data items. Additionally, users can interact with the program to examine visuals and transfer them to other widgets (Kukasvadiya and Divecha, 2017; Naik and Samant, 2016).

4. Experimental results

4.1. Analysis of ground truth data

In this study, the NASA-TLX score was used as a reliable measure to identify mental fatigue states. The findings presented in Table 5 demonstrate the descriptive statistics derived from the ground-truth analysis. Notably, subjective mental fatigue was significantly higher at the end of the experiment than at the beginning, exhibiting an increase from 11.25 (SD = 2.77) to 65.25 (SD = 4.85). Additionally, the results listed in Table 5 indicate that the operators experienced progressively higher levels of mental fatigue as the excavation operation continued.

4.2. Machine learning-based classification results for multimodal data

This study utilized a novel approach to identify and classify mental fatigue states in construction equipment operators by integrating input data from multiple sensors and employing machine learning techniques. Three machine learning models (ANN, k-NN, and DT) were used to classify mental fatigue into alert, mild, and fatigue states. In addition to the EEG data, the input data included electrodermal activity (EDA) and geometric measurements of facial features. The data was fused as inputs for the machine-learning models used in the study. Furthermore, input data from multiple sensors was fused into various combinations, including (a) EEG and EDA, (b) EEG and FF, (c) EDA and FF, and (d) EEG, EDA, and FF. The results, as shown in Tables 5–7, indicated that the machine learning models achieved classification accuracies ranging from 56.5% to 97.1%. However, the decision tree models achieved the highest accuracies for all input data combinations, ranging from 85.0% to 97.1%. The findings of this study indicate that the decision models outperformed the other machine learning models investigated in terms of accuracy when trained on input data from multiple sensors of operators over three increasingly demanding phases of work.

4.2.1. Neural network (NN)

The evaluation metrics and confusion matrix presented in Table 6 and Fig. 5 indicate the performance of an Artificial Neural Network (ANN) model for identifying different levels of mental fatigue in construction equipment operators. Overall, the evaluation metrics demonstrated the good performance of the model for different input data fusions. However, the performance of the model was slightly lower than that of Decision Tree (DT) models. The ANN model achieved an accuracy ranging from 73.5% to 96.6% for all input data combinations, with the highest accuracy of 96.6% achieved using FF and EDA as the input data. The model's classification performance ranged from 93.96% to 98.347% in terms of precision, with FS and MDS representing the highest values of the correctly identified fatigue levels. Additionally, higher recall and precision indicate that the model yielded fewer false negatives and false positives, respectively. Specificity and F1-score measures ranged between 96.783% and 99.168%, and 95.979% and 98.892%, respectively. In addition, a confusion matrix was utilized to determine whether the classes were misclassified or confused with others. As demonstrated in Fig. 5, the high values of the diagonal elements imply that the model correctly distinguished between the three mental fatigue classifications. The other cells indicated incidents that were incorrectly classified. Alert states were more often misclassified than mild and fatigue states. It was confused with MFS in 23 instances, as shown in Fig. 5(a). However, the misclassification rate remains remarkably low compared with the number of classified instances. In addition, the results were similar for the combination of all three sensor datasets (EEG, EDA, and FF), with an overall classification accuracy of 94.7%. MFS was the most misclassified state, being confused with AS and FS in 14 and 10 instances, respectively. Moreover, the combination of EEG and FF exhibited slightly less accuracy than the above-mentioned two combinations, with an overall accuracy of 87.8%. The fourth combination, with EEG and EDA as input data, attained the lowest overall accuracy (73.8%) among the four

Table 5
Ground truth of mental fatigue.

	Baseline	Mental Fatigue States		
		Alert State	Mild Fatigue State	Fatigue State
Subjective Assessment NASA-TLX Score (0–100)	11.25 (2.77)	30.81 (2.99)	45.00 (4.27)	65.25 (4.85)

Table 6
Performance assessment metrics for ANN models.

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
FF-EDA				
Accuracy	96.6	97.222	96.667	99.259
Precision		93.963	97.619	98.347
Recall		98.082	92.134	99.443
Specificity		96.783	98.895	99.168
F1-score		95.979	94.798	98.892
EEG-EDA				
Accuracy	73.8	85.093	78.519	83.981
Precision		76.289	67.514	77.515
Recall		81.096	67.135	72.981
Specificity		87.133	84.116	89.459
F1-score		78.619	67.324	75.179
EEG-FF				
Accuracy	87.8	93.981	88.426	93.148
Precision		91.667	84.894	86.632
Recall		90.411	78.932	93.871
Specificity		95.804	93.093	92.788
F1-score		91.034	81.805	90.107
EEG-EDA-FF				
Accuracy	94.7	96.481	95.093	97.870
Precision		94.550	93.162	96.409
Recall		95.068	91.854	97.214
Specificity		97.203	96.685	98.197
F1-score		94.809	92.504	96.810

Table 7
Performance assessment metrics for k-NN models.

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
FF-EDA				
Accuracy	94.4	97.037	94.537	97.130
Precision		92.802	96.552	94.086
Recall		98.904	86.517	97.493
Specificity		96.084	98.481	96.949
F1-score		95.756	91.259	95.759
EEG-EDA				
Accuracy	56.5	67.778	70.000	75.185
Precision		51.731	56.061	64.000
Recall		69.589	41.573	57.939
Specificity		66.853	83.978	83.773
F1-score		59.346	47.742	60.819
EEG-FF				
Accuracy	87.5	93.889	88.241	92.873
Precision		86.553	87.055	88.950
Recall		96.986	75.562	89.694
Specificity		92.308	94.475	94.452
F1-score		91.473	80.902	89.320
EEG-EDA-FF				
Accuracy	85.8	92.685	86.389	92.593
Precision		84.048	85.185	88.430
Recall		96.712	71.067	89.415
Specificity		90.629	93.923	94.175
F1-score		89.936	77.489	88.920

