

RESEARCH

Open Access



Internet of things and ensemble learning-based mental and physical fatigue monitoring for smart construction sites

Bubryur Kim¹, K. R. Sri Preethaa^{2,3}, Sujeen Song⁴, R. R. Lukacs⁵, Jinwoo An⁶, Zengshun Chen⁷, Euijung An^{5*} and Sungho Kim⁵

*Correspondence:
babyofbear@gmail.com

¹ School of Space Engineering Sciences, Kyungpook National University, 80 Daehak-Ro, Buk-Gu, Daegu 41566, Korea

² Intelligent Construction Automation Center, Kyungpook National University, 80, Daehak-Ro, Buk-Gu, Daegu 41566, Korea

³ School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India 632014

⁴ Earth Turbine, 36, Dongdeok-Ro 40-Gil, Jung-Gu, Daegu 41905, Korea

⁵ Department of Robot and Smart System Engineering, Kyungpook National University, 80 Daehak-Ro, Buk-Gu, Daegu 41566, Korea

⁶ Department of Civil Engineering, College of Engineering and Computer Science, The University of Texas Rio Grande Valley, Edinburg, TX 78539, USA

⁷ School of Civil Engineering, Chongqing University, Chongqing 400044, China

Abstract

The construction industry substantially contributes to the economic growth of a country. However, it records a large number of workplace injuries and fatalities annually due to its hesitant adoption of automated safety monitoring systems. To address this critical concern, this study presents a real-time monitoring approach that uses the Internet of Things and ensemble learning. This study leverages wearable sensor technology, such as photoplethysmography and electroencephalography sensors, to continuously track the physiological parameters of construction workers. The sensor data is processed using an ensemble learning approach called the ChronoEnsemble Fatigue Analysis System (CEFAS), comprising deep autoregressive and temporal fusion transformer models, to accurately predict potential physical and mental fatigue. Comprehensive evaluation metrics, including mean square error, mean absolute scaled error, and symmetric mean absolute percentage error, demonstrated the superior prediction accuracy and reliability of the proposed model compared to standalone models. The ensemble learning model exhibited remarkable precision in predicting physical and mental fatigue, as evidenced by the mean square errors of 0.0008 and 0.0033, respectively. The proposed model promptly recognizes potential hazards and irregularities, considerably enhancing worker safety and reducing on-site risks.

Keywords: Smart construction sites, Internet of things, Ensemble learning, Fatigue monitoring, Safety management, Multivariate time series forecasting

Introduction

The construction industry is a global economic driver, substantially contributing to the gross domestic product of both developed and emerging economies [1]. Despite its economic importance, the industry is burdened with a high incidence of workplace accidents and safety hazards [2], establishing it as one of the most hazardous sectors [3]. Notably, these accidents are not only detrimental in terms of human cost but also lead to consequential declines in productivity, increased repair costs, and delays in project schedules [4–6]. Construction workers routinely perform physically demanding tasks [7–10], which increases their risk for chronic illnesses, cardiovascular diseases, and

other health issues [8, 10–12]. In addition to the high injury rate, their poor lifestyle habits and physical conditions further exacerbate health risks [13]. The Bureau of Labor Statistics data [14] highlights these issues, positioning construction as one of the sectors with the highest fatal occupational injury rates in the USA. Factors contributing to construction accidents include inadequate site management, the absence of necessary safety precautions, environmental hazards, unsafe behavior, and physical exhaustion [15, 16]. Among these, unsafe worker behavior and fatigue majorly contribute to accident incidence [10–12, 17–21]. Moreover, attention lapses due to fatigue and heavy workload when operating construction equipment are key precursors to accidents [22, 23].

Furthermore, the dynamic and hazardous nature of construction sites, coupled with the high prevalence of mental and physical fatigue among workers, constitutes a significant risk factor for accidents and injuries. The incidence of fatigue among construction personnel is alarmingly high, ranging from 10 to 40% across various trades [21, 24, 25], with 20–40% of different craft workers consistently surpassing the widely recognized physiological limits for manual labor [24]. This high prevalence can be attributed to the physically strenuous nature of construction tasks, which often involve long hours, shiftwork, and challenging environmental conditions [26, 27]. Higher levels of mental and physical fatigue were significantly correlated with greater musculoskeletal pain [28]. This indicates that fatigue not only increases the risk of accidents but also exacerbates physical discomfort and potential long-term musculoskeletal damage. Occupational fatigue is recognized as a major safety and health concern, particularly among workers at higher risk of injuries and illnesses due to inescapable tiring conditions. More than 43% of laborers suffer from sleep deprivation, notably those who work nights or have long, irregular shifts, contributing to fatigue [29]. The pervasive nature of fatigue in this sector is further evidenced by its association with a third of all occupational injuries, akin to impairment levels experienced under the influence of alcohol [26]. A 12-h workday is linked to a 37% higher risk of sustaining injuries [30, 31]. Additionally, when comparing day shifts to evening and night shifts, the incidence of accidents is 18% higher and 30% higher, respectively [32]. Fatigue has been empirically linked to impaired physical and cognitive function, leading to an increased risk of incidents as well as injuries [21, 33]. Construction workers with severe fatigue have been found to have 1.77 times higher odds of experiencing occupational injuries compared to those without fatigue [25]. Furthermore, 7.7% of non-fatal injuries among Colorado's construction workers were shown to be related to fatigue [34]. In a study of 328 construction workers and 69 incidents, fatigue-related impairment was associated with a significantly elevated incident rate of 9.6 vs. 0.8 incidents per 1000 person-hours worked [13]. Mental fatigue, especially prevalent among roles that entail high cognitive demands, leads to decrements in cognitive performance. This is marked by a substantial rise in the time it takes to detect hazards and an increase in the instances of missed detections by operators as subjective mental fatigue levels rise [26, 27]. The failure to detect hazards has been identified as a leading cause of accidents involving construction equipment, with 50% of fatal accidents being related to equipment operation [27]. The probability ratios for reporting a vehicle accident and a near-miss incident after a long work shift, as opposed to a standard-length shift, are 2.3 and 5.9, respectively [35]. The economic consequences of fatigue-related accidents are substantial, as they not only affect productivity—costing employers

between \$1200 to \$3100 per employee annually—but also result in an estimated annual cost of \$136.4 billion in the United States alone due to health-related lost productive work time [29, 36]. The precise assessment of fatigue's direct influence on construction safety is challenging to quantify, largely attributed to the lack of established real-time fatigue monitoring methodologies. This deficiency highlights the imperative need for the development of fatigue monitoring technologies to mitigate the risks associated with mental and physical fatigue and ensure the safety and well-being of construction workers [37].

Smart technologies have revolutionized sectors such as healthcare; however, their adoption in construction is surprisingly limited [12, 38]. In light of these concerns, the primary research question guiding this study is as follows: how can the integration of the Internet of Things (IoT) and ensemble learning methodologies enhance real-time fatigue monitoring in the construction industry? This study aims to address this question by providing an innovative, technology-driven solution to improve safety measures and mitigate health risks in the construction industry.

Related work

Overview of wearable technology in the construction industry

Wearable technology that incorporates components such as computers, software, electronics, and sensors seamlessly embedded in clothing or accessories that can be easily worn on the body (e.g., smart glasses, watches, or garments) has recently garnered considerable attention across a multitude of industries due to its capability to monitor and analyze real-time biological, health, and safety parameters [39, 40]. The most prominent application of such devices is in the healthcare sector, where wearable health devices integrated with sophisticated biosensor systems enable the perpetual surveillance of an individual's physiological parameters, which include, but are not limited to, heart rate (HR), heart-rate variability (HRV), blood volume pressure (BVP), and dermal temperature, thereby providing instantaneous feedback [12, 21, 41].

To harness these opportunities for proactive safety management, various wearable sensing technologies have been extensively researched. This research includes the deployment of motion sensors such as inertial measurement units and an array of physiological sensors comprising heart-rate monitors, electrodermal activity sensors, skin temperature measurement devices, eye-tracking systems, and brainwave-monitoring apparatuses. These technologies aim to enhance safety measures by providing real-time data on workers' physical and mental states. Wearable technology has also been applied in the construction industry to improve the safety and health of construction workers. Applications in construction include monitoring physical activities such as walking, running, and lifting through embedded motion sensors and utilizing physiological sensors for continuous health status monitoring to identify potential risks like fatigue [8, 42, 43].

Moreover, wearable technology combined with sophisticated data analytical methods, such as machine learning and deep learning algorithms, can be used to analyze large amounts of data acquired by wearable devices and extract meaningful insights for safety and health management in the construction sector [2, 3, 44]. For instance, machine learning algorithms can predict fatigue risk based on physiological data collected by wearable devices [45, 46]. This prediction capability can help prevent fatigue-related

accidents, thereby enhancing the overall safety of construction workers. However, the adoption of wearable technology in construction faces challenges related to data accuracy, worker privacy, and comfort, which necessitate the development of robust data analysis techniques and addressing worker concerns to promote its usage [47–49]. To overcome these challenges, advanced wearable systems that combine multiple sensors and data analysis techniques have been widely researched. These systems comprehensively and accurately monitor the safety and health of construction workers [38–40, 50, 51]. For instance, a wearable system that integrates motion and physiological sensors was recently proposed to simultaneously monitor the physical activities and physiological status of construction workers [8].

In conclusion, while wearable technology has demonstrated immense potential for enhancing the safety and well-being of construction personnel, further research is required to overcome the existing challenges and realize its full potential. In this regard, integrating wearable technology with the latest data analytics methods, such as ensemble learning, is a promising avenue for future research.

Wearable technology for fatigue detection

Construction work is labor-intensive and involves performing strenuous tasks over extended periods, which causes fatigue and affects the mental and physical health of construction workers while considerably increasing the risk of on-site accidents [7, 10, 18–21]. This necessitates the development of reliable fatigue detection systems for the construction industry. Fatigue detection traditionally relies on self-reported fatigue measures or the supervision of a site manager but yields inaccurate and biased results due to subjectivity [18]. Real-time fatigue detection, which is pivotal in proactive accidents, presents a significant challenge due to its inherent difficulty. Wearable technology is utilized for this purpose. The devices, which are equipped with an array of sensors, such as heart-rate sensors and brainwave monitors, can continuously monitor workers' physiological status and detect signs of fatigue [45, 52]. Deep learning algorithms are used to analyze sensor data, predict fatigue risk, and deliver timely warnings to prevent accidents [53]. The integration of machine learning techniques with wearable technology allows for fatigue detection, enhancing the timeliness of interventions and helping to prevent fatigue-related accidents [20, 52, 54].

Photoplethysmography sensor

A photoplethysmography (PPG) sensor is a noninvasive technology that measures cardiovascular parameters such as HR and changes in blood volume. Wearable devices containing PPG sensors, such as smartwatches, have been used for fatigue detection in the construction industry [20, 22]. The sensor detects variations in blood volume changes with each heartbeat, which are affected by physiological changes that occur during fatigue [21]. This technology is easy to use, unobtrusive, and can continuously monitor fatigue, among other advantages. However, PPG sensors may produce inaccurate readings due to motion artifacts, environmental light interference, and variable sensor placement [23]. PPG sensors measure the HR, BVP, and HRV, which are associated with fatigue [19]. HR denotes the frequency of heartbeats in a minute, whereas HRV refers to the fluctuation in the intervals between consecutive heartbeats. A lower HRV and a

higher HR are indicative of fatigue [20]. HR can be calculated directly from the time intervals between successive heartbeats. HRV is typically calculated using the root mean square of the successive differences to analyze the intervals between normal-to-normal (NN) intervals.

Electroencephalography sensor

Electroencephalography (EEG) sensors monitor brain activity through the scalp and are often used in sleep studies to identify changes in different sleep stages, including the onset of fatigue [21, 55]. Research indicates that EEG correlation patterns shift markedly across sleep stages, with significant interhemispheric synchrony during non-REM sleep, particularly in the theta and alpha bands, which are indicative of fatigue [56, 57]. Additionally, a reduction in EEG complexity from awake states to deeper non-REM stages was found, suggesting a potential analog for physiological fatigue detection [58]. These findings align with similar transitions observed during general anesthesia, where shifts in alpha frequency could indicate recovery phases, drawing parallels with sleep transitions that may predict fatigue recovery in real-world scenarios [59]. Additionally, with increasing sleep deprivation, higher synchrony among EEG channels across various brain regions was observed, suggesting that increased connectivity serves as a compensatory mechanism to maintain cognitive functions despite fatigue [60]. In the construction industry, EEG sensors can directly detect changes in brainwave patterns in real-time to indicate the onset of fatigue [21]. Although these sensors are valuable tools, their usage can be challenging due to their susceptibility to noise, the need for specialized expertise for data interpretation, and concerns related to comfort and practicality of use in an active work environment [61]. The brainwave frequencies associated with fatigue states are theta (4–7 Hz) and alpha (8–12 Hz) waves. An increase in theta and alpha wave activities is indicative of fatigue [62]. These frequencies can be calculated directly from the EEG sensor data using the fast Fourier transform.

Different sensor combinations

To improve fatigue detection accuracy, different types of sensors can be combined in wearable devices [63]. For instance, cardiovascular and emotional changes associated with fatigue can be comprehensively monitored by combining PPG and electrodermal activity sensors [7, 21]. Similarly, physiological changes caused by fatigue can be detected by integrating EEG and skin temperature sensors [21]. However, the development and calibration of systems that integrate multiple sensors present challenges in terms of complexity, power consumption, and potential data redundancy [64]. Furthermore, privacy and comfort-related concerns might increase when multiple sensors are involved, potentially affecting worker acceptance [48] of the technology.