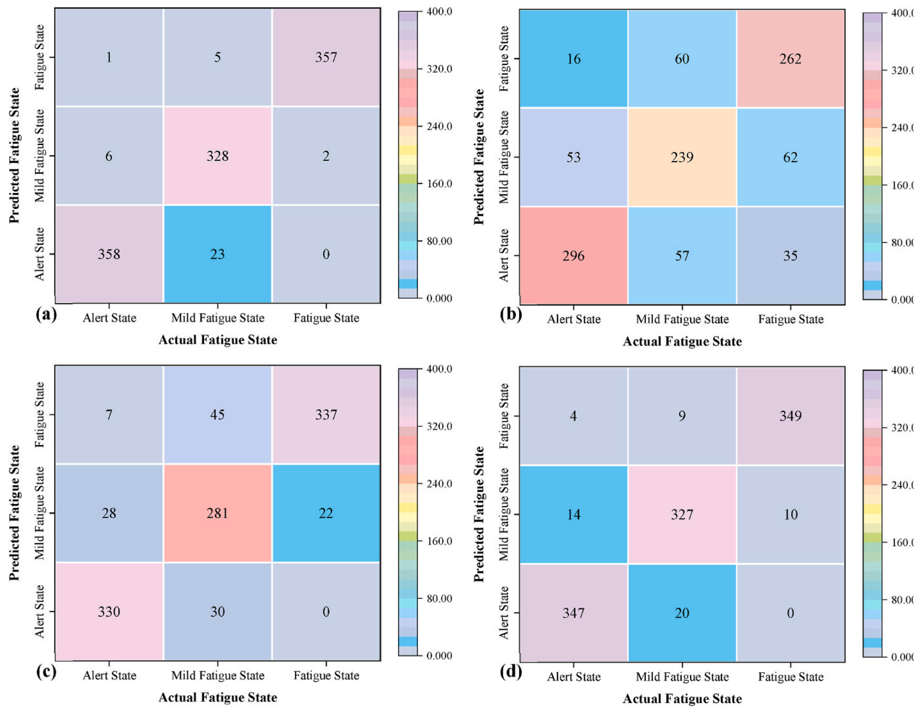

Fig. 5. Confusion matrix (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF.

combinations. This combination also exhibited the highest number of misclassified classes among all combinations, with MFS being confused with AS and FS in 53 and 62 instances, respectively.

4.2.2. K-nearest neighbors (kNN)

Table 7 presents the evaluation matrix, and Fig. 6 shows the confusion matrix of the k-nearest neighbor model. When used on all possible combinations of input data, k-NN performed inferior to ANN and DT in determining different levels of mental fatigue in construction equipment operators. Nonetheless, the overall accuracy, except for one combination of input data, was greater than 80%. The k-NN model attained overall performance accuracy values ranging from 56.5% to 94.4% for all input data combinations. The model attained an overall accuracy of 94.4% when FF and EDA were employed as the input data. Consequently, the MFS indicated higher instances of correctly identified fatigue levels with a precision of slightly above 96.5%. However, AS and FS exhibited slightly less effect on the k-NN model with a precision of 92.802% and 94.086%, respectively, as shown in Table 2. The model

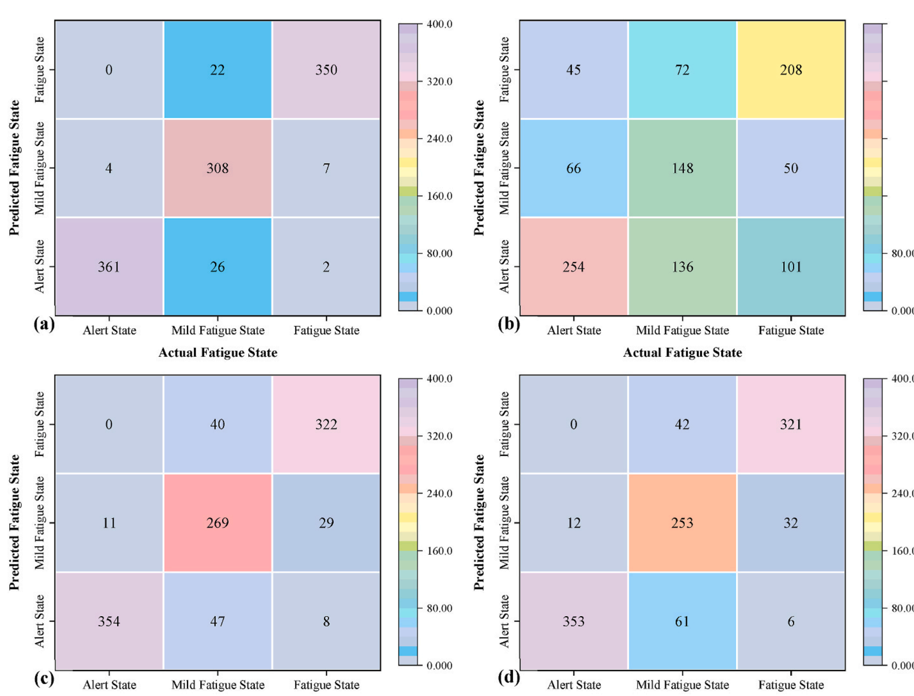

Fig. 6. Confusion matrix (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF.

attained the highest values of precision and recall for the aforementioned combination, indicating that it yielded fewer false positives and negatives. Similarly, specificity and F1-score measures had values ranging between 96.084% and 98.481% and 91.259% and 95.759%, respectively, indicating that an operator identified as being in a particular fatigue state was, in fact, in that fatigue state. Furthermore, using EEG and FF as input data, the model achieved an overall accuracy of 87.5%, with classification precision values ranging from 86.553% to 88.950%. Interestingly, the model attained higher specificity values ranging between 92.308% and 94.452%, indicating that the operator who identified any fatigue state actually experienced that state. Similarly, a comparable overall accuracy of 85.8% was achieved when the input data from all the three sensors were combined. Consequently, the classification performance in terms of precision was between 84.048% and 88.430%. According to the confusion matrix in Fig. 6, it can be observed that the confusion among the mental fatigue states was modest except for the combination of EEG and EDA. The misclassification rate for this combination is exceptionally high, as shown in Fig. 6(b). Notably, AS and FS were the most recognized states, as shown in Fig. 6 (a), (c), and 6(d). AS was recognized in 361 (FF and EDA), 354 (EEG and FF), and 353 (EEG, EDA, and FF) positive cases, respectively. Furthermore, when we see the confusion matrix demonstrated in Fig. 6(b), AS was confused 136 and 101 times with MFS and FS, respectively.

4.2.3. Decision tree (DT)

Table 8 and Fig. 7 present the evaluation metrics and confusion matrix for the decision tree (DT) model, which includes the correct classifications displayed in the diagonal cells for a more detailed evaluation of the classification performance. Compared to the ANN and k-NN models, the DT model achieved the highest overall accuracy, ranging between 85.0% and 97.1% for all input data combinations. It is important to note that using EEG and EDA as input data resulted in an accuracy of 85.0%, whereas all other input data combinations achieved an accuracy above 96.0%. When using data from all sensors as inputs, AS had the most accurately classified instances at 97.568%. In contrast, FS had the lowest percentage of accurately classified instances compared to AS and MFS (94.370%). Additionally, the model produced a high number of false negatives and false positives for FS compared to other fatigue stages, with 21 instances of confusion with MFS. However, this confusion number is modest compared with other combinations of input data. The specificity and F1-score measures ranged between 97.087% and 98.741%, and 94.084% and 98.231%, respectively, indicating that an operator identified as being in a particular fatigue state was indeed in that fatigue state. Moreover, the FF and EDA, and EEG and FF input data combinations also resulted in higher instances of correctly identified fatigue levels, with modest confusion among mental fatigue states. However, using EEG and EDA as input data resulted in higher confusion among mental fatigue states. Nonetheless, the confusion among the states was still modest compared to the findings indicated by the ANN and k-NN models, as shown in Tables 6 and 7, respectively. Fig. 7(a)–(d) show that AS and FS were recognized with 364 and 352 (FF and EDA), 331 and 294 (EEG and FF), 363 and 347 (EEG and FF), and 361 and 352 (EEG, EDA, and FF) positive instances, respectively.

Table 8
Performance assessment metrics for DT models.

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
FF-EDA				
Accuracy	96.9	99.538	96.944	97.407
Precision		98.913	97.640	94.370
Recall		99.726	92.978	98.050
Specificity		99.441	98.895	97.087
F1-score		99.318	95.252	96.175
EEG-EDA				
Accuracy	85.0	92.037	88.426	89.537
Precision		86.423	82.535	85.965
Recall		90.685	82.303	81.894
Specificity		92.727	91.436	93.343
F1-score		88.503	82.419	83.880
EEG-FF				
Accuracy	97.1	98.796	97.778	97.685
Precision		97.059	97.977	96.389
Recall		99.452	95.225	96.657
Specificity		98.462	99.033	98.197
F1-score		98.241	96.581	96.523
EEG-EDA-FF				
Accuracy	96.2	98.796	96.204	97.407
Precision		97.568	96.736	94.370
Recall		98.904	91.573	98.050
Specificity		98.741	98.481	97.087
F1-score		98.231	94.084	96.175

5. Discussion

Construction equipment operations demand a high level of attention and cognitive effort from the operators. These operations are complex and often require multitasking, quick decision making, and precise control. Operators should remain alert and focused for extended periods, which can lead to mental fatigue. Mental fatigue is a state of reduced mental performance that results from prolonged cognitive activity. This can impair an operator's ability to perform tasks, react to stimuli, and make decisions, which increases the risk of accidents and injuries on construction sites. Therefore, it is crucial to noninvasively monitor the mental fatigue of construction equipment operators to minimize equipment-related incidents and ensure safe working conditions. Previous studies used a single-modal data approach to detect and classify mental fatigue. However, it is unclear which physiological measure is the best indicator of mental fatigue. Thus, the objective of this study is to evaluate a new approach that uses machine learning and multimodal sensor data collected from equipment operators to recognize and classify different types of mental fatigue states during equipment operation. Three types of data from operators-electroencephalography, electrodermal activity, and geometric measurement of facial features-were gathered during an onsite operation on actual construction sites. The study then compared the performance of three types of machine learning models, artificial neural networks (ANN), k-nearest neighbors (k-NN), and decision trees (DT), for training the input data collected from multiple sensors. Furthermore, this study is the first to propose a machine learning-based approach for recognizing and classifying mental fatigue states, including alert, mild fatigue, and fatigue states, in construction equipment operators under sustained attention by integrating of multiple sensor data. The results showed that mental fatigue can be accurately classified in construction equipment operators with varying levels of mental fatigue, that is, alert, mild fatigue, and fatigue states, while integrating the acquired data from multiple sensors.

5.1. Multimodal data integration and machine learning-based models

In the current study, the performances of the three machine learning models were compared, and it was found that the decision tree (DT) model outperformed the other two models, with an overall accuracy ranging from 85.0% to 97.1% when using different combinations of input data. The precision, recall specificity, and F1-score of the DT model ranged from 94.370% to 97.568%, 91.573%–98.904%, 97.087%–98.741%, and 94.084%–98.231%, respectively, when integrating data from all sensors. The other two input combinations, EDA and FF, and EEG and FF, also showed high values of assessment metrics, with overall accuracies of 96.9% and 97.1%, respectively. Based on the analysis of the confusion matrix, the alert state (AS), mild fatigue state (MFS), and fatigue state (FS) had a relatively small number of instances that were misclassified. For example, when using a combination of FF and EDA as inputs, the number of misclassified instances for AS, MFS,

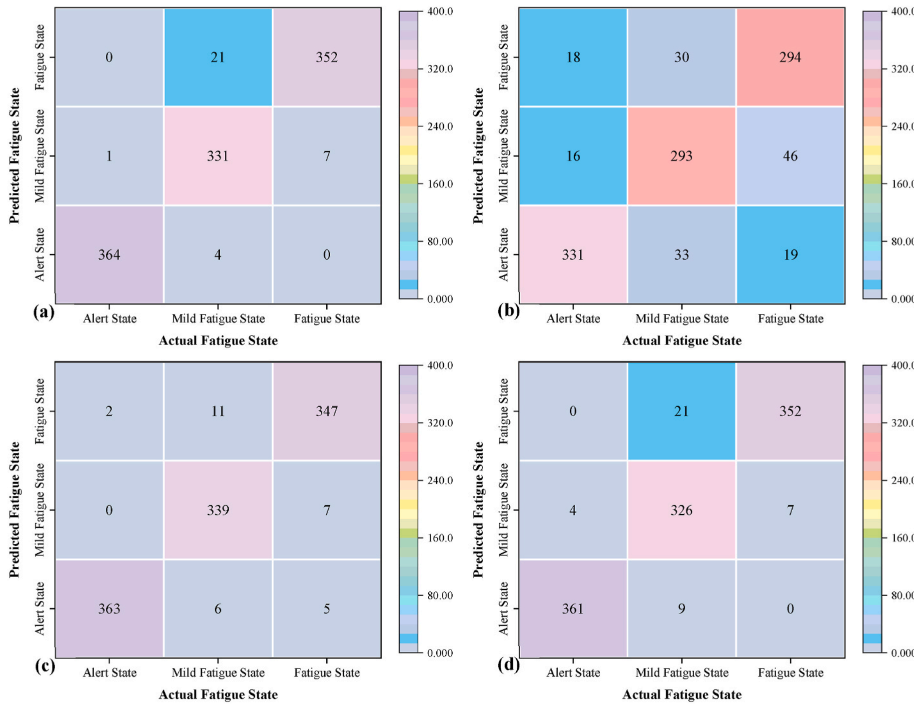

Fig. 7. Confusion matrix (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF.

and FS were 4, 8, and 21, respectively. When EEG and FF were used as inputs, the misclassified instances for AS, MFS, and FS were 11, 7, and 13, respectively. Similarly, when EEG, EDA, and FF were used as inputs, the misclassified instances for AS, MFS, and FS were 9, 11, and 21, respectively. However, the confusion matrix revealed that the EEG and EDA combination had a larger number of misclassified instances than the other three combinations. Nonetheless, the misclassification rate was still lower than that of the ANN and k-NN models. The findings indicate that the integration of multiple measures can be utilized to identify and categorize mental fatigue in equipment operators.