As each sensor type and combination offers unique strengths and limitations for fatigue detection in the construction industry, their combination can be used to overcome the associated challenges and enhance accuracy and reliability [18, 21]. Using a combination of different sensors in wearable devices, various fatigue-associated metrics, such as HR, HRV, and theta and alpha waves [65], can be simultaneously measured to monitor a worker's fatigue level comprehensively and reliably.

Computational approaches for fatigue detection

The integration of computational methods in fatigue detection has seen a surge. While individual machine learning models have carved a path for innovative fatigue assessment, they are not without their challenges. Analytical approaches leveraging machine learning or simple statistical techniques, including logistic and linear regression, have been used to explore the data derived from wearable sensors and classify fatigue levels in construction workers [20, 21]. Various computational algorithms, including decision trees, boosted trees, support vector machines (SVM), random forest (RF), and artificial neural networks [54, 66], have been used for fatigue detection. However, these methods pose challenges, such as overfitting and feature selection [66]. One study utilized physiological parameters such as heart rate, breathing rate, and skin temperature, measured by textile-based wearable sensors. The method incorporated penalized logistic regression and multiple linear regression models, as well as supervised and unsupervised machine learning approaches, including Artificial Neural Network (ANN), to develop a real-time fatigue assessment system; however, the study's lack of generalizability beyond bar benders questions its broader applicability. The method also suffered from a high quantity of motion artifacts, and the lack of correlation with blood lactate levels also posed additional limitations [20].

Another approach involved a non-invasive method using a wearable respiration sensor and random forest classifier, which showed high accuracy and was less affected by motion artifacts. Even so, there was potential variability in sensor performance, and the study heavily relied on a subjective questionnaire for labeling, which could introduce bias [7]. Deep learning algorithms applied in the study [18] for motion capture and biomechanical analysis were hindered by environmental complexities and the dependency of high-quality RGB cameras, which could potentially compromise the accuracy of the model under less ideal conditions. A study [10] assessed workers' physical demand through physiological signals captured by wearable biosensors, employing a Gaussian kernel support vector machine for classifying physical demand levels. The limitations highlighted the need for validation on a larger, more diverse subject pool and the bias of subjective fatigue assessments. Another study [67] developed an automatic biomechanical workload estimation method using computer vision and smart insoles with pressure sensors, but the error rate of the load estimation exceeded 15%, and the noise along with signal artifacts affecting accuracy were not examined; thus, there is a potential for inaccuracies in real-world construction environments. A study also employed EMG and IMU sensors along with a recurrent neural network for continuous fatigue monitoring. The focus of this study is on work severity classification rather than determining real-time physical fatigue. Despite the good results, the potential variability in individual worker characteristics such as work experience, age, and health status can affect the system's accuracy. Furthermore, continuously monitoring fatigue levels for multiple tasks performed in short intervals can be challenging [68]. Wang et al. [69] implemented a wearable EEG system for monitoring attention and vigilance; however, the system's effectiveness can vary with individual workers. A different research involved monitoring

different physiological parameters such as heart rate and breathing rate, but the study was limited by the small sample size, which may not represent a broader population of construction workers [70]. A study [71] looked at wearable-device-based data for measuring construction workers' psychological status, but the initial combined data modeling may not be sufficiently accurate for all individuals. In addition, the K-means clustering algorithm was adapted to classify the physical data into only three distinct classes. [72] used a wristband-type wearable health device to collect heart rate data and measure the physical demands of construction personnel. The study, however, does not account for all individual, environmental, and mental factors affecting physical demands, which could influence the results. Wearable eye-tracking technology can also be used to evaluate the impact of mental fatigue on hazard detection abilities. Yet, the study was only conducted in a controlled lab environment with only a few participants. Additionally, the use of wearable tracking technology may not capture mental fatigue accurately [42].

Ensemble learning operates on the principle of leveraging a collection of "weak learners," which are individual models trained to address the same issue, to formulate a stronger predictive model through their aggregation. Weaker learners can be combined to develop a model that yields precise and robust results [73, 74]. One study investigated these ensemble classifiers for physical exertion modeling along with other models including K-nearest neighbor (KNN), SVM, and DT. However, the experiments were only conducted in a controlled environment that did not reflect the complexity of actual construction work, focusing on a single task [19]. In another study, boosted trees, bagged trees, and RUSBoosted trees were proposed for fatigue classification. The prediction accuracy of the individual models was compared to the ensemble methods, resulting in enhanced fatigue detection accuracy for the ensemble models. Even so, the overall prediction accuracy was only 82.60%, 79.10%, and 80.60% for the boosted trees, bagged trees, and RUSBoosted trees, respectively. The study was limited by the reliability of subjective Borg's RPE scale for fatigue assessment [21]. Ensemble learning was also used to analyze human activity data gathered by smartphone sensors. Results revealed that ensemble learning can enhance the accuracy of human activity recognition, providing more reliable results than several stand-alone algorithms [50].

The computational approaches using standalone models, while promising, still face challenges such as overfitting, feature selection, bias, lack of generalizability, and the nascent stage of application in wearable devices for safety and health monitoring in the construction industry [3, 9, 66, 74]. Ensemble learning, which combines predictions of machine learning models trained on different sets of sensor data, has shown potential for higher fatigue detection accuracy [21, 50]. However, selecting appropriate models for the ensemble and combining their predictions remains a challenge.

Challenges in fatigue detection

The aim to develop an effective fatigue assessment approach for construction workers has led to the development of various computational and non-computational fatigue detection methods. In spite of the improvements, each approach comes with its own unique set of challenges and limitations. One challenge is the accurate detection and

assessment of fatigue [7, 10, 18–21, 67, 68, 71, 75]. In addition, noise and signal artifacts can affect the accuracy of wearable sensing technology [72, 75]. There is also a scarcity of studies that have directly examined the relationship between changes in physiological metrics and physical fatigue in construction workers [57].

Approaches utilizing wearable sensors to measure physiological parameters such as heart rate, breathing rate, and skin temperature face several limitations. The measurements suffer from the high quantity of movement artifacts, affecting the accuracy during the actual construction tasks [20], and environmental influences that skew the data [21, 51]. These wearable sensors can also cause discomfort or distraction to construction workers, affecting the adoption of these technologies [48]. Furthermore, several experiments were conducted only in controlled laboratory settings, not reflecting actual construction site conditions [27]. Variations in sensor placement can lead to significant data discrepancies, questioning the reliability of the measurements [2, 44]. Some studies also suffer from a lack of generalizability [69].

Machine learning models, such as regression models and neural networks, are not immune to overfitting and feature selection issues, which can affect their generalizability and real-world application [20]. The necessity for large datasets to train these models often poses an obstacle, as does the variability in individual responses to fatigue [7]. Privacy concerns related to continuous physiological monitoring also cannot be overlooked [19]. The generalizability of findings is a recurrent concern, with studies often limited to specific tasks or controlled environments that do not mirror the complexities of actual construction sites [7, 10, 18–21]. Fatigue detection systems struggle with specificity, frequently being unable to differentiate between fatigue and other physiological states, which is crucial for accurate assessments [76]. Technologies such as 3D motion capture and deep learning networks require optimal conditions for operation, and their performance is limited by the quality of the equipment and the algorithms' robustness [18]. The higher accuracy of fatigue classification often requires multiple sensor data, yet using multiple sensors adds complexity to the system [21].

This study aims to address the following challenges, contributing toward more effective fatigue monitoring in the construction sector:

1. The development of an IoT sensor and ensemble learning-based ChronoEnsemble Fatigue Analysis System (CEFAS), which enables the monitoring of both physical and mental fatigue. The sensors can be effortlessly integrated into safety helmets and worn on the wrist. This integration into everyday safety gear ensures that the workers can comfortably perform their duties while their fatigue levels are continuously monitored.
2. The ensemble learning model extracts data from various sensors and applies sophisticated learning algorithms to predict physical and mental fatigue (PF and MF) not only in real-time but also proactively. Since people often do not realize when they are fatigued, early detection is crucial. By identifying the signs of fatigue before they become apparent, the model enables preventative measures to be taken, significantly reducing the risk of accidents, unsafe behaviors, and injuries in the construction sector.

3. A deep ensemble learning model is developed for accurately and reliably detecting MF and PF. Unlike traditional machine learning methods that categorize fatigue into discrete classes, this model provides a continuous and precise measure of fatigue. The ensemble approach can also handle noisy data, reduce overfitting, and improve feature selection.
4. To address the lack of openly available datasets suitable for training deep learning models in this domain, we assembled an exhaustive dataset in laboratory settings. This dataset contains physiological signal measurements recorded by photoplethysmography and electroencephalography sensors. The dataset provided a solid foundation for training and testing the deep ensemble learning model.

In summary, the proposed approach can considerably enhance the health and safety outcomes for construction personnel and contribute to a safer and healthier work environment in the construction industry. Through these contributions, we aspire to advance the current state-of-the-art in fatigue detection, ensuring safer work environments and improved worker well-being. The remainder of this paper is organized as follows: “Methodology” discusses the methodology, including the methodological process, dataset building, ensemble model, and prototype implementation. “Results and discussion” highlights the results and discusses the evaluation of the proposed approach. “Conclusions and future work” explores the limitations and future scope of the proposed approach and concludes the study.

Methodology

Herein, a deep ensemble learning-based fatigue detection approach is proposed for construction workers. The research combined PPG and EEG sensor data with an ensemble learning model to monitor and predict physical fatigue and mental fatigue in real-time. Through meticulous data collection, preprocessing, transformation, and the implementation of sophisticated models such as DeepAR and TFT, the study endeavors to accurately forecast fatigue levels. The methodological procedure is outlined below.

Methodological process

First, the research problem, i.e., the necessity for an enhanced safety management system in the construction industry, was identified. Then, the use of wearable technology, biosensor systems, and ensemble learning was analyzed by an exhaustive literature review. This study was meticulously constructed to ensure a comprehensive methodology for data accumulation and analysis. The methodological process is shown in Fig. 1. Following an exhaustive literature review on wearable technology, biosensor systems, and ensemble learning, we systematically selected appropriate wearable devices, such as PPG and EEG sensors, accompanied by a microcontroller.

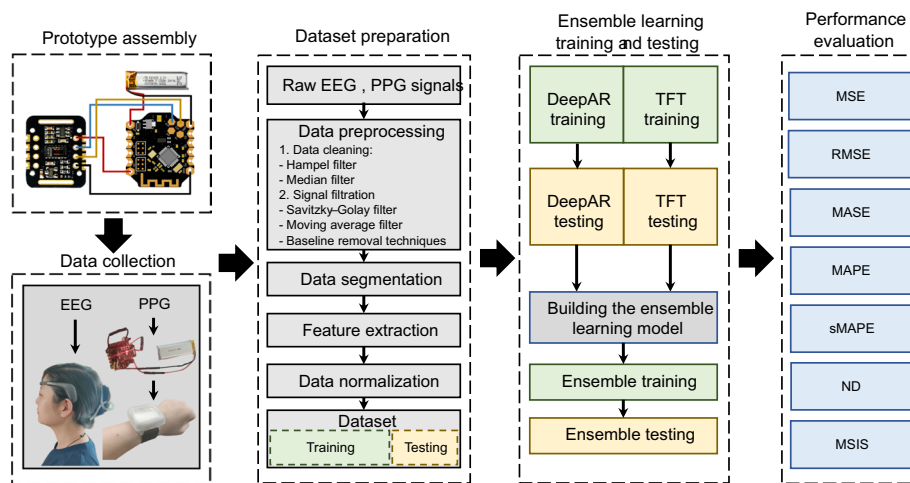


Fig. 1 Methodological process overview

Table 1 Mean age, mean height, mean weight, and standard deviation of parameters and the count of participants by gender

Gender	Mean age (years)	Mean height (cm)	Mean weight (kg)	Std. age	Std. height	Std. weight	Count
Male	25.33	174.5	67.92	1.21	5.61	10.14	6
Female	29.4	159.2	55	6.19	7.98	11.73	5

Subsequent to this selection, data was collected using these wearable sensors. The collected data were preprocessed to construct a robust and usable dataset, which was used to train the developed ensemble learning model.

After the training phase, the performance of the model was thoroughly analyzed using various evaluation metrics to evaluate the efficacy of the proposed approach. Lastly, a prototype was developed and validated under real-time conditions. Each phase of this methodological process was executed with meticulous attention to the research objectives and ethical considerations intrinsic to the study.

Dataset formulation and refinement

The dataset used for training the ensemble learning model comprises physiological metrics recast into a time-series configuration. This transformation is vital for the proficient employment of the deep autoregressive (DeepAR) and temporal fusion transformer (TFT) models designed to manage the time-series data. Various physiological metrics, including HR, BVP, HRV, and EEG signals, were systematically collected from 11 participants, comprising males with a mean age of 25.33 years and a standard deviation of 1.21 and females with a mean age of 29.4 years and a standard deviation of 6.19. Table 1 lists the details of the participants.

Data was acquired within a specified timeframe to ensure continuous and uninterrupted collection. The data were combined into a temporally annotated dataset using meticulous procedures, such as data acquisition, preprocessing, transformation, and segmentation, to ensure its robustness and suitability for subsequent analysis.

Data acquisition protocol

The initial preparations were completed in approximately 10 min. Before data collection, instruments were rigorously checked to ensure their operational readiness. Participants were acquainted with the objectives and protocols of the study and given an informed consent form for endorsement. The volunteers filled out a comprehensive demographic and health-related questionnaire. This step ensured that the participants met the inclusion criteria of the study and identified any potential exclusions.