5.2. Comparison with studies in non-construction domain

In this study, machine learning was utilized for the first time to recognize and categorize mental fatigue in equipment operators by integrating multiple types of data. The findings indicate that similar to previous studies in other fields, combining data from various sources can be used to identify mental fatigue. However, the current study performed better than studies in other domains in terms of performance metrics. For instance, Ding et al. (2020) achieved an accuracy of 58.5% when using a fusion of ECG and EDA for classifying mental workload with neural networks; however, combining all physiological measures as input data increased the accuracy to 78.3%. In another study by Xu et al. (2015), a combination of ECG, GSR, SpO2, electroencephalography, and electromyography was used to differentiate cognitive tasks, and achieved an accuracy of 73.0% with support vector machines. Similarly, Hirachan et al. (2022) fused data from four sensors, including ECG and EDA, to distinguish cognitive workloads, and achieved an accuracy of 74.0% with DT models. The DT model achieved an accuracy of 68.0% when using single-modal data. Majid et al. (2022) found that combining data from multiple physiological modalities, such as electroencephalography, galvanic skin response, and photoplethysmography, increased the perceived stress classification accuracy to 95.0% for two stress classes and 77.5% for three classes. Similarly, Jaiswal et al. (2022) utilized a fusion of input data from four sensors, namely EEG, ECG, EDA, and EMG, to detect cognitive fatigue, and achieved an accuracy of 77.2% using a random forest model. While the current study achieved higher accuracy than previous studies in other domains, making an exact comparison is challenging because of differences in experimental protocols and the nature of tasks performed. Nevertheless, the findings suggest that the current approach has significant potential for improving mental fatigue assessment for construction operators and workers, which could help reduce the occurrence of injuries and accidents at construction sites.

5.3. Comparison with studies in construction industry

The integration of data from multiple sensors yielded a higher classification accuracy compared to previous studies in the construction domain that only used single-modal data. Table 9 shows a comparison of the results of the current study with other relevant methods in the literature. Prior studies in construction have focused on classifying stress or fatigue using machine learning with single-modal data, achieving acceptable accuracy. In contrast, the current approach is significantly different because it integrates the input data from multiple sensors in various combinations for mental fatigue classification. For example, Jeon and Cai (2022) used a two-step ensemble approach to classify hazard recognition and cognitive states using single-modal EEG data, and achieved 82.3% accuracy with the LightGBM classifier. Jebelli et al. (2019a) used the OMTL-Von Neumann method for stress recognition in construction workers and achieved 77.61% accuracy, while another study by Jebelli et al. (2018c) used non-linear support vector machines to classify construction worker stress with 71.1% accuracy using single-modal EEG data on a construction site. However, these studies differed from the current approach because they did not focus on prolonged tasks or mental fatigue in construction equipment operators. Additionally, direct comparison with these studies may be challenging owing to variations in experimental setups, task nature, number of subjects, and subject differences.

6. Practical implications

The findings of this study have important implications in improving the health and safety of construction operators at construction sites.

Table 9
Comparison of classification accuracies in construction industry studies.

Reference	No. of subjects	Mode(s)	Stress or Fatigue (Levels)	Stimulus (Data collection settings)	Classification Method	Accuracy (%)
Jeon and Cai (2022)	30	EEG	Hazard (3)	Simulated environment. (Laboratory setting)	LightGBM	82.3
Jebelli et al. (2018b)	11	EEG	Stress (2)	Working on ladder (Construction site)	Fully connected NN	79.26
Jebelli et al. (2018c)	8	EEG	Stress (2)	Working on ladder (Construction site)	Non-linear support vector machine	71.1
Aryal et al. (2017)	12	Beta 1 channel	Fatigue (4)	Psychomotor Vigilance Task (indoor simulated)	Boosted trees	82.60
Jebelli et al. (2019a)	5	EEG	Stress (2)	Working on ladder (construction site)	OMTL-VonNeuman	77.61
Current study	16	EEG, EDA and FF	Fatigue (3)	Excavation Operation (Construction Site)	ANN k-NN DT	94.7 85.8 96.2

First, unlike previous studies that used single-modal data to detect mental stress or fatigue, this study demonstrated that it is possible to fuse data from multiple sensors to classify the mental fatigue levels of construction operators accurately. This suggests that practitioners and researchers can use a single system with multiple sensors to detect mental fatigue in equipment operators. Second, the findings of this study show that wearable electroencephalography, electrodermal activity sensors, and a mobile camera can be used to collect onsite experimental data for detecting mental fatigue, which has practical implications for real-time fatigue management of construction workers. Third, construction managers can use the insights from this study to develop a framework for managing worker shifts by observing equipment operators every 30–45 min and introducing breaks between shifts to allow them to recuperate from mental fatigue. Moreover, the findings of this study can be applied to other cognitive failures such as mental stress, mental workload, hazard identification, and emotions, which can aid in better incident management for construction workers experiencing cognitive issues.

7. Limitations and future research

The proposed study is the first to classify mental fatigue in construction equipment operators using machine-learning-based models and multimodal sensor data. While this study's findings have extended the understanding of mental fatigue monitoring by utilizing multimodal data as input, there remain caveats that should be addressed in subsequent studies. Second, this study did not use multimodal data to establish thresholds for the different levels of mental fatigue. Future research may leverage these thresholds to recognize and classify mental fatigue states, depending on whether they can be established and applied to all construction equipment operators. Second, the study used machine-learning models and multimodal sensor data to categorize mental fatigue in equipment operators, but the features were manually crafted from various sensors and then combined for classification purposes. Future research should utilize deep learning techniques or a combination of multiple deep learning techniques and raw multimodal data to identify mental fatigue in operators without the manual crafting of features. Unsupervised learning may also be employed in future studies to learn the features related to operators' mental fatigue on unlabeled multimodal sensor data. Third, the sample size was small, with only sixteen equipment operators and three levels of mental fatigue. Although the sample size was based on previous research, it may restrict the application of the proposed approach in the construction industry. To make the results more generalizable to the entire population of operators, future research should gather extensive datasets that represent a range of mental fatigue states. Fourth, this study evaluated mental fatigue using only three levels. Future studies should assess performance using more classes of mental fatigue for a better understanding of mental fatigue in real time. Fifth, the ground truth of mental fatigue was based on operators' subjective assessment, which may have been influenced by personal biases. Although the operators were familiar with the evaluation method of mental fatigue level, subjective assessment can still be considered a limitation owing to its lack of objectivity. However, it is a reliable technique for annotating data, despite its potential shortcomings. Sixth, this study focused solely on excavation operators as equipment operators. Subsequent research should replicate these results for operators of different types of construction equipment, such as cranes, dozer, and grader operators. In general, it is crucial to collect a large dataset with sufficient samples from various groups of equipment operators to identify additional mental fatigue states that are essential for training, testing, and constructing a comprehensive model for construction operations. Lastly, privacy concerns are valid and real concerns when implementing a system for mental fatigue recognition among construction workers. While the system proposed in the current study holds the potential to revolutionize occupational health by enabling real-time monitoring and proactive interventions, there are also inherent risks related to device hacking, data breaches, privacy issues, and data mismanagement. However, addressing such concerns is beyond the scope of this study. Future studies should focus on evaluating and mitigating these risks to ensure the successful on-field deployment of such a system. This involves carefully assessing the system architecture and hardware to ensure robustness against privacy and data security concerns. Measures should be implemented to secure the collected data, including encryption during transmission and storage, to prevent unauthorized access. Obtaining informed consent from the construction workers is essential. They should be informed of the purpose of data collection and how it will be used solely for on-site safety management. Transparency in the process helps build trust and confidence among workers regarding the protection and security of their data.