Baseline data were acquired in 20 min. The mental fatigue of participants was assessed using scales such as the Stanford sleepiness scale, the Swedish occupational fatigue inventory (SOFI) [77], and the multidimensional fatigue inventory (MFI) [78]. Then, a 10-min resting-state EEG focused on the prefrontal cortex (PFC) was recorded using the Muse 2 headband. The HR and HRV were simultaneously documented. A standardized reaction time test was then administered via a dedicated computational interface to gauge mental alertness. This was followed by the PF induction phase, lasting 60 min. Participants began with a 30-min cycling task, segmented by intensities: low (10 min), medium (10 min), and high (10 min). Reaction times were evaluated after each segment. Then, they performed a 10-min jump rope exercise, followed by ball tasks such as squat throws and wall passes, for another 10 min.

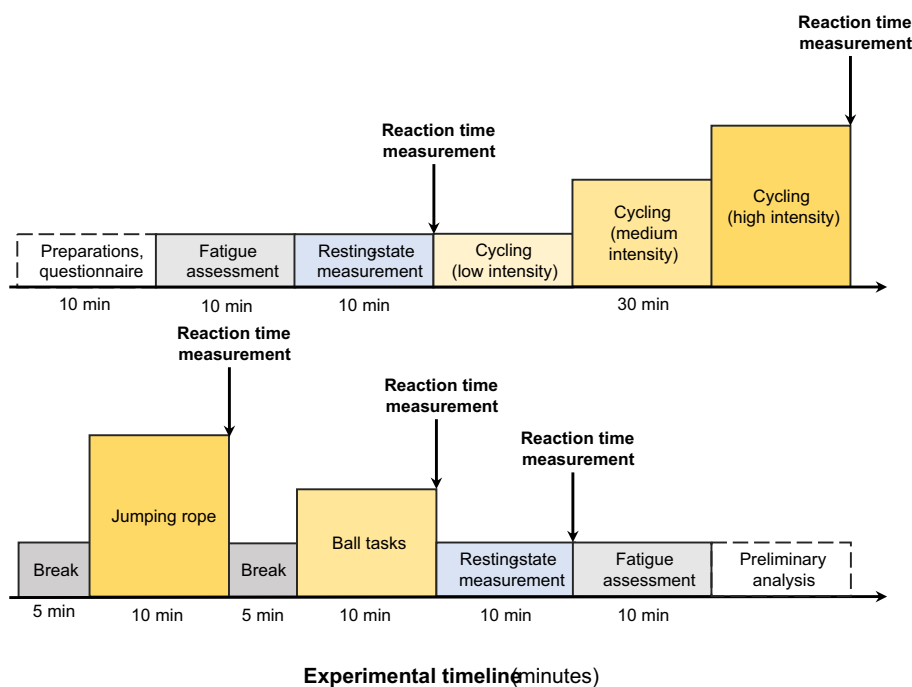


Fig. 2 Experimental procedure of the fatigue assessment

A 5-min recuperative interval was provided between each activity, and reaction times were assessed after each exercise.

Post-exertion data were acquired for 40 min. Participants re-administered the Stanford sleepiness scale, SOFI, and MFI to gauge the shifts in mental fatigue. A follow-up 10-min resting-state PFC EEG was then performed. The post-activity PPG and EEG signals were recorded. Lastly, a reaction-time test was conducted to detect any changes in cognitive responsiveness. The data collation and preliminary analysis were completed in approximately 10 min. All sensor data was curated and structured into a dataset. We ensured a consistent mapping of participant IDs across various metrics, including subjective ratings, EEG recordings, cardiovascular metrics, reaction times, and physical performance indicators. Figure 2 depicts the experimental procedure for the assessment.

Data preprocessing

The original physiological data extracted from sensors was processed in several stages to ascertain its appropriateness for the ensemble learning model. Missing data points, outliers, and erroneous readings, which could distort the model performed, were cleaned using data visualization and statistical analysis. The multi-stage processing involved several filters and techniques, each addressing different types of noise and artifacts. Firstly, the Hampel filter was deployed to correct for sensing failures and outliers typically resulting from transient spikes and artifacts. This filter operates by examining each data point relative to its neighbors, replacing those deviating by more than a threshold—set three times the median absolute deviation—with the median of surrounding values, thus protecting the integrity of physiological signals. Additionally, a third-order one-dimensional median filter was utilized to ensure data integrity. Unlike traditional linear filters, the median filter preserves the fidelity of the temporal sequence by replacing each data point with the median of a predefined window of neighboring points and mitigating the influence of singular aberrant values or spikes that may distort the true signal. This non-linear filtering approach was particularly adept at maintaining the edge features of signals, essential for the accurate characterization of time-dependent physiological states in both EEG and PPG data. Additionally, data-cleaning techniques were used to remove spikes and aberrant readings. For smoothing the signals while conserving significant features like peak heights and troughs, we employed the Savitzky-Golay filter. By fitting a polynomial to the moving window of data points, this filter refines the waveform without distorting its essential morphology. This step was critical for subsequent analytical steps such as feature extraction. We also incorporated a moving average filter to address short-term fluctuations due to random signal noise. By computing the mean within a moving window, this filter effectively mitigates the impact of transient noise while allowing the underlying trend of the data to be discerned with greater clarity.

Furthermore, baseline removal techniques were used to remove the direct current (DC) component and standardize the signal within a normal range. This enhances the signal quality for subsequent analytical steps. For EEG signals, we addressed the issue of baseline wander due to slow drifts by applying a high-pass filter with a cutoff frequency

lower than the frequencies of interest. This effectively removed the drifts while leaving the higher frequency brain wave components intact. In PPG signals, baseline wander is often attributable to respiration, body movements, and other physiological variations. To rectify this, we considered using techniques like ensemble empirical mode decomposition (EEMD), which adaptively removes baseline trends based on the signal’s intrinsic oscillatory modes. These baseline removal techniques were crucial for compensating for low-frequency non-stationarities, potentially induced by sensor movements or attachment issues. By standardizing the signals and aligning them closer to their zero reference line, we facilitated a consistent interpretation of the physiological data, which proved significantly beneficial when comparing temporal segments across individual subjects.

Data segmentation

Signal segmentation followed the preprocessing step. The primary challenge was to determine the optimal window size for fatigue detection, for which window sizes of 1–20 s, with 1-s increments, were tested. After determining the optimal window size, the cleaned and filtered data were segmented into overlapping windows corresponding to the optimal size. Feature extraction was first performed for data transformation. Features from PPG signals were retrieved, considering time and frequency domains, to compile a comprehensive dataset for analysis. Statistical measures, such as mean, variance, and standard deviation, were computed for various physiological parameters derived from the sensor data. Then, feature selection was performed using techniques such as the backward-elimination wrapper method. As the feature set was extensive, dimensionality reduction methods, such as principal component analysis, were used. Lastly, data transformation was performed using the Z-normalization technique to scale the selected features to a standardized range, such as [0, 1]. Similar steps were followed for EEG processing, emphasizing filtering and artifact removal, visual assessment, feature extraction from frequency bands, and data segmentation. The specifics of the final dataset are given in Table 2.

Proposed system architecture

To accurately predict PF and ME, a robust and complex system that uses DeepAR and TFT models is used. DeepAR, a probabilistic forecasting model developed by Amazon, uses long short-term memory (LSTM) cells to proficiently capture typical dynamics within time series and incorporates time-dependent covariates influencing the target variables. Conversely, the TFT model, designed for multivariate time-series forecasting, uses self-attention mechanisms, LSTM components, and gating mechanisms to decipher complex temporal dynamics within various time series. Both models accurately capture different dependencies and patterns within the dataset; therefore, an ensemble model is incorporated into the system. This model combines DeepAR and TFT prediction; it avoids overfitting of data and

Table 2 Details of the dataset

Dataset size	No. of variables	Variables	Time frame	Source
~ 79,200 s	11	HRV, HR, BVP, Gamma, Beta, Alpha1, Alpha2, Theta, Delta, PF, MF	11 × 2 h frequency = “1S”	Self-collected using EEG and PPG sensors

ensures robustness and excellent prediction performance. Thus, this system architecture can handle diverse and heterogeneous time-series data, which is essential for predicting PF and MF.

DeepAR model

DeepAR, a probabilistic forecasting model developed by Amazon, is aptly designed to handle multiple time series exhibiting common seasonal patterns and characteristics [79]. As shown in Fig. 3, the architecture leverages recurrent neural networks, namely the LSTM cells, as the foundation for the model architecture.

For the probabilistic forecast of an input vector x_t at time t , the hidden state of the LSTM, denoted as h_t , is computed using Eq. (1).

$$h_t = LSTM(x_t, h_{t-1}; \theta). \quad (1)$$

The likelihood $p(y_t|\theta)$ of observing a value y_t at a time t is parameterized by a neural network, as shown in Eq. (2).

$$p(y_t|\theta) = Distribution, \quad (2)$$

where θ represents the model parameters and NN denotes the neural network that maps the hidden state of the LSTM to the parameters of a chosen distribution. The training objective of DeepAR is to maximize the log-likelihood of the observed data, as shown in Eq. (3).

$$L(\theta) = \sum_t \log(p(y_t|\theta)). \quad (3)$$

DeepAR proficiently captures typical dynamics within a time series and comprehensively analyzes seasonality and trend patterns. Further, the model is equipped to incorporate time-dependent covariates that may influence the target variable. The LSTM architecture learns from the past values of the time series to predict future values. Moreover, the model outputs a probability distribution over the prospective values, enabling the quantification of uncertainty within the forecasts.

The implementation process of the DeepAR model is explained in Algorithm 1, in which the target variables denote the PF and MF scores. As such, the model employs additional covariates—HRV, HR, gamma, beta, alpha1, alpha2, theta, and delta brain waves—to augment the forecasting of PF and MF scores. During training, the architecture potentially deciphers the relationship between these covariates and the target variables. The model estimates uncertainty by generating a full distribution of probable outcomes—an advantageous feature for estimating fluctuating fatigue levels. The model also accurately captures intricate temporal dependencies in physiological signals and fatigue scores. Multiple time series were generated concurrently for all participants, i.e., multiple measurements were gathered from multiple individuals.

Algorithm 1 DeepAR algorithm

DeepAR Algorithm:
<ol style="list-style-type: none"> 1. Let X be the input time-series data. 2. Preprocess the time-series data. <ol style="list-style-type: none"> a. Normalize the time-series data to a suitable scale, typically between 0 and 1. b. If necessary, transform the time series to be stationary (mean and variance do not change over time). c. Split the time series into training and test sets. 1. Define the DeepAR model. <ol style="list-style-type: none"> a. Initialize the model with a specified number of layers and hidden units in each layer. b. Define the LSTM cells for each layer in the model. 2. Train the DeepAR model. <ol style="list-style-type: none"> a. Pass the training set through the DeepAR model. <ol style="list-style-type: none"> i. For each t step in the time series: <ol style="list-style-type: none"> 1. Pass the current and past values as the input to the LSTM cells. 2. The LSTM cells output a hidden state and cell state for the next t step. ii. Pass the final output of the LSTM cells through a dense layer to predict the parameters of the chosen distribution. b. Calculate the loss between the predicted distribution and the actual values. c. Backpropagate the loss through the model and update the model parameters. 3. Test the DeepAR model. <ol style="list-style-type: none"> a. Pass the test set through the trained DeepAR model. b. Calculate the loss between the predicted distribution and the actual values in the test set. 4. If the loss is acceptable \rightarrow the model is ready for forecasting. Else, adjust the model parameters (number of layers, hidden units, etc.) and go back to step 3. 5. For forecasting: <ol style="list-style-type: none"> a. Pass the entire time series (including future time steps with unknown values) through the trained DeepAR model. b. Output a distribution for each future time step, which can be used to generate forecast samples.

TFT model

The TFT model (Fig. 4) was particularly developed for forecasting multivariate time series using self-attention mechanisms, LSTM components, and gating mechanisms [80]. This model deciphers complex temporal dynamics, including both local and long-term dependencies within multiple time series. An integral attribute of the TFT model is its ability to

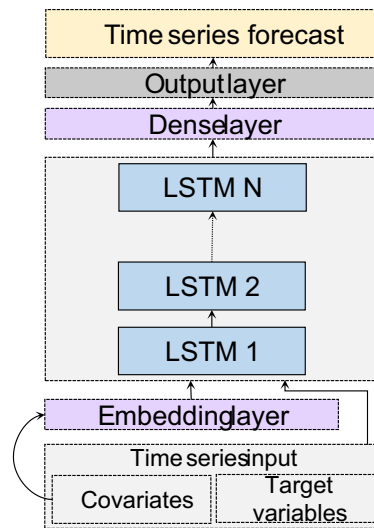


Fig. 3 DeepAR architecture

execute quantile forecasts, which enables it to predict outcomes across a continuum of probabilistic quantiles. These prediction intervals represent the range of values within which the actual fatigue scores are expected to fall with a certain probability, thereby generating a distributional forecast that encapsulates a spectrum of fatigue levels, each associated

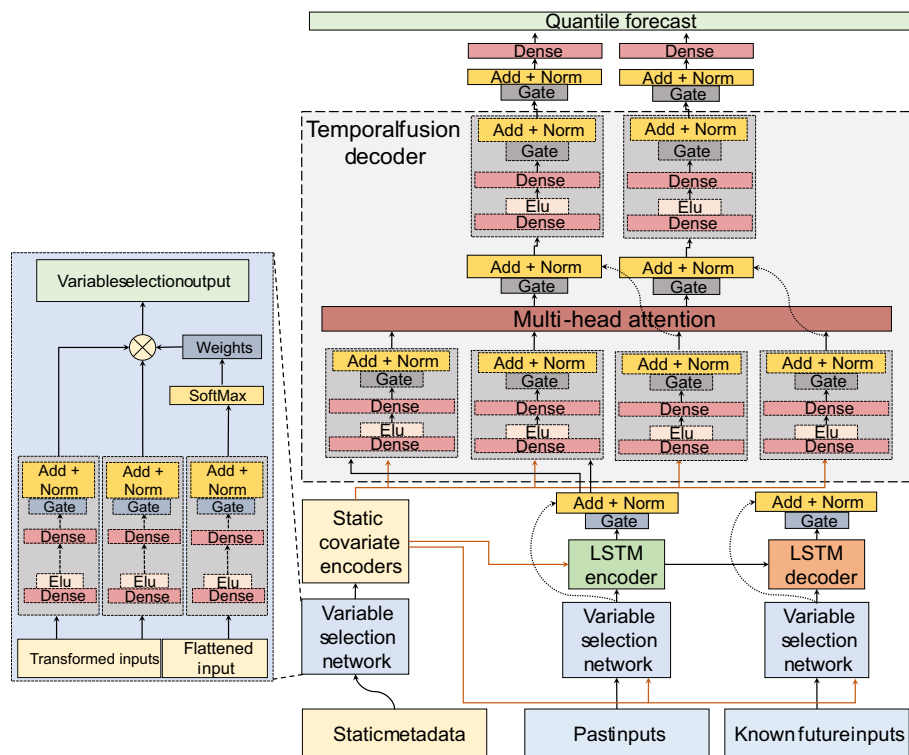


Fig. 4 Temporal fusion transformer architecture

with a specific quantile, rather than yielding a discrete point estimate. The ability to forecast across multiple quantiles is indispensable in the field of construction safety management, providing a nuanced understanding of potential fatigue scenarios. Such a probabilistic approach is instrumental in formulating comprehensive risk mitigation strategies by quantifying the uncertainty inherent in fatigue predictions.

The TFT model employs attention mechanisms to weigh the importance of different steps. Given an input sequence $X = \{x_1, x_2, \dots, x_t\}$, the self-attention mechanism computes a weighted sum of the inputs, as shown in Eq. (4).

$$A_t = \sum_{i=1}^t \alpha_{ti} x_i, \quad (4)$$

where the attention weight α_{ti} is determined using Eq. (5).

$$\alpha_{ti} = \frac{\exp(\text{score}(x_t, x_i))}{\sum_{j=1}^t \exp(\text{score}(x_t, x_j))}. \quad (5)$$

The score function can be a dot product or other similarity measure between input vectors. The TFT model also employs gating mechanisms to control the flow of information, as shown in Eq. (6).

$$G_t = \sigma(W_g x_t + b_g). \quad (6)$$

Here, σ is the sigmoid activation function, and W_g and b_g are the gating parameters. Its capability to assign varied weights to different points in the time series according to their relevance enhanced the model's proficiency in understanding intricate interdependencies between variables across different temporal contexts.

The TFT model can adeptly handle diverse and heterogeneous time-series data. This is a characteristic of our custom dataset, where training is performed on a multitude of time series, each demonstrating unique attributes and distributions. Three types of features are distinctive to the TFT model: time-dependent data with known future inputs, time-dependent data known only up to the present, and static or categorical variables. These features allow the model to flexibly integrate various types of auxiliary information for the forecasting and modeling of multivariate time series. The TFT model is vital for our CEFAS model as it captures the local and long-term temporal dependencies between MF and PF scores and physiological variables, including HRV, HR, and gamma, beta, alpha1, alpha2, theta, and delta brain waves. These physiological signals display complex temporal dynamics, necessitating the development of a model that can accurately interpret data beyond the capabilities of traditional LSTM and RNN models. The TFT model can accommodate an array of signal types and ranges, including continuous

variables such as HRV and HR. Algorithm 2 illustrates the computational steps of the TFT model.

Algorithm 2 TFT algorithm

TFT Algorithm	
1.	Let X be the input time-series data.
2.	Preprocess the time-series data.
a.	Normalize the time-series data to a suitable scale, typically between 0 and 1.
b.	If necessary, transform the time series to be stationary (mean and variance do not change over time).
c.	Split the time series into training and test sets.
3.	Define the TFT model.
a.	Initialize the model with a specified number of layers and hidden units in each layer.
b.	Define the data processing layers:
i.	Variable selection networks to identify relevant features.
ii.	Gated linear units (GLUs) for input processing.
iii.	Temporal processing layers, including positional encoding and temporal convolutional layers.
c.	Define the Transformer-style self-attention mechanism.
d.	Define the gated skip connections to facilitate information flow.
e.	Define the output layer to predict the future values of the time series.
4.	Train the TFT model.
a.	Pass the training set through the TFT model.
i.	Process the input data through the variable selection networks, GLUs, and temporal processing layers.
ii.	Pass the processed data through the self-attention mechanism and gated skip connections.
iii.	Predict the future values of the time series with the final output layer.
b.	Calculate the loss between the predicted values and the actual values.
c.	Backpropagate the loss through the model and update the model parameters.
5.	Test the TFT model.
a.	Pass the test set through the trained TFT model.
b.	Calculate the loss between the predicted values and the actual values in the test set.
6.	If the loss is acceptable, the model is ready for forecasting.
	Else , adjust the model parameters (number of layers, hidden units, etc.) and go back to step 3 .
7.	For forecasting:
a.	Pass the entire time series (including future time steps with unknown values) through the trained TFT model.
b.	Output a prediction for each future time step.

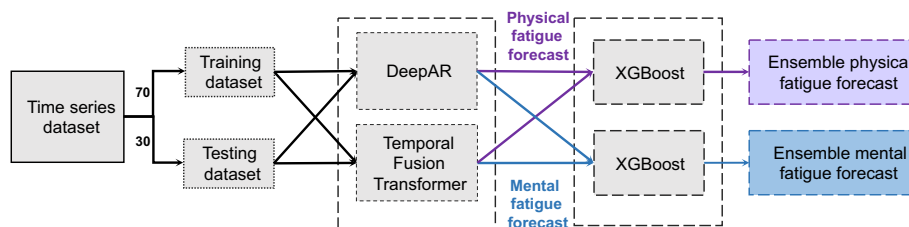


Fig. 5 Schematic of the ensemble learning model topology

Ensemble model

The ensemble model uses a sophisticated machine learning technique that leverages the strengths of the base models [73, 74], DeepAR and TFT, to analyze time-series datasets and generate more accurate forecasts for PF and MF. Herein, the time-series dataset was first used to train the base models, DeepAR and TFT, which then separately created forecasts for PF and MF. Then, a two-step meta-learning process was initiated: the PF forecasts generated by the base models were used as inputs for one meta learner, and the mental fatigue forecasts were channeled into another meta learner. This dual meta-learning strategy allowed for a more refined final forecast. In the ensemble learning architecture, forecasts from base models, DeepAR and Temporal Fusion Transformer, are integrated using XGBoost meta-learners to yield refined ensemble predictions for both physical and mental fatigue, thereby mitigating overfitting risks as opposed to a single deep learning model. For contradictory forecasts, the meta-model can ensure a consensus among these different predictions. The ensemble learning model is illustrated in Fig. 5.

The eXtreme gradient boosting (XGBoost) algorithm [81] sophisticatedly adopts the gradient boosting algorithm to enhance the efficiency of expansive machine learning tasks. It originates from the theoretical framework of the decision tree algorithm, systematically refining model predictions through iterative learning from previous errors. Given a differentiable loss function $L(y, \hat{y})$, where y is the true value and \hat{y} is the predicted value, XGBoost constructs an additive model as shown in Eq. (7).

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i), \quad (7)$$

where $f_t(x)$ is the decision tree and t is the number of iterations. The primary objective is to select f_t that minimizes the overall loss, as shown in Eq. (8).

$$obj^{(t)} = \sum_{i=1}^n L, \quad (8)$$

where $\Omega(f_t)$ is the regularization term that imposes a penalty on the model complexity and prevents overfitting. It incorporates both L1 (Lasso) and L2 (Ridge) regularization, rendering the model more robust to potential overfitting. In terms of computational efficiency, XGBoost is designed for parallel and distributed computing, ensuring rapid model training. The implementation of the ensemble model is given in Algorithm 3, wherein XGBoost is utilized as the meta-learner.

Algorithm 3 Proposed ensemble learning algorithm

Proposed Ensemble Learning Algorithm
<ol style="list-style-type: none"> 1. Let X be the input time-series data. 2. Preprocess the time-series data. <ol style="list-style-type: none"> a. Normalize the time-series data to a suitable scale between 0 and 1. b. If necessary <ol style="list-style-type: none"> i. Transform the time series to be stationary (mean and variance do not change over time). c. Split the time series into training, validation, and test sets. 3. Define the TFT model and DeepAR model. <ol style="list-style-type: none"> a. Initialize both models with a specified number of layers and hidden units in each layer. b. Define the data processing layers for each model. 4. Train the TFT model and DeepAR model. <ol style="list-style-type: none"> a. Pass the training set through both models. b. Calculate the loss between the predicted values and the actual values for each model. c. Backpropagate the loss through each model and update the model parameters. 5. Validate the TFT model and DeepAR model. <ol style="list-style-type: none"> a. Pass the test set through the trained models. b. Calculate the loss between the predicted values and the actual values in the test set for each model. 6. If the loss is acceptable for both models → the models are ready for ensemble learning. Else → Adjust the model parameters (number of layers, hidden units, etc.) and go back to step 3. 7. For ensemble learning predictions: <ol style="list-style-type: none"> a. Combine the predictions from both models. <ol style="list-style-type: none"> i. Pass the entire time series through the trained TFT model and DeepAR model. ii. Output a prediction for each future time step by each model. iii. Combine the predictions from both models using the ensemble method. iv. Let the combined predictions from the TFT model and DeepAR model represent the final output of the ensemble learning model as represent the forecasted values of the time series.

Experimental setup

The experimental setup uses a distinct combination of wearable technologies and an ensemble learning model for comprehensive fatigue monitoring. The research was conducted in a high-performance computational environment powered by an AMD Ryzen 3990X 64-core processor and an NVIDIA GeForce RTX 3090 integrated GPU, complemented by 256 GB of RAM and a 1 TB SSD single drive. This system ensures the seamless execution of complex algorithms and the efficient handling of large datasets, thereby laying a solid groundwork for the analytical processes. Multiple sensors, including the Muse 2 headband electroencephalography sensor and MAX30102 heart-rate and pulse

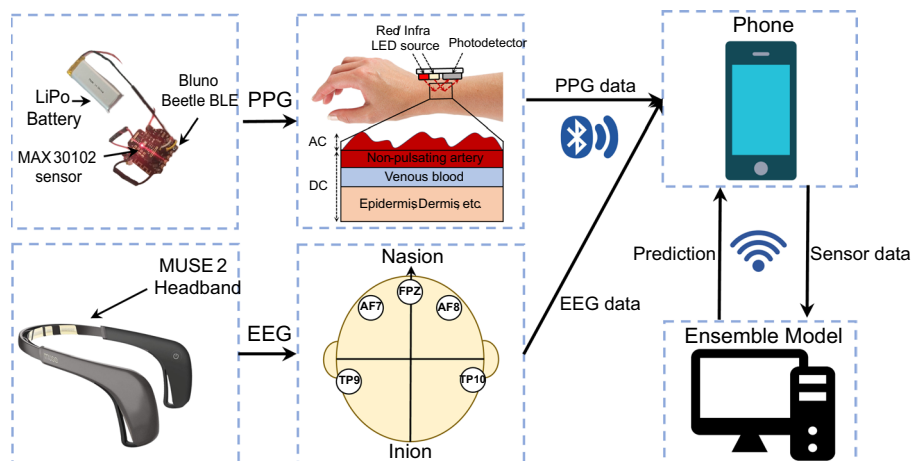


Fig. 6 Experimental setup and process

oximetry sensor, were integrated to capture physiological signals. The setup is shown in Fig. 6.

To streamline data collection and processing, a Bluno Beetle Bluetooth low-energy (BLE) microcontroller is incorporated into the wearable device because it is lightweight and offers high power and programmability. This setup was proposed for real-time monitoring of individuals’ health status. Moreover, it was easy and comfortable to use, which is a critical factor in ensuring its sustained application in construction environments.

Prototype development

The prototype was developed via a structured pathway by integrating wearable technologies and an ensemble learning model in five critical steps. First, a high-performance computational environment was established. Then, the Muse 2 headband and MAX30102 sensor were integrated into the system for EEG readings and heart rate and pulse oximetry monitoring, respectively. A microcontroller was then installed to

Table 3 Configuration and use of sensors and microcontroller in the prototype

Sensor type	Configuration		Use in prototype
Muse 2 headband (EEG sensor)	Positioning	Forehead	Captures electroencephalography data to monitor mental fatigue
	Sampling rate	256 Hz	
	Sample depth	12-bit	
	Connection method	Wireless communication via Bluetooth	
MAX30102 sensor	Positioning	Wrist	Monitors HR and pulse oximetry, providing data on PF and overall health status
	Connection method	two-wire setup with the microcontroller	
	Sampling rate	10 Hz	
Bluno Beetle BLE Microcontroller	Connection method	Bluetooth BLE (BT 4.0)	Facilitates the streamlined collection of data from the wearable sensor
	Dimensions	28.8 mm × 33.1 mm	
		10 g	
	Microcontroller	ATmega328 operating at a clock frequency of 16 MHz	

streamline data collection and capture physiological signals in real-time. The first stage leveraged the ensemble learning model to process the acquired data, facilitating the extrapolation of PF and MF scores from the sensor-derived data.