8. Conclusions

This study introduced a novel approach for classifying mental fatigue in construction equipment operators using supervised machine learning and multimodal sensor data fusion. Sixteen equipment operators participated in an excavation operation on a construction site, with subjective assessment using NASA-TLX as the ground truth for mental fatigue. Simultaneous measurement of EEG and EDA using wearable Muse headbands and E4 watches, respectively, and video recording for the geometric measurement of facial features were conducted during the experiment. Mental fatigue was induced by a monotonous and prolonged excavation task. After the experiment, the features were extracted and integrated as the input data from multiple sensors. Three supervised machine-learning models, including an artificial neural network (ANN), k-nearest neighbors (k-NN), and a decision tree (DT), and four combinations of multimodal data were used to classify three levels of mental fatigue: alert, mild fatigue, and fatigue. Additionally, the performance of these models was evaluated using assessment metrics, such as accuracy, precision, recall, specificity, and F1-score. The experimental findings indicate that the DT model outperformed the other models for all combinations of multimodal data, achieving an overall accuracy of 96.9% (FF and EDA), 85.0% (EEG and EDA), 97.1% (EEG and EDA), and 96.2% (EEG, EDA, and FF). Although the overall accuracy of the ANN and k-NN models was inferior to that of the DT model, their performance was still better than that of previous studies conducted with single-modal data. These findings support the use of the DT model and the fusion of data from multiple sensors to classify mental fatigue states during construction equipment operations, aiding in the development of a single real-time system of multiple sensors and machine learning to classify mental fatigue in operators. Additionally, the system will improve safety and health management at construction sites by enabling safety managers to track the mental fatigue level of operators in real-time, thereby reducing injuries and accidents at construction sites.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This study acknowledges the following funding grants: 1. General Research Fund (GRF) Grant (15201621) titled “Monitoring and managing fatigue of construction plant and equipment operators exposed to prolonged sitting”; 2. General Research Fund (GRF) Grant (15210720) titled “The development and validation of a noninvasive tool to monitor mental and physical stress in construction workers”; and 3. Research Institute for Intelligent Wearable System (RI-IWEAR) - Strategic Supporting Scheme (CD47) titled “An automated assessment of construction equipment operators’ mental fatigue based on facial expressions”.

References

Antwi-Afari, M.F., Anwer, S., Umer, W., Mi, H.-Y., Yu, Y., Moon, S., Hossain, M.U., 2023. Machine learning-based identification and classification of physical fatigue levels: a novel method based on a wearable insole device. Int. J. Ind. Ergon. 93, 103404.
Arsalan, A., Majid, M., Butt, A.R., Anwar, S.M., 2019. Classification of perceived mental stress using A commercially available EEG headband. IEEE J. Biomed. Health Inform. 23, 2257–2264.
Aryal, A., Ghahramani, A., Becerik-Gerber, B., 2017. Monitoring fatigue in construction workers using physiological measurements. Autom. ConStruct. 82, 154–165.
Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y., 2015. Physical human activity recognition using wearable sensors. Sensors 15, 31314–31338.
Baltrušaitis, T., Robinson, P., Morency, L., 2016. OpenFace: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision, pp. 1–10 (WACV), 7-10 March 2016.
Birhane, G.E., Yang, L., Geng, J., Zhu, J., 2022. Causes of construction injuries: a review. Int. J. Occup. Saf. Ergon. 28, 343–353.
Bitkina, O.V., Park, J., Kim, H.K., 2021. The ability of eye-tracking metrics to classify and predict the perceived driving workload. Int. J. Ind. Ergon. 86, 103193.
Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., 2012. Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 6442–6445.
Braithwaite, J., 2013. A guide for analysing electrodermal activity & skin conductance responses for psychological experiments/J. Jason Braithwaite, Derrick G Watson, Robert Jones, Mickey Rowe. In: –Selective Attention & Awareness Laboratory Behavioural Brain Sciences Centre. University of Birmingham, UK (Tech. Rep).
Cai, Q., Wang, H., Li, Z., Liu, X., 2019. A survey on multimodal data-driven smart healthcare systems: approaches and applications. IEEE Access 7, 133583–133599.
Cannard, C., Wahbeh, H., Delorme, A., 2021. Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp. 3603–3610.
Chae, J., Hwang, S., Seo, W., Kang, Y., 2021. Relationship between rework of engineering drawing tasks and stress level measured from physiological signals. Autom. ConStruct. 124, 103560.
Charles, R.L., Nixon, J., 2019. Measuring mental workload using physiological measures: a systematic review. Appl. Ergon. 74, 221–232.
Chen, J., Taylor, J.E., Comu, S., 2017. Assessing task mental workload in construction projects: a novel electroencephalography approach. J. Construct. Eng. Manag. 143, 04017053.
Cheng, Q., Wang, W., Jiang, X., Hou, S., Qin, Y., 2019. Assessment of driver mental fatigue using facial landmarks. IEEE Access 7, 150423–150434.
Choi, B., Jebelli, H., Lee, S., 2019. Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers’ perceived risk. Saf. Sci. 115, 110–120.
Choi, J., Gu, B., Chin, S., Lee, J.-S., 2020. Machine learning predictive model based on national data for fatal accidents of construction workers. Autom. ConStruct. 110, 102974.
Collet, C., Salvia, E., Petit-Boulanger, C., 2014. Measuring workload with electrodermal activity during common braking actions. Ergonomics 57, 886–896.
Das, S., Maiti, J., Krishna, O.B., 2020. Assessing mental workload in virtual reality based EOT crane operations: a multi-measure approach. Int. J. Ind. Ergon. 80, 103017.
Dasari, D., Shou, G., Ding, L., 2013. EEG Index for Time-On-Task Mental Fatigue in Real Air Traffic Controllers Obtained via Independent Component Analysis.
Demšar, J., Curk, T., Erjavec, A., Gorup, č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., 2013. Orange: data mining toolbox in Python. J. Mach. Learn. Res. 14, 2349–2353.
Ding, Y., Cao, Y., Duffy, V.G., Wang, Y., Zhang, X., 2020. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. Ergonomics 63, 896–908.
Dissanayake, U.C., Steuber, V., Amirabdollahian, F., 2022. EEG spectral feature modulations associated with fatigue in robot-mediated upper limb gross and fine motor interactions. Front. Neurorob. 15, 192.
Doudou, M., Bouabdallah, A., Berge-Cherfaoui, V., 2020. Driver drowsiness measurement technologies: current research, market solutions, and challenges. Int. J. Intell. Trans. Syst. Res. 18, 297–319.
Dziuda, Ł., Baran, P., Zieliński, P., Murawski, K., Dziwosz, M., Krej, M., Piotrowski, M., Stablewski, R., Wojdas, A., Strus, W., Gasiul, H., Kosobudzki, M., Bortkiewicz, A., 2021. Evaluation of a fatigue detector using eye closure-associated indicators acquired from truck drivers in a simulator study. Sensors 21, 6449.
Eoh, H.J., Chung, M.K., Kim, S.-H., 2005. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. Int. J. Ind. Ergon. 35, 307–320.
Fan, X., Zhou, Q., Liu, Z., Xie, F., 2015. Electroencephalogram assessment of mental fatigue in visual search. Bio Med. Mater. Eng. 26, S1455–S1463.
Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P.G., Marias, K., Tsiknakis, M., 2017. Stress and anxiety detection using facial cues from videos. Biomed. Signal Process Control 31, 89–101.
Greco, A., Valenza, G., Lanata, A., Scilingo, E.P., Citi, L., 2015. cvxEDA: a convex optimization approach to electrodermal activity processing. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 63, 797–804.
Grier, R.A., 2015. How high is high? A meta-analysis of NASA-TLX global workload scores. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications Sage CA, Los Angeles, CA, pp. 1727–1731.
Gromer, M., Salb, D., Walzer, T., Madrid, N.M., Seepold, R., 2019. ECG sensor for detection of driver’s drowsiness. Procedia Comput. Sci. 159, 1938–1946.
Han, Y., Jin, R., Wood, H., Yang, T., 2019. Investigation of demographic factors in construction employees’ safety perceptions. KSCE J. Civ. Eng. 23, 2815–2828.
Han, Y., Yin, Z., Zhang, J., Jin, R., Yang, T., 2020. Eye-tracking experimental study investigating the influence factors of construction safety hazard recognition. J. Construct. Eng. Manag. 146, 04020091.
Hancock, P.A., Matthews, G., 2019. Workload and performance: associations, insensitivities, and dissociations. Hum. Factors 61, 374–392.
Hart, S.G., 2006. 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, pp. 904–908.
Hasanzadeh, S., Esmaeili, B., Dodd, M.D., 2018. Examining the relationship between construction workers’ visual attention and situation awareness under fall and tripping hazard conditions: using mobile eye tracking. J. Construct. Eng. Manag. 144, 04018060.
Hirachan, N., Mathews, A., Romero, J., Rojas, R.F., 2022. Measuring cognitive workload using multimodal sensors. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) 4921–4924, 11-15 July 2022.
Hse, H.A.S.E., 2020. Construction Statistics in Great Britain. Available from. http://www.hse.gov.uk/statistics/industry/construction.pdf. (Accessed 30 September 2021).
Hu, J., Min, J., 2018. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. Cognitive neurodynamics 12, 431–440.
Hu, Z., Chan, W.T., Hu, H., Xu, F., 2023. Cognitive factors underlying unsafe behaviors of construction workers as a tool in safety management: a review. J. Construct. Eng. Manag. 149, 03123001.
Hwang, S., Jebelli, H., Choi, B., Choi, M., Lee, S., 2018. Measuring workers’ emotional state during construction tasks using wearable EEG. J. Construct. Eng. Manag. 144, 04018050.
Jaafar, M.H., Arifin, K., Aiyub, K., Razman, M.R., Ishak, M.I.S., Samsurijan, M.S., 2018. Occupational safety and health management in the construction industry: a review. Int. J. Occup. Saf. Ergon. 24, 493–506.