Sensors and microcontroller The developed prototype comprised various sensors. The Muse 2 headband electroencephalography sensor was used to obtain EEG signals from TP9, AF7, AF8, and TP10 channels. The MAX30102 heart-rate and pulse oximetry sensor was placed in the wristband to record PPG signals.

The Bluno Beetle BLE microcontroller was also integrated into the wristband configuration to regulate the collection and processing of sensor data. This microcontroller, specifically tailored for wearable technology applications, is lightweight and offers enhanced power and programmability; therefore, it is used herein. A lithium-ion battery was also used as a compact and dependable power solution. The Bluno Beetle BLE wirelessly transmits the collected data to the computer, providing flexibility in data management and processing. Operating as a Bluetooth master or slave, this module permitted the wristband to engage in wireless serial port communication. With a transmission range extending up to 10 m, this BLE technology enables the real-time transmission of sensor data to a smartphone, allowing the wearer to manage the wristband through a user interface. These sensors and modules (Table 3) constitute a complex network that can accurately capture and process physiological signals. A comprehensive fatigue score can be determined to ultimately enhance the safety and health of the wearers.

Prototype validation

The prototype was verified comprehensively to affirm its functionality in real-time operational conditions. The prototype was subjected to testing, wherein sensor data was obtained. The fatigue predictions generated by the proposed model were compared with the established fatigue scoring benchmarks. The performance of the ChronoEnsemble Fatigue Analysis System was evaluated using metrics such as mean square error (MSE), mean absolute scale error (MASE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), normalized deviation (ND), and mean scaled interval score (MSIS). Prototype validation revealed its salient outcomes, such as high correlation coefficients between the model-predicted outcomes and the actual fatigue scores, as well as the precise tracking of PF and MF fluctuations in real-time conditions. The ensemble model effectively identified the fatigue scores within the sensor data, indicating its proficiency in recognizing and responding to the complex combination of physiological signals that contributed to fatigue. The validation process was predominantly focused on performance metrics, but it was instrumental in detecting areas that required potential enhancements, thereby ensuring its continuous development.

Results and discussion

Performance evaluation metrics

The efficacy of the ensemble model was assessed using performance metrics such as MSE, RMSE, MASE, MAPE, sMAPE, ND, and MSIS. The MSE denotes the mean of

the squared discrepancies between the predicted and true values, indicating the error variance of the model. MSE is calculated as follows (Eq. (9)).

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2, \quad (9)$$

where y and \hat{y} denote the actual and predicted values, respectively.

The RMSE is a standard metric used in supervised learning to measure the accuracy of predictive models. To determine the RMSE, the residual for each observation, defined as the discrepancy between the predicted and observed values, is determined. Then, the squared norm of each residual is computed. The mean of these squared residuals is then calculated, and its square root is calculated to obtain the RMSE. True measurements for each predicted data point are required for calculating the RMSE. The RMSE is represented in Eq. (10):

$$RMSE = \sqrt{\sum \frac{(y - \hat{y})^2}{n}}, \quad (10)$$

where n denotes the total number of observations, y denotes the true values, and \hat{y} denotes the predicted values corresponding to the observations.

The MASE is a metric used to compare the prediction accuracies of different methods. It is computed by normalizing the mean absolute error by the mean absolute difference of successive actual values, thus facilitating model comparisons. It is mathematically expressed in Eq. (11).

$$MASE = \frac{\frac{1}{n} \sum |y - \hat{y}|}{\frac{1}{n-1} \sum |y[t] - y[t-1]|}, \quad (11)$$

where n is the total number of observations; y and \hat{y} are the true and predicted values at a specific t time point, respectively; $y[t]$ and $y[t-1]$ are the actual values at time t and $t-1$, respectively.

The MAPE is the average of the absolute percentage discrepancies between the predicted and actual values, as shown in Eq. (12). It is used as a measure of the prediction accuracy of a forecasting model.

$$MAPE = \frac{1}{n} \sum \left(\left| \frac{y - \hat{y}}{y} \right| * 100 \right), \quad (12)$$

where n is the number of observations in the dataset and y and \hat{y} denote the true and predicted values, respectively, at a particular time point.

The sMAPE is a symmetric version of MAPE (Eq. (13)), wherein the mean absolute percentage discrepancy is normalized by the sum of the predicted and absolute values of the actual value. It also addresses some issues encountered with the MAPE.

$$sMAPE = \frac{2}{n} \sum \left(\frac{y - \hat{y}}{(|y| + |\hat{y}|)} \right), \quad (13)$$

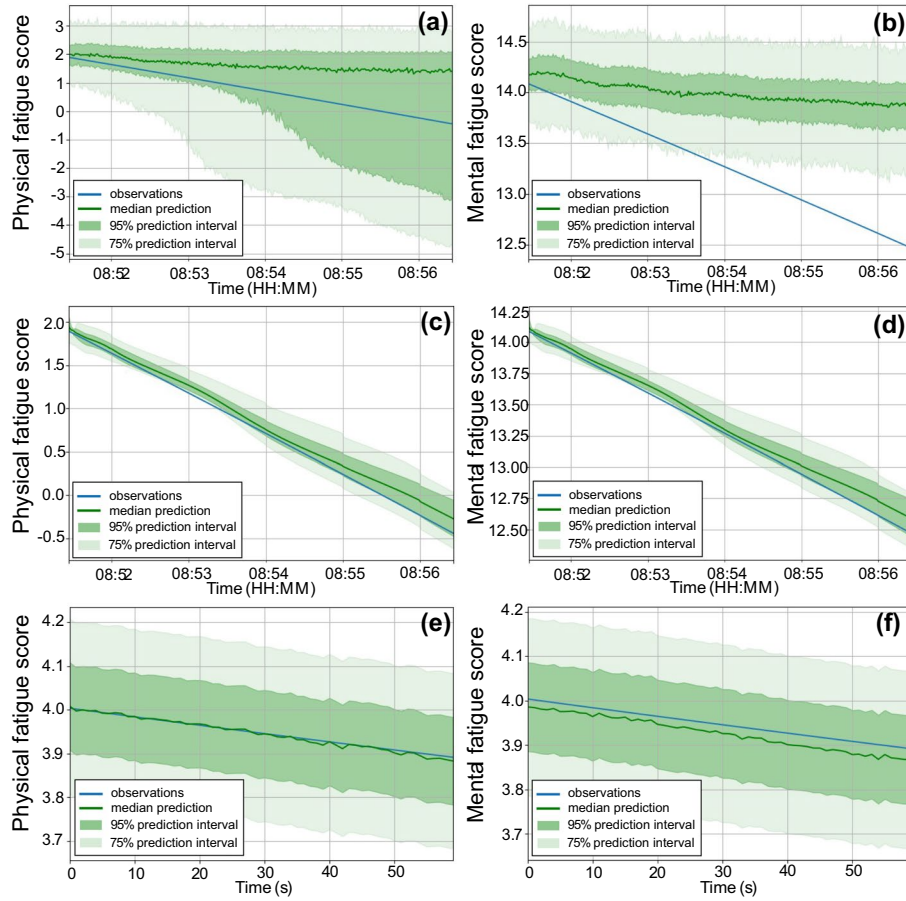


Fig. 7 Visual evaluation of the physical and mental fatigue forecasts on the test dataset. **a** DeepAR physical fatigue forecast, **b** DeepAR mental fatigue forecast, **(c)** temporal fusion transformer physical fatigue forecast, **(d)** temporal fusion transformer mental fatigue forecast, **(e)** CEFAS physical fatigue forecast, and **(f)** CEFAS mental fatigue forecast

where n denotes the total number of observations in a dataset and y and \hat{y} denote the actual and predicted values at a specific time point, respectively.

The ND (Eq. (14)) scales the disparity between the predicted and true values by the range of the actual values:

$$ND = \frac{\sum |y - \hat{y}|}{\sum |y|}, \tag{14}$$

where y and \hat{y} denote the true and predicted values at a distinct time point, respectively.

The MSIS considers the average width of the predicted interval normalized by the mean absolute discrepancy between the predicted and actual values, as shown in Eq. (15). In other words, it considers the average width of the predicted interval normalized by the mean absolute error.

$$MSIS = \frac{1}{n} \frac{\sum (upperbound - lowerbound)}{\sum |y - \hat{y}|}, \tag{15}$$

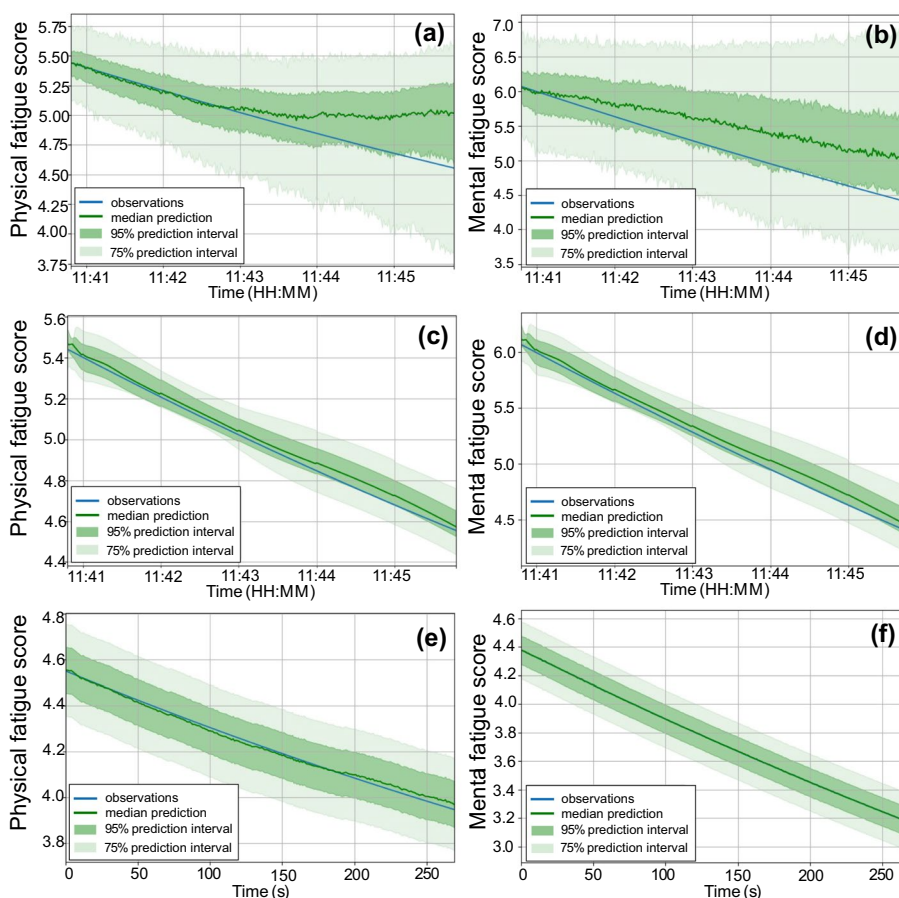


Fig. 8 Visual evaluation of the physical and mental fatigue forecasts on the training dataset. **a** DeepAR physical fatigue forecast, **b** DeepAR mental fatigue forecast, **c** temporal fusion transformer physical fatigue forecast, **d** temporal fusion transformer mental fatigue forecast, **e** CEFAS physical fatigue forecast, and **f** CEFAS mental fatigue forecast

where the lower and upper bounds refer to the bounds of the prediction interval, n is the number of observations in the dataset, y is the true value at a specific time point, and \hat{y} is the predicted value at a specific time point.

Model performance

The ensemble learning model is comprehensively evaluated here. Its performance was analyzed in terms of accuracy, precision, and resilience to variations and deviations using several evaluation metrics. The results were juxtaposed with those of the DeepAR and TFT models to confirm the superior efficacy of the proposed model over these standalone models.