Jabbar, R., Al-Khalifa, K., Kharbeche, M., Alhajyaseen, W., Jafari, M., Jiang, S., 2018. Real-time driver drowsiness detection for android application using deep neural networks techniques. Proc. Comput. Sci. 130, 400–407.

Jaiswal, A., Zadeh, M.Z., Hebri, A., Makedon, F., 2022. Assessing Fatigue with Multimodal Wearable Sensors and Machine Learning. *arXiv preprint arXiv: 2205.00287*.

Jap, B.T., Lal, S., Fischer, P., Bekiaris, E., 2009. Using EEG spectral components to assess algorithms for detecting fatigue. Expert Syst. Appl. 36, 2352–2359.

Jebelli, H., Choi, B., Kim, H., Lee, S., 2018a. Feasibility study of a Wristband-type wearable sensor to understand construction workers' physical and mental status. Construct. Res. Cong. 2018.

Jebelli, H., Hwang, S., Lee, S., 2018b. EEG-based workers' stress recognition at construction sites. Autom. ConStruct. 93, 315–324.

Jebelli, H., Khalili, M.M., Hwang, S., Lee, S., 2018c. A supervised learning-based construction workers' stress recognition using a wearable electroencephalography (EEG) device. Construct. Res. Cong. 2018.

Jebelli, H., Khalili, M.M., Lee, S., 2019a. A continuously updated, computationally efficient stress recognition framework using electroencephalogram (EEG) by applying online multitask learning algorithms (OMTL). IEEE J. Biomed. Health Inform. 23, 1928–1939.

Jebelli, H., Khalili, M.M., Lee, S., 2019b. Mobile EEG-based workers' stress recognition by applying deep neural network. In: MUTIS, I., HARTMANN, T. (Eds.), Advances in Informatics and Computing in Civil and Construction Engineering, 2019//. Springer International Publishing, Cham, pp. 173–180.

Jeelani, I., Albert, A., Han, K., Azevedo, R., 2019. Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. J. Construct. Eng. Manag. 145, 04018115.

Jeon, J., Cai, H., 2022. Multi-class classification of construction hazards via cognitive states assessment using wearable EEG. Adv. Eng. Inf. 53, 101646.

Ji, Y., Wang, S., Zhao, Y., Wei, J., Lu, Y., 2019. Fatigue state detection based on multi-index fusion and state recognition network. IEEE Access 7, 64136–64147.

Kaduk, S.I., Roberts, A.P.J., Stanton, N.A., 2021. Driving performance, sleepiness, fatigue, and mental workload throughout the time course of semi-automated driving—experimental data from the driving simulator. Human Fact. Ergonom. Manufact. Service Indus. 31, 143–154.

Ke, J., Du, J., Luo, X., 2021a. The effect of noise content and level on cognitive performance measured by electroencephalography (EEG). Autom. ConStruct. 130, 103836.

Ke, J., Zhang, M., Luo, X., Chen, J., 2021b. Monitoring distraction of construction workers caused by noise using a wearable Electroencephalography (EEG) device. Autom. ConStruct. 125, 103598.

Khalid, U., Sagoo, A., Benachir, M., 2021. Safety Management System (SMS) framework development – mitigating the critical safety factors affecting Health and Safety performance in construction projects. Saf. Sci. 143, 105402.

Kines, P., Andersen, L.P., Spangenberg, S., Mikkelsen, K.L., Dyreborg, J., Zohar, D., 2010. Improving construction site safety through leader-based verbal safety communication. J. Saf. Res. 41, 399–406.

King, D.E., 2009. Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. 10, 1755–1758.

Koc, K., Gurgun, A.P., 2022. Scenario-based automated data preprocessing to predict severity of construction accidents. Autom. ConStruct. 140, 104351.