The forecasts of physical and mental fatigue on the test dataset by DeepAR, as illustrated in Fig. 7a, b, exhibit notable fluctuations. In Fig. 7a, the physical fatigue score seems to be relatively stable over a short period of time, as the observation lies within the 95% prediction interval. However, the broad prediction intervals suggest that the model only captures a general trend, and there is uncertainty in the forecasts. Although the mental fatigue score is decreasing over time (Fig. 7b), the predictions lie below the

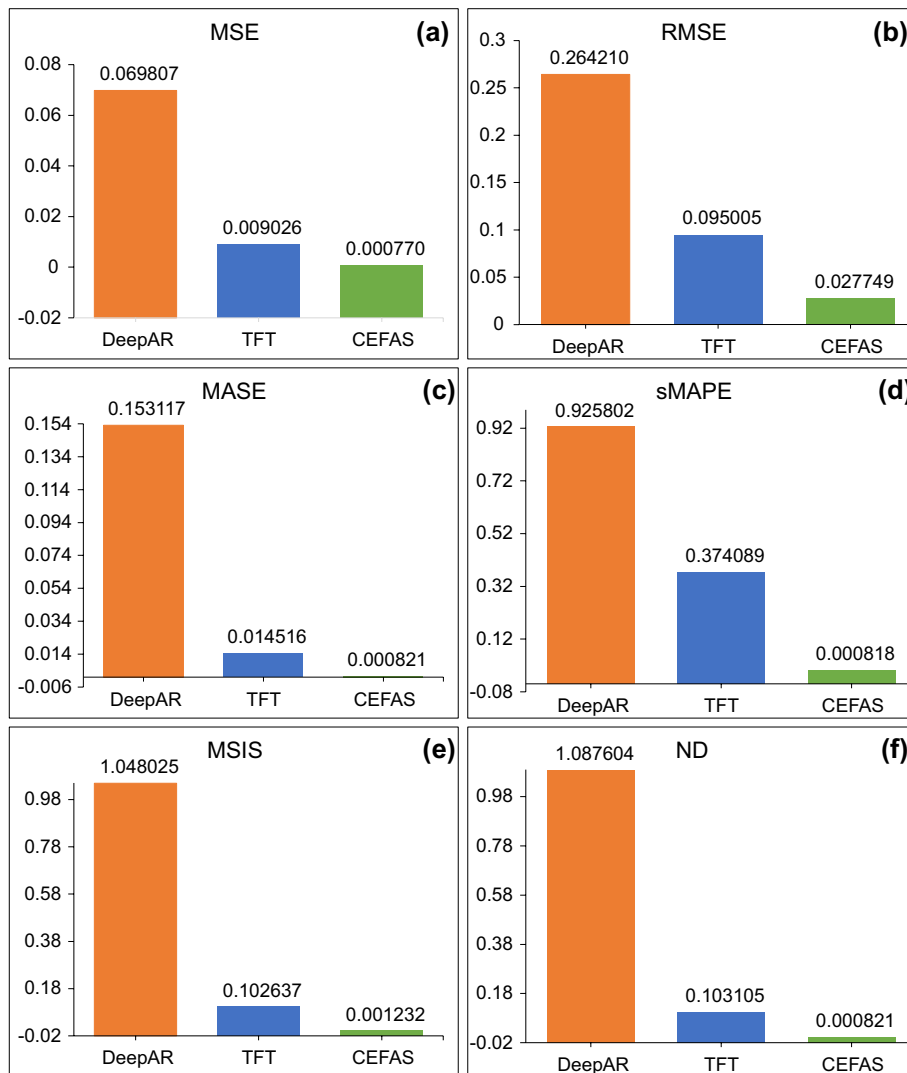


Fig. 9 Comparative analysis of forecasting performance metrics for DeepAR, TFT, and CEFAS models in predicting physical fatigue. **a** MSE, **b** RMSE, **c** MASE, **d** sMAPE, **e** MSIS, and **f** ND

observations. Thus, the model systematically underestimates the mental fatigue score. In Fig. 7b, the prediction intervals are also relatively wide, indicating uncertainty again. This highlights the challenges that standalone models might encounter when attempting to understand intrinsic data patterns. The Temporal Fusion Transformer, shown in Fig. 7c, d, provides forecasts with a more consistent downward trajectory. The model’s median prediction closely follows the observations, and the model provides narrower prediction intervals compared to the DeepAR predictions, which is indicative of its ability to manage temporal correlations well and higher confidence in the forecasts. Figure 7e, f capture the outcomes from CEFAS, representing the ensemble learning predictions. As shown in Fig. 7e, f, the ensemble model is able to forecast the physical as well as the mental fatigue precisely and give stable predictions over time. For the mental fatigue forecasts, the CEFAS slightly underestimates the fatigue score. The tight

prediction intervals suggest the model's confidence in the predictions for both PF and MF. Overall, the CEFAS produces a more stable prediction by balancing the errors from the DeepAR and TFT models. The combined inputs of the ensemble model led to a reduction in variance.

The visual evaluation of the forecasts on the training dataset is illustrated in Fig. 8. The observations within Fig. 8a, b show significant variation from the actual values, even when the forecasts were made on the training dataset. This could indicate the DeepAR model's sensitivity to the immediate past and potential overfitting to noise within the data. The prediction intervals also indicate a high level of uncertainty in the physical fatigue levels. The physical fatigue scores in Fig. 8a oscillate with a narrower band, suggesting fewer volatile traits, whereas the mental fatigue scores in Fig. 8b display wider swings, underscoring that the model is unable to capture the mental fatigue scores precisely due to the inherent nature of mental fatigue dynamics. TFT (Fig. 8c, d) provides a more confident forecast for both physical and mental fatigue prediction with a strong trend and narrow intervals, which indicates that it clearly understands the temporal dynamics of the data. The ensemble learning model (Fig. 8e) shows even narrower intervals for physical fatigue scores than either of the base models, which demonstrates the capability of the ensemble to synthesize information and correct over- or under-estimations by individual models when it is necessary. As shown in Fig. 8f, the narrow intervals of the prediction denote the ensemble model's ability to learn predictive cues from both DeepAR and TFT models. The ensemble effectively integrates the DeepAR and TFT models' traits to capture the probabilistic nature of the data and to understand complex temporal patterns.

A comparative analysis of the physical fatigue forecasting performance was conducted, as illustrated in Fig. 9. The MSE and RMSE values were determined to evaluate the physical fatigue prediction accuracy of the proposed model. The ensemble model, represented by the CEFAS, consistently registers lower MSE values of 0.0007 for PF (Fig. 9a). Such outcomes can be attributed to the ensemble's iterative refinement process, inherent to gradient boosting mechanisms. The DeepAR model, registering an MSE value of 0.0698 for PF, potentially suggests its inherent limitations in deciphering certain intricate data dynamics.

The TFT model, while achieving an MSE of 0.0639 for PF, marginally surpassing the DeepAR, still does not approach the precision of the ensemble model's results. The result likely stems from the ensemble's ability to combine and optimize diverse model strengths, mitigating individual weaknesses and enhancing overall precision. With RMSE metrics of 0.0277 for PF, as illustrated in Fig. 9b, the proposed approach effectively amalgamates various analytical perspectives, emphasizing its capability of producing consistent forecasting results. The RMSE values achieved by the DeepAR model, 0.2642 for PF, reveal potential limitations in its alignment, as a higher RMSE indicates a greater discrepancy between the predicted values and actual observations. Meanwhile, TFT, despite a commendable RMSE of 0.2642, doesn't attain the precision observed in the ensemble forecasts.

In the context of sMAPE (Fig. 9c), the ensemble model's result of 0.0054 for PF articulates robustness across diverse scales. TFT, while outperforming DeepAR with a MASE value of 0.056576 for PF, may indicate a proclivity towards overfitting

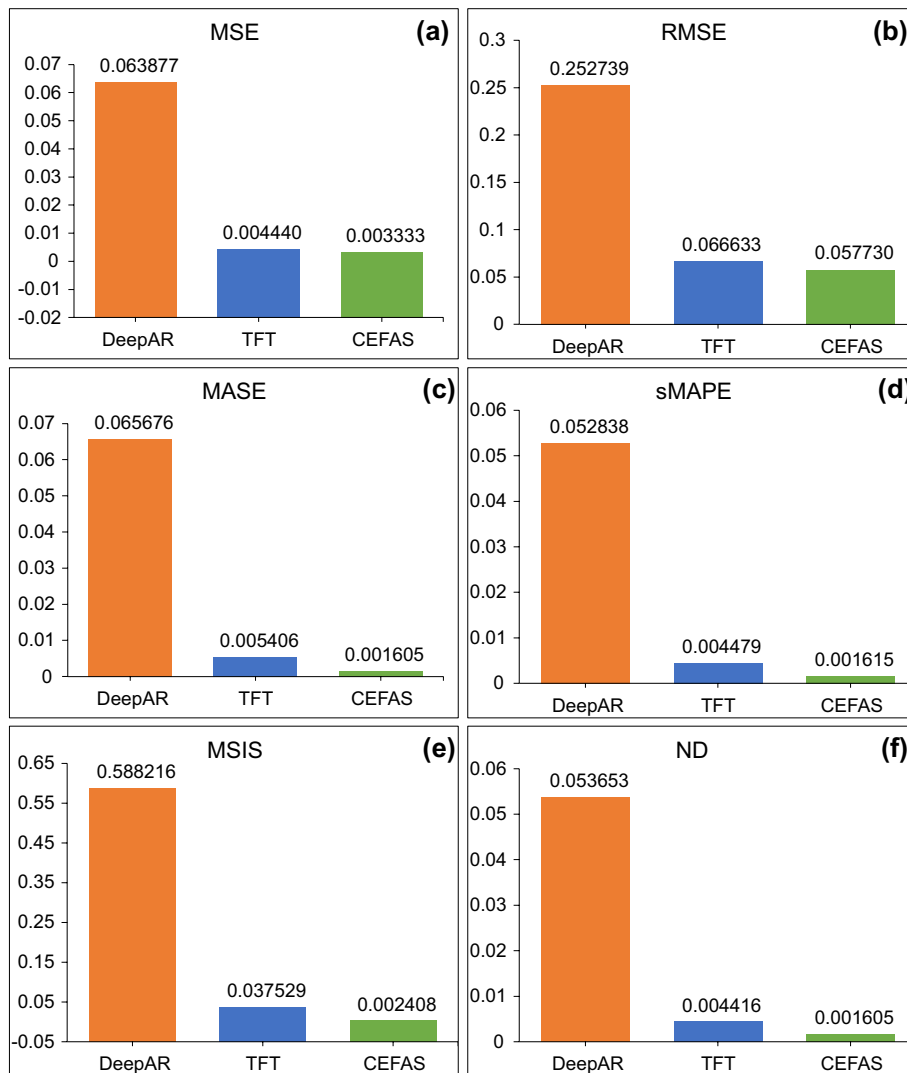


Fig. 10 Comparative analysis of forecasting performance metrics for DeepAR, TFT, and CEFAS models in predicting mental fatigue. **a** MSE, **b** RMSE, **c** MASE, **d** sMAPE, **e** MSIS, and **f** ND

temporal patterns. DeepAR’s escalated MASE metric of 0.1531 for PF elucidates its broader inefficacies in addressing scaled discrepancies. The MASE values underscore the proposed approach’s advanced forecasting capabilities and its ability to minimize percentage and scaled errors.

In terms of sMAPE, as reflected in Fig. 8d, the 0.0008 value for PF signifies that the predicted values are exceptionally close to the actual observations as compared to the standalone models. The low sMAPE value of the ensemble model emphasizes its consistency in handling the intricacies of the data and capturing underlying data patterns. TFT’s performance, indicated by a sMAPE of 0.3740 for PF, suggests intermittent challenges in deciphering unpredictable data trajectories. DeepAR, with its considerably higher sMAPE value of 0.9258 for PF, reiterates its vulnerabilities in symmetric error capture.

The ensemble model, boasting MSIS metrics (Fig. 9e) of 0.0012 for PF and an ND value (Fig. 9f) of 0.0008 for PF prediction, proves its expansive comprehension and reliability of its prediction intervals. The MSIS values suggest that the ensemble model's predictions are more reliable, given the narrow intervals they operate within. The ensemble model exhibits the lowest ND scores, indicating the model's increased resilience to variations and exemplary alignment with the actual observations. The minimal deviation proves that the ensemble model is not only accurate on an absolute scale but also consistently accurate relative to the actual data's range.

TFT's metrics, encompassing an MSIS value of 0.0375 and an ND of 0.0536 for PF, while surpassing DeepAR, elucidate occasional limitations in capturing data nuances. In comparison with the standalone models, the significantly lower ND values of the ensemble model underline its superior prediction fidelity, as it manages to maintain constant alignment with the actual data.

Figure 10 shows a comparative analysis of forecasting the performance metrics for the three models. The lower MSE value (Fig. 10a) indicates a closer fit between the observed values and the prediction. The CEFAS has the lowest value for MF prediction with 0.0033, which is due to its potential to correct the errors in the base models' forecasts and minimize the sum of squared deviations. As seen in the figure, TFT also performs well with an MSE value of 0.0044, as the model can leverage temporal patterns effectively. The DeepAR's high MSE of 0.0639 suggests that the model is unable to capture the complexity of mental fatigue as effectively as the others.

Similarly, the RMSE (Fig. 10b) values confirm the previous results, with CEFAS showing the highest accuracy with a value of 0.0577. There is a substantial difference between the MASE scores (Fig. 10c) of the base models and the ensemble model. The MASE value of 0.0016 for CEFAS indicates that the absolute errors from the CEFAS forecasts are much smaller than those from the naïve model. This is because the gradient boosting nature of the XGBoost meta-learner enables the capture of nonlinear relationships and interactions between the features, thus effectively predicting mental fatigue levels. The TFT model uses attention mechanisms to weigh different parts of the time-series data differently, potentially allowing it to focus on more relevant past information when predicting future points, as proven by a low 0.0054 MASE score. The DeepAR MASE value is the highest with 0.0657, which implies that for specific nuances of the mental fatigue data, its method for forecasting is less efficient than the ensemble learning model.

The sMAPE (Fig. 10d) values suggest CEFAS is the most accurate, with the lowest percentage error of 0.0016, indicating its strong predictive capabilities and robustness to outliers in mental fatigue forecasting. TFT also performs well with a value of 0.0045, showcasing its efficient time-series modeling, likely due to its attention mechanisms. DeepAR's higher 0.0528 sMAPE value implies the model is less precise in this context, potentially due to the complex nature of mental fatigue that may not align well with DeepAR's probabilistic autoregressive approach.

The MSIS (Fig. 10e) measures the accuracy of prediction intervals. CEFAS's very low MSIS value of 0.002408 indicates its interval forecasts are both narrow and accurate, reflecting high confidence in its predictions. TFT has a higher MSIS score of 0.037529, suggesting its prediction intervals are less precise than CEFAS's but still relatively

Table 4 Comparative analysis of average performance metrics between the proposed model and existing prediction methods

Refs.	Method	PF	MF	MSE	RMSE	MASE	MAPE	sMAPE	ND	MSIS
[82]	SONFIN	Yes	Yes	-	0.360	-	-	-	-	-
[83]	GELM	Yes	Yes	-	0.0712	-	-	-	-	-
[84]	ANN	Yes	No	-	0.857	-	0.0128	-	-	-
[85]	Multimodal DAE	No	Yes	-	0.094 ± 0.017	-	-	-	-	-
[86]	DANN	No	Yes	-	0.1427	-	-	-	-	-
[86]	ADDA	No	Yes	-	0.1405	-	-	-	-	-
[87]	DResNet	No	Yes	-	0.1420 ± 0.0402	-	-	-	-	-
[87]	DG-DANN	No	Yes	-	0.1470 ± 0.0444	-	-	-	-	-
[88]	LSTM-CapsAtt	No	Yes	-	0.1089 ± 0.0696	-	-	-	-	-
[89]	FedLSTM	Yes	Yes	0.4	0.6124	-	0.2499	-	-	-
[89]	FedGRU	Yes	Yes	0.3864	0.6026	-	0.2642	-	-	-
[90]	DNNSN	No	Yes	-	0.09	-	-	-	-	-
[91]	Regression	Yes	No	-	0.10	-	-	-	-	-
[91]	ANN	Yes	No	-	0.24	-	-	-	-	-
[92]	LSTM	Yes	No	-	0.0734	-	0.0577	-	-	-
[92]	BiLSTM	Yes	No	-	0.0735	-	0.0585	-	-	-
[92]	GRU	Yes	No	-	0.0739	-	0.0585	-	-	-
[92]	BiGRU	Yes	No	-	0.0782	-	0.0607	-	-	-
[93]	PSO-SVR	Yes	No	-	-	-	0.1153	-	-	-
[93]	ELM	Yes	No	-	-	-	0.1977	-	-	-
[94]	Regression with MMD and GRU	No	Yes	-	0.27	-	-	-	-	-
[95]	CAE + CNN + RNN	No	Yes	-	0.08	-	-	-	-	-
CEFAS	Ensemble learning model	Yes	Yes	0.0021	0.0427	0.0012	0.0012	0.0012	0.0012	0.0018

SONFIN: Self-Organizing Neural Fuzzy System, GELM: Graph regularized Extreme Learning Machine, DANN: Domain-Adversarial Neural Network, ADDA: Adversarial Discriminative Domain Adaptation, DResNet: Deep Residual Network, DG-DANN: Domain Generalization using Domain Adversarial Neural Networks, LSTM-CapsAtt: Long Short-Term Memory with Capsule Attention, FedLSTM: Federated Long Short-Term Memory, FedGRU: Federated Gated Recurrent Unit, DNNSN: Double-layered Neural Network with Subnetwork Nodes, DAE: Deep AutoEncoder, ANN: Artificial Neural Network, LSTM: Long Short-Term Memory, PSO-SVR: Particle Swarm Optimization-Support Vector Regression, BiLSTM: Bidirectional LSTM, BiGRU: Bidirectional GRU, CAE: Convolutional AutoEncoder, CNN: Convolutional Neural Network, RNN: Recurrent Neural Networks

informative. DeepAR's much higher MSIS of 0.588216 implies its prediction intervals are quite wide, indicating less certainty in its forecasts.

In terms of the ND of the forecasts (Fig. 10f), a lower score means a model's predictions are closer to actual values. CEFAS's lowest ND score of 0.001605 indicates it has the smallest deviation from the actual observations in predicting MF, likely due to its efficient handling of complex patterns in the data. TFT has a slightly higher ND score of 0.004416, showing good but less precise predictions than CEFAS. DeepAR's highest ND score of 0.053653 suggests it is less accurate and less effective in capturing the data dynamics related to mental fatigue. The tight alignment between the predictions and the observations suggests that the ensemble model is robust, translating to increased resilience against data variations, which is crucial for time-series forecasting tasks where the data can exhibit intricate temporal patterns.

Table 4 presents a comprehensive comparative analysis of the proposed ensemble model, CEFAS, against various state-of-the-art methods for fatigue prediction. The proposed CEFAS model demonstrates superior performance across multiple metrics compared to existing prediction methods, exhibiting an average MSE of 0.0021, RMSE of 0.0427, and MAPE of 0.0012.

Among the compared methods, the Self-Organizing Neural Fuzzy System (SONFIN) [82] and Graph regularized Extreme Learning Machine (GELM) [83] are capable of predicting both PF and MF. However, their RMSE values of 0.360 and 0.0712, respectively, are higher than those of CEFAS, indicating lower prediction accuracy. Similarly, the ANN [84] and the regression method with Maximum Mean Discrepancy (MMD) and Gated Recurrent Unit (GRU) [94] have RMSE values of 0.857 and 0.27, respectively, which are significantly higher than CEFAS.

Several deep learning-based methods, such as the Multimodal Deep AutoEncoder (DAE) [85], Domain-Adversarial Neural Network (DANN) [86], Adversarial Discriminative Domain Adaptation (ADDA) [86], Deep Residual Network (DResNet) [87], and Domain Generalization using Domain Adversarial Neural Networks (DG-DANN) [87], focus solely on predicting mental fatigue. While these models exhibit lower RMSE values compared to the traditional machine learning methods, they still underperform when compared to CEFAS. The Long Short-Term Memory with Capsule Attention (LSTM-CapsAtt) [88], Federated Long Short-Term Memory (FedLSTM) [89], and Federated Gated Recurrent Unit (FedGRU) [89] models incorporate temporal information for fatigue prediction. However, their RMSE values range from 0.1089 to 0.6124, indicating lower accuracy.

Other notable methods include the Double-layered Neural Network with Subnetwork Nodes (DNNSN) [90], regression [91], ANN [91], LSTM [92], Bidirectional LSTM (BiLSTM) [92], GRU [92], Bidirectional GRU (BiGRU) [92], Particle Swarm Optimization-Support Vector Regression (PSO-SVR) [93], and Extreme Learning Machine (ELM) [93]. While these methods show promise in predicting either physical or mental fatigue, their performance metrics are not as competitive as CEFAS.

In conclusion, the proposed CEFAS ensemble learning model demonstrates state-of-the-art performance in predicting both physical and mental fatigue, outperforming a wide range of existing methods across multiple evaluation metrics. This remarkable performance, which signifies a substantial improvement over individual base learners, encourages

the continued exploration and application of ensemble models for time-series forecasting, particularly in the context of fatigue prediction.

Conclusions and future work

In this study, the validity and effectiveness of ensemble learning models in processing IoT sensor data for real-time mental fatigue and physical fatigue monitoring were evaluated. Furthermore, the correlations between the prediction capabilities of different models and the improvement in their prediction accuracy were determined after their integration into an ensemble learning-based ChronoEnsemble Fatigue Analysis System. This study was performed in two stages. First, a study was conducted to verify the reliability, accuracy, and improvement potential of the proposed ensemble learning model in processing and analyzing sensor data. Second, a study was conducted using real sensor data to determine the precision and reliability of the proposed model for predicting fatigue states. During the assessment, the CEFAS algorithm was used for predicting PF and MF. In physical fatigue prediction, the algorithm yielded MSE and RMSE values of 0.0007 and 0.0277, respectively. In mental fatigue prediction, the algorithm yielded MSE and RMSE values of 0.0033 and 0.0577, respectively. These evaluation metrics highlighted the efficacy of the algorithm, indicating excellent prediction accuracy. The results from real-world datasets demonstrated that the ensemble model that uses the DeepAR and TFT models as the base learners and XGBoost as the meta learner yields reliable real-time predictions based on sensor data.

While our findings offer significant insights into the use of IoT and ensemble learning for fatigue monitoring, they are subject to certain limitations that must be considered. Firstly, the dataset used for training and validating CEFAS was limited to 11 participants. Although the ensemble model demonstrated high precision in predicting fatigue, a larger and more diverse dataset would enhance the generalizability of our approach. Secondly, our methodology relies solely on wearable sensor data, which can introduce potential biases related to sensor placement, individual physiological variances, and environmental impacts. Despite implementing a range of preprocessing steps—including baseline removal and various filtering methods—to standardize inputs and mitigate noise from multiple sources, these procedures do not completely eliminate variabilities induced by environmental conditions and individual differences. Additionally, our experimental protocols were conducted under controlled conditions to maximize the reliability and validity of our findings; however, applying these results to the dynamic environments of actual construction sites may necessitate further adaptations through extensive field validations and experiments.

Future research should explore the potential of integrating additional models to enhance forecasting accuracy. Moreover, the model's performance may vary with different sensor configurations due to its reliance on data from a specific set of sensors. Thus, future studies should utilize a diverse array of sensor data to confirm the model's applicability. An additional objective for future research is to expand the dataset to include a larger and more diverse participant pool, along with various construction settings. Integrating these limitations, the study illustrates that while the current ensemble learning model yields reliable real-time predictions for fatigue monitoring, these future

advancements could further support real-time risk assessment, particularly in hazardous work environments.

The present study predominantly focused on the ensemble model and its application in predicting MF and PF; therefore, the future research directions pointed out herein are crucial for improving its applicability. The continued validation of the prediction results in other domains using extensive IoT sensor data, in addition to exploring its wide potential applications, will be essential to exploring the full potential of the ensemble learning model.

Abbreviations

IoT	Internet of things
DeepAR	Deep autoregressive
TFT	Temporal fusion transformer
MSE	Mean square error
RMSE	Root mean square error
MASE	Mean absolute error
MAPE	Mean absolute percentage error
sMAPE	Symmetric mean absolute percentage error
ND	Normalized deviation
MSIS	Mean Scaled Interval Score
PPG	Photoplethysmography
EEG	Electroencephalography

Author contributions

Conceptualization, B. K.; methodology, R. L.; software, R. L. and S. K.; validation, B. K., K. R. S. P., R. L., and S. K.; formal analysis, R. L., and K. R. S. P.; investigation, R. L. and K. R. S. P.; resources, B. K.; data curation, R. L., K. R. S. P., and S. K.; writing—original draft preparation, R. L. and K. R. S. P.; writing—review and editing, B. K., J. A., G. H., K. R. S. P., S. S., Z. C., E. A., and R. L.; visualization, R. L.; supervision B. K.; project administration, R. L., K. R. S. P., E. A., and S. K.; funding acquisition, B. K. All authors have read and agreed to the published version of the manuscript.

Funding

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2023-00251002). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1C1C1005409). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1009781).

Data availability

The data that support the findings of this study are available from the corresponding author, B. K., upon reasonable request.

Declarations

Competing interests

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.

Declaration of generative AI in scientific writing

During the preparation of this work, the author(s) did not use any AI-assisted technologies in the writing process.

Received: 29 November 2023 Accepted: 6 August 2024

Published online: 16 August 2024

References

1. Yu Y, Umer W, Yang X, Antwi-Afari M. Posture-related data collection methods for construction workers: a review. *Autom Constr.* 2021;124: 103538. <https://doi.org/10.1016/j.autcon.2020.103538>.
2. Khallaf R, Khallaf M. Classification and analysis of deep learning applications in construction: a systematic literature review. *Autom Constr.* 2021;129: 103760. <https://doi.org/10.1016/j.autcon.2021.103760>.
3. Ahn C, Lee S, Sun C, Jebelli H, Yang K, Choi B. Wearable sensing technology applications in construction safety and health. *J Constr Eng Manag.* 2019;145:03119007. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001708](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001708).
4. Choi B, Lee S. An empirically based agent-based model of the sociocognitive process of construction workers' safety behavior. *J Constr Eng Manag.* 2018;144:04017102. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001421](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001421).