Krauss, T.P., Shure, L., Little, J., 1994. Signal Processing Toolbox for Use with MATLAB®: User's Guide. The MathWorks.

Kukasvadiya, M.S., Divecha, N.H., 2017. Analysis of data using data mining tool orange. Int. J. Exp. Diabetes Res. 5, 1836–1840.

Lee, B.G., Choi, B., Jebelli, H., Lee, S., 2021. Assessment of construction workers' perceived risk using physiological data from wearable sensors: a machine learning approach. J. Build. Eng. 42, 102824.

Lee, G., Lee, S., 2022. Feasibility of a mobile electroencephalogram (EEG) sensor-based stress type classification for construction workers. Construct. Res. Cong. 2022.

Li, G., Huang, S., Xu, W., Jiao, W., Jiang, Y., Gao, Z., Zhang, J., 2020a. The impact of mental fatigue on brain activity: a comparative study both in resting state and task state using EEG. BMC Neurosci. 21, 1–9.

Li, H., Wang, D., Chen, J., Luo, X., Li, J., Xing, X., 2019a. Pre-service fatigue screening for construction workers through wearable EEG-based signal spectral analysis. Autom. ConStruct. 106, 102851.

Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., Hou, J., 2020b. Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. Autom. ConStruct. 109, 103000.

Li, J., Li, H., Wang, F., Cheng, A.S.K., Yang, X., Wang, H., 2021. Proactive Analysis of Construction Equipment Operators' Hazard Perception Error Based on Cognitive Modeling and a Dynamic Bayesian Network, vol. 205. Reliability Engineering & System Safety, 107203.

Li, J., Li, H., Wang, H., Umer, W., Fu, H., Xing, X., 2019b. Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology. Autom. ConStruct. 105, 102835.

Li, J., Yu, M., Wang, H., 2018. A taxonomy of performance shaping factors for shield tunnel construction. Eng. Construct. Architect. Manag.

Li, X., Zeng, J., Chen, C., Chi, H.-L., Shen, G.Q., 2022. Smart Work Package Learning for Decentralized Fatigue Monitoring through Facial Images. Computer-Aided Civil and Infrastructure Engineering (n/a).

Liu, P., Chi, H.-L., Li, X., Guo, J., 2021a. Effects of dataset characteristics on the performance of fatigue detection for crane operators using hybrid deep neural networks. Autom. ConStruct. 132, 103901.

Liu, Q., Liu, Y., Chen, K., Wang, L., Li, Z., Ai, Q., Ma, L., 2021b. Research on channel selection and multi-feature fusion of EEG signals for mental fatigue detection. Entropy 23, 457.

Liu, X., Li, G., Wang, S., Wan, F., Sun, Y., Wang, H., Bezerianos, A., Li, C., Sun, Y., 2021c. Toward practical driving fatigue detection using three frontal EEG channels: a proof-of-concept study. Physiol. Meas. 42, 044003.

Ma, L., Guo, H., Fang, Y., 2021. Analysis of construction workers' safety behavior based on myers-briggs type indicator personality test in a bridge construction project. J. Construct. Eng. Manag. 147, 04020149.

Maior, C.B.S., Das Chagas Moura, M.J., Santana, J.M.M., Lins, I.D., 2020. Real-time classification for autonomous drowsiness detection using eye aspect ratio. Expert Syst. Appl. 158, 113505.

Majid, M., Arsalan, A., Anwar, S.M., 2022. A Multimodal Perceived Stress Classification Framework Using Wearable Physiological Sensors. *arXiv preprint arXiv:2206.10846*.

Mehmood, I., Li, H., Qarout, Y., Umer, W., Anwer, S., Wu, H., Hussain, M., Fordjour Antwi-Afari, M., 2023. Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data. Adv. Eng. Inf. 56, 101978.

Mehmood, I., Li, H., Umer, W., Arsalan, A., Saad Shakeel, M., Anwer, S., 2022. Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators. Adv. Eng. Inf. 54, 101777.

Milstein, N., Gordon, I., 2020. Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. Front. Behav. Neurosci. 14.

Morales, J.M., Díaz-Piedra, C., Rieiro, H., Roca-González, J., Romero, S., Catena, A., Fuentes, L.J., Di Stasi, L.L., 2017. Monitoring driver fatigue using a single-channel electroencephalographic device: a validation study by gaze-based, driving performance, and subjective data. Accid. Anal. Prev. 109, 62–69.

Murphy, K.P., 2012. Machine Learning: a Probabilistic Perspective. MIT press.

Naik, A., Samant, L., 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Comput. Sci. 85, 662–668.

Noghabaei, M., Han, K., Albert, A., 2021. Feasibility study to identify brain activity and eye-tracking features for assessing hazard recognition using consumer-grade wearables in an immersive virtual environment. J. Construct. Eng. Manag. 147, 04021104.

Ojha, A., Shakerian, S., Habibnezhad, M., Jebelli, H., 2023. Feasibility verification of multimodal wearable sensing system for holistic health monitoring of construction workers. In: Walbridge, S., Nik-Bakht, M., Ng, K.T.W., Shome, M., Alam, M.S., El Damatty, A., Lovegrove, G. (Eds.), Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021, 2023//. Springer Nature Singapore, Singapore, pp. 283–294.

Orfanidis, S.J., 1995. Introduction to Signal Processing. Prentice-Hall, Inc.

Özdemir, A.T., Barshan, B., 2014. Detecting falls with wearable sensors using machine learning techniques. Sensors 14, 10691–10708.

Poh, M.-Z., Swenson, N.C., Picard, R.W., 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans. Biomed. Eng. 57, 1243–1252.

Posada-Quintero, H.F., Chon, K.H., 2020. Innovations in electrodermal activity data collection and signal processing: a systematic review. Sensors 20, 479.

Posada-Quintero, H.F., Florian, J.P., Orjuela-Cañ̃n, A.D., Chon, K.H., 2018. Electrodermal activity is sensitive to cognitive stress under water. Front. Physiol. 8.

Prabaswari, A.D., Basumerda, C., Utomo, B.W., 2019. The mental workload analysis of staff in study program of private educational organization. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 012018.

Raufi, B., Longo, L., 2022. An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload. Front. Neuroinf. 16.

Saedi, S., Fini, A.A.F., Khanzadi, M., Wong, J., Sheikhkhoshkar, M., Banaei, M., 2022. Applications of electroencephalography in construction. Autom. ConStruct. 133, 103985.

Sarkar, S., Pramanik, A., Maiti, J., Reniers, G., 2020. Predicting and analyzing injury severity: a machine learning-based approach using class-imbalanced proactive and reactive data. Saf. Sci. 125, 104616.

Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Soares, G., De Lima, D., Neto, A.M., 2019. A mobile application for driver's drowsiness monitoring based on PERCLOS estimation. IEEE Ltin America Transact. 17, 193–202.

Stancin, I., Frid, N., Cifrek, M., Jovic, A., 2021. EEG signal multichannel frequency-domain ratio indices for drowsiness detection based on multicriteria optimization. Sensors 21, 6932.

Sweeney, K.T., Ward, T.E., Mcloone, S.F., 2012. Artifact removal in physiological signals—Practices and possibilities. IEEE Trans. Inf. Technol. Biomed. 16, 488–500.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., Zhang, T., 2019. A systematic review of physiological measures of mental workload. Int. J. Environ. Res. Publ. Health 16, 2716.

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., Picard, R., 2015. Automatic identification of artifacts in electrodermal activity data. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 1934–1937.

Techera, U., Hallowell, M., Littlejohn, R., Rajendran, S., 2018. Measuring and predicting fatigue in construction: empirical field study. J. Construct. Eng. Manag. 144, 04018062.

Tehrani, B.M., Wang, J., Truax, D., 2021. Assessment of Mental Fatigue Using Electroencephalography (EEG) and Virtual Reality (VR) for Construction Fall Hazard Prevention. Construction and Architectural Management, Engineering (ahead-of-print).

Tyas, A.E., Wibawa, A.D., Purnomo, M.H., 2020. Theta, alpha and beta activity in the occipital based on EEG signals for mental fatigue in high school students. In: 2020 International Conference on Smart Technology and Applications (ICoSTA), pp. 1–7, 20-20 Feb. 2020.
Umer, W., 2022. Simultaneous monitoring of physical and mental stress for construction tasks using physiological measures. J. Build. Eng. 46, 103777.
Umer, W., Li, H., Yantao, Y., Antwi-Afari, M.F., Anwer, S., Luo, X., 2020. Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures. Autom. ConStruct. 112, 103079.
Utomo, D., Yang, T.-H., Thanh, D.T., Hsiung, P.-A., 2019. Driver fatigue prediction using different sensor data with deep learning. In: 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS). IEEE, pp. 242–247.
Vahdatikhaki, F., El Ammari, K., Langroodi, A.K., Miller, S., Hammad, A., Doree, A., 2019. Beyond data visualization: a context-realistic construction equipment training simulators. Autom. ConStruct. 106, 102853.
Vidya, K.S., Ng, E., Acharya, U.R., Chou, S.M., San Tan, R., Ghista, D.N., 2015. Computer-aided diagnosis of myocardial infarction using ultrasound images with DWT, GLCM and HOS methods: a comparative study. Comput. Biol. Med. 62, 86–93.
Wagstaff, A.S., Sigstad Lie, J.-A., 2011. Shift and night work and long working hours – a systematic review of safety implications. Scand. J. Work. Environ. Health 173–185.
Walambe, R., Nayak, P., Bhardwaj, A., Kotecha, K., 2021. Employing multimodal machine learning for stress detection. J. Healthc Eng. 2021, 9356452.
Wan, W., Cui, X., Gao, Z., Gu, Z., 2021. Frontal EEG-based multi-level attention states recognition using dynamical complexity and extreme gradient boosting. Front. Hum. Neurosci. 15.
Wang, C., Guragain, B., Verma, A.K., Archer, L., Majumder, S., Mohamud, A., Flaherty-Woods, E., Shapiro, G., Almashor, M., Lenné, M., 2019a. Spectral analysis of EEG during microsleep events annotated via driver monitoring system to characterize drowsiness. IEEE Trans. Aero. Electron. Syst. 56, 1346–1356.
Wang, D., Li, H., Chen, J., 2019b. Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals. Autom. ConStruct. 100, 11–23.
Wang, H., Wu, C., Li, T., He, Y., Chen, P., Bezerianos, A., 2019c. Driving fatigue classification based on fusion entropy analysis combining EOG and EEG. IEEE Access 7, 61975–61986.
Wang, M., Zhao, Y., Liao, P.-C., 2022. EEG-based work experience prediction using hazard recognition. Autom. ConStruct. 136, 104151.
Wang, Y., Huang, Y., Gu, B., Cao, S., Fang, D., 2023. Identifying mental fatigue of construction workers using EEG and deep learning. Autom. ConStruct. 151, 104887.
Wang, Z., Zhou, X., Wang, W., Liang, C., 2020. Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video. Int. J. Machine Learn Cybernetics 11, 923–934.
Witten, I.H., Frank, E., 2002. Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record 31, 76–77.
Wu, E.Q., Deng, P.Y., Qiu, X.Y., Tang, Z., Zhang, W.M., Zhu, L.M., Ren, H., Zhou, G.R., Sheng, R.S.F., 2021. Detecting fatigue status of pilots based on deep learning network using EEG signals. IEEE Transact. Cognit. Develop. Syst. 13, 575–585.
Xing, X., Li, H., Li, J., Zhong, B., Luo, H., Skitmore, M., 2019. A multicomponent and neurophysiological intervention for the emotional and mental states of high-altitude construction workers. Autom. ConStruct. 105, 102836.
Xing, X., Zhong, B., Luo, H., Rose, T., Li, J., Antwi-Afari, M.F., 2020. Effects of physical fatigue on the induction of mental fatigue of construction workers: a pilot study based on a neurophysiological approach. Autom. ConStruct. 120, 103381.
Xu, Q., Nwe, T.L., Guan, C., 2015. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE J. Biomed. Health Inform. 19, 275–281.
Yang, J., Ye, G., Xiang, Q., Kim, M., Liu, Q., Yue, H., 2021a. Insights into the mechanism of construction workers' unsafe behaviors from an individual perspective. Saf. Sci. 133, 105004.
Yang, Y., Gao, Q., Song, X., Song, Y., Mao, Z., Liu, J., 2021b. Facial expression and EEG fusion for investigating continuous emotions of deaf subjects. IEEE Sensor. J. 21, 16894–16903.
You, F., Li, X., Gong, Y., Wang, H., Li, H., 2019. A real-time driving drowsiness detection algorithm with individual differences consideration. IEEE Access 7, 179396–179408.
Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58, 1–17.
Zhang, G., Etemad, A., 2021. Capsule attention for multimodal EEG-EOG representation learning with application to driver vigilance estimation. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 1138–1149.
Zhang, Y., Zhang, M., Fang, Q., 2019. Scoping review of EEG studies in construction safety. Int. J. Environ. Res. Publ. Health 16, 4146.
Zhao, G., Liu, Y.-J., Shi, Y., 2018. Real-time assessment of the cross-task mental workload using physiological measures during anomaly detection. IEEE Trans. Human-Machine Syst. 48, 149–160.
Zhao, L., Li, M., He, Z., Ye, S., Qin, H., Zhu, X., Dai, Z., 2022. Data-driven learning fatigue detection system: a multimodal fusion approach of ECG (electrocardiogram) and video signals. Measurement 201, 111648.
Zhu, Q., Xu, X., Yuan, N., Zhang, Z., Guan, D., Huang, S.-J., Zhang, D., 2020. Latent correlation embedded discriminative multi-modal data fusion. Signal Process. 171, 107466.
Zhu, X., Ye, S., Zhao, L., Dai, Z., 2021. Hybrid attention cascade network for facial expression recognition. Sensors 21, 2003.