5. Zhang P, Li N, Jiang Z, Fang D, Anumba C. An agent-based modeling approach for understanding the effect of worker-management interactions on construction workers' safety-related behaviors. *Autom Constr.* 2019;97:29–43. <https://doi.org/10.1016/j.autcon.2018.10.015>.
6. Liang Q, Zhou Z, Ye G, Shen L. Unveiling the mechanism of construction workers' unsafe behaviors from an occupational stress perspective: a qualitative and quantitative examination of a stress–cognition–safety model. *Saf Sci.* 2022;145: 105486. <https://doi.org/10.1016/j.ssci.2021.105486>.
7. Sadat-Mohammadi M, Shakerian S, Liu Y, Asadi S, Jebelli H. Non-invasive physical demand assessment using wearable respiration sensor and random forest classifier. *J Build Eng.* 2021;44: 103279. <https://doi.org/10.1016/j.jobe.2021.103279>.
8. Lee W, Lin K-Y, Seto E, Migliaccio G. Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction. *Autom Constr.* 2017. <https://doi.org/10.1016/j.autcon.2017.06.012>.
9. Awolusi I, Marks E, Hollowell M. Wearable technology for personalized construction safety monitoring and trending: review of applicable devices. *Autom Constr.* 2018;85:96–106. <https://doi.org/10.1016/j.autcon.2017.10.010>.
10. Jebelli H, Choi B, Lee S. Application of wearable biosensors to construction sites. II: Assessing workers' physical demand. *J Constr Eng Manag.* 2019;145:04019080. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001710](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001710).
11. CPWR. The Construction Chart Book-Sixth Edition. Silver Spring, MD: The Center for Construction Research and Training (CPWR); 2018. Available from: <http://www.cpwrc.com/publications/construction-chart-book>.
12. Mohamed Z, Mostafa K, Oetomo A, Hegazy T, Morita P. Wearable devices: cross benefits from healthcare to construction. *Autom Constr.* 2022;142: 104501. <https://doi.org/10.1016/j.autcon.2022.104501>.
13. Powell R, Copping A. Sleep deprivation and its consequences in construction workers. *J Constr Eng Manag.* 2010;136(10):1086–92. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000211](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000211).
14. BLS. 2021 Bureau of Labor Statistics, Census of Occupational Fatal Injuries Summary. Available from: <https://www.bls.gov/news.release/cfoi.nr0.htm>. 2021.
15. Eteifa S, El-adaway I. Using social network analysis to model the interaction between root causes of fatalities in the construction industry. *J Manage Eng.* 2018;34:04017045. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000567](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000567).
16. Golizadeh H, Hon C, Drogemuller R, Hosseini MR, Al-Hussein M, Wong A. Digital engineering potential in addressing causes of construction accidents. *Autom Constr.* 2018;95:284–95. <https://doi.org/10.1016/j.autcon.2018.08.013>.
17. Asilian H, Khosravi Y, Hassanzadeh Rangji N, Hajizadeh E, Behzadan A. A qualitative investigation of factors influencing unsafe work behaviors on construction projects. *Work.* 2018;61:1–13. <https://doi.org/10.3233/WOR-182799>.
18. Yu Y, Li H, Yang X, Kong L, Luo X, Wong A. An automatic and non-invasive physical fatigue assessment method for construction workers. *Autom Constr.* 2019;103:1–12. <https://doi.org/10.1016/j.autcon.2019.02.020>.
19. Umer W, Li H, Yu Y, Antwi-Afari M, Anwer S, Luo X. Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures. *Autom Constr.* 2020. <https://doi.org/10.1016/j.autcon.2020.103079>.
20. Anwer S, Li H, Antwi-Afari M, Umer W, Mehmood I, Al-Hussein M, Wong A. Test-retest reliability, validity, and responsiveness of a textile-based wearable sensor for real-time assessment of physical fatigue in construction bar-benders. *J Build Eng.* 2021;44: 103348. <https://doi.org/10.1016/j.jobe.2021.103348>.
21. Aryal A, Ghahramani A, Becerik-Gerber B. Monitoring fatigue in construction workers using physiological measurements. *Autom Constr.* 2017;82:154. <https://doi.org/10.1016/j.autcon.2017.03.003>.
22. Cheng WC, Tserng H. Real-time individual workload management at tunnel worksite using wearable heart rate measurement devices. *Autom Constr.* 2021;134: 104051. <https://doi.org/10.1016/j.autcon.2021.104051>.
23. Hwang S, Seo J, Jebelli H, Lee S. Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (PPG) sensor embedded in a wristband-type activity tracker. *Autom Constr.* 2016;71:372. <https://doi.org/10.1016/j.autcon.2016.08.029>.
24. Abdelhamid TS, Everett JG. Physiological demands during construction work. *J Constr Eng Manage.* 2002;128(5):427–37. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2002\)128:5\(427\)](https://doi.org/10.1061/(ASCE)0733-9364(2002)128:5(427)).
25. Techera U, Hollowell M, Stambaugh N, Littlejohn R. Causes and consequences of occupational fatigue: meta-analysis and systems model. *J Occup Environ Med.* 2019;61(6):461–73. <https://doi.org/10.1097/JOM.0000000000001564>.
26. Xing X, Zhong B, Luo H, Rose T, Li J, Antwi-Afari MF. Effects of physical fatigue on the induction of mental fatigue of construction workers: a pilot study based on wearable electroencephalography. *Autom Constr.* 2020;120: 103381. <https://doi.org/10.1016/j.autcon.2020.103381>.
27. Li J, Li H, Umer W, Fu H, Xing X. Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology. *Autom Constr.* 2019;105: 102835. <https://doi.org/10.1016/j.autcon.2019.102835>.
28. Hawley DJ, Wolfe F. Fatigue and musculoskeletal pain. *Phys Med Rehabil Clin N Am.* 1997;8:101–11.
29. Cunningham T, Guerin R, Ferguson J, Cavalleri J. Work-related fatigue: a hazard for workers experiencing disproportionate occupational risks. *Am J Ind Med.* 2022;65:913. <https://doi.org/10.1002/ajim.23325>.
30. Dembe A, Erickson JB, Delbos RG, Banks SM. The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. *Occup Environ Med.* 2005;62(9):588–97. <https://doi.org/10.1136/oem.2004.016667>.
31. Vegso S, Cantley L, Slade M, Taiwo O, Sircar K, Rabinowitz P, et al. Extended work hours and risk of acute occupational injury: a case-crossover study of workers in manufacturing. *Am J Ind Med.* 2007;50(8):597–603. <https://doi.org/10.1002/ajim.20486>.
32. Smith CS, Folkard D, Tucker P, Evans MS. Work schedules, health, and safety. In: Quick JC, Tetrick LE, editors. *Handbook of occupational health psychology*. 2nd edn. American Psychological Association; 2011. p. 185–204. <https://doi.org/10.1037/12498-011>.
33. Zhang M, Murphy LA, Fang D, Caban-Martinez AJ. Influence of fatigue on construction workers' physical and cognitive function. *Occup Med (Lond).* 2015;65(3):245–50. <https://doi.org/10.1093/occmed/kqu215>.
34. Glazner JE, Bondy J, Lezotte DC, Lipscomb H, Guarini K. Factors contributing to construction injury at Denver International Airport. *Am J Ind Med.* 2005;47(1):27–36. <https://doi.org/10.1002/ajim.20108>.

35. Barger LK, Cade BE, Ayas NT, Cronin JW, Rosner B, Speizer FE, et al. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med*. 2005;352(2):125–34. <https://doi.org/10.1056/NEJMoa041401>.
36. Ricci JA, Chee E, Lorandeanu AL, Berger J. Fatigue in the U.S. workforce: prevalence and implications for lost productive work time. *J Occup Environ Med*. 2007;49(1):1–10. <https://doi.org/10.1097/01.jom.0000249782.60321.2a>.
37. Hallowell M. Safety risk perception in construction companies in the Pacific Northwest of the USA. *Constr Manag Econ*. 2010;28(4):403–13. <https://doi.org/10.1080/01446191003587752>.
38. Xu M, Nie X, Li H, Cheng J, Mei Z. Smart construction sites: a promising approach to improving on-site HSE management performance. *J Build Eng*. 2022;49: 104007. <https://doi.org/10.1016/j.jobe.2022.104007>.
39. Lou Z, Wang L, Jiang K, Wei Z, Shen G. Reviews of wearable healthcare systems: Materials, devices and system integration. *Mater Sci Eng R Rep*. 2020;140: 100523. <https://doi.org/10.1016/j.mser.2019.100523>.
40. Kubiček J, Fiedorova K, Vilimek D, Cerny M, Penhaker M, Janura M, Rosický J. Recent trends, construction and applications of smart textiles and clothing for monitoring of health activity: a comprehensive multidisciplinary review. *IEEE Rev Biomed Eng*. 2020;15:36. <https://doi.org/10.1109/RBME.2020.3043623>.
41. Ahn J, Ku Y, Kim HC. A novel wearable EEG and ECG recording system for stress assessment. *Sensors*. 2019;19:1991. <https://doi.org/10.3390/s19091991>.
42. Dobkin B. Wearable motion sensors to continuously measure real-world physical activities. *Curr Opin Neurol*. 2013;26:602. <https://doi.org/10.1097/WCO.0000000000000026>.
43. Patel V, Chesmore A, Legner C, Pandey S. Trends in workplace wearable technologies and connected-worker solutions for next-generation occupational safety, health, and productivity. 2022.
44. Abioye S, Oyedele L, Akanbi L, Ajayi A, Davila Delgado M, Bilal M, Akinadé O, Ahmed A. Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges. *J Build Eng*. 2021;44: 103299. <https://doi.org/10.1016/j.jobe.2021.103299>.
45. Martins N, Annaheim S, Spengler C, Rossi R. Fatigue monitoring through wearables: a state-of-the-art review. *Front Physiol*. 2021;12: 790292. <https://doi.org/10.3389/fphys.2021.790292>.
46. Beéck T, Meert W, Schutte K, Vanwaseele B, Davis J. Fatigue prediction in outdoor runners via machine learning and sensor fusion. 2018:606–615. <https://doi.org/10.1145/3219819.3219864>.
47. Wong T, Man SS, Chan A. Critical factors for the use or non-use of personal protective equipment amongst construction workers. *Saf Sci*. 2020;126: 104663. <https://doi.org/10.1016/j.ssci.2020.104663>.
48. Choi B, Hwang S, Lee S. What drives construction workers' acceptance of wearable technologies in the workplace?: Indoor localization and wearable health devices for occupational safety and health. *Autom Constr*. 2017;84:31–41. <https://doi.org/10.1016/j.autcon.2017.08.005>.
49. Cho S, Ensari I, Weng C, Kahn M, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid review (Preprint). *JMIR Mhealth Uhealth*. 2020;9: e20738. <https://doi.org/10.2196/20738>.
50. Chen Z, Jiang C, Xie L. A novel ensemble ELM for human activity recognition using smartphone sensors. *IEEE Trans Ind Inform*. 2019;15(5):2691–9. <https://doi.org/10.1109/TII.2018.2869843>.
51. Rao A, Radanovic M, Liu Y, et al. Real-time monitoring of construction sites: sensors, methods, and applications. *Autom Constr*. 2022;136: 104099. <https://doi.org/10.1016/j.autcon.2021.104099>.
52. Mehmood I, Li H, Qarout Y, et al. Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data. *Adv Eng Inf*. 2023;56: 107978. <https://doi.org/10.1016/j.aei.2023.101978>.
53. Zhang J, Chen M, Peng Y, et al. Wearable biosensors for human fatigue diagnosis: a review. *Bioeng Transl Med*. 2022;8. <https://doi.org/10.1002/btm2.10318>.
54. Karvekar S, Abdollahi M, Rashedi E. Smartphone-based human fatigue level detection using machine learning approaches. *Ergonomics*. 2021;64:1–28. <https://doi.org/10.1080/00140139.2020.1858185>.
55. Saedi S, Fini A, Khanzadi M, Wong J, Sheikhhoshkar M, Banaei M. Applications of electroencephalography in construction. *Autom Constr*. 2022;133:103985. <https://doi.org/10.1016/j.autcon.2021.103985>.
56. Guevara M, Lorenzo I, Arce C, Ramos J, Corsi-Cabrera M. Inter- and intrahemispheric EEG correlation during sleep and wakefulness. *Sleep*. 1995;18:257–65. <https://doi.org/10.1093/sleep/18.4.257>.
57. Corsi-Cabrera M, Guevara M, Arce C, Ramos-Loyo J. Inter and intrahemispheric EEG correlation as a function of sleep cycles. *Prog Neuro-Psychopharmacol Biol Psychiatry*. 1996;20:387–405. [https://doi.org/10.1016/0278-5846\(96\)00004-8](https://doi.org/10.1016/0278-5846(96)00004-8).
58. Kobayashi T, Misaki K, Nakagawa H, Madokoro S, Ota T, Ihara H, Tsuda K, Umezawa Y, Murayama J, Isaki K. Correlation dimension of the human sleep electroencephalogram. *Psychiatry Clin Neurosci*. 2000;54:11–6. <https://doi.org/10.1046/j.1440-1819.2000.00629.x>.
59. Sun C, Longrois D, Holcman D. Spectral EEG correlations from the different phases of general anesthesia. *Front Med*. 2023;10:1009434. <https://doi.org/10.3389/fmed.2023.1009434>.
60. Kar S, Routray A. Effect of sleep deprivation on functional connectivity of EEG channels. *IEEE Trans Syst Man Cybern Syst*. 2013;43:666–72. <https://doi.org/10.1109/TSMCA.2012.2207103>.
61. Li H, Wang D, Chen J, et al. Pre-service fatigue screening for construction workers through wearable EEG-based signal spectral analysis. *Autom Constr*. 2019;106: 102851. <https://doi.org/10.1016/j.autcon.2019.102851>.
62. Craig A, Tran Y, Wijesuriya N, Nguyen H. Regional brain wave activity changes associated with fatigue. *Psychophysiology*. 2012;49:574–82. <https://doi.org/10.1111/j.1469-8986.2011.01329.x>.
63. Nweke H, Wah T, Mujtaba G, Al-Garadi M. Data fusion and multiple classifier systems for human activity detection and health monitoring: review and open research directions. *Inf Fusion*. 2018;46:147–70. <https://doi.org/10.1016/j.inffus.2018.06.002>.
64. Jebelli H, Choi B, Lee S. Application of wearable biosensors to construction sites. I: Assessing Workers' stress. *J Constr Eng Manag*. 2019;145:04019079. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001729](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001729).
65. Sikander G, Anwar S. Driver fatigue detection systems: a review. *IEEE Trans Intell Transp Syst*. 2018. <https://doi.org/10.1109/TITS.2018.2868499>.
66. Mukhopadhyay SC. Wearable sensors for human activity monitoring: a review. *IEEE Sens J*. 2015;15:1321–30. <https://doi.org/10.1109/JSEN.2014.2370945>.

67. Yu Y, Li H, Umer W, Dong C, Yang X, Skitmore M, Wong A. Automatic biomechanical workload estimation for construction workers by computer vision and smart insoles. *J Comput Civ Eng*. 2019;33. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000827](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000827).
68. Bangaru SS, Wang C, Aghazadeh F. Automated and continuous fatigue monitoring in construction workers using forearm EMG and IMU wearable sensors and recurrent neural network. *Sensors*. 2022;22(24):9729. <https://doi.org/10.3390/s22249729>.
69. Wang D, Chen J, Zhao D, et al. Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. *Autom Constr*. 2017;82:122–37. <https://doi.org/10.1016/j.autcon.2017.02.001>.
70. Gatti U, Schneider S, Migliaccio G. Physiological condition monitoring of construction workers. *Autom Constr*. 2014;44:227–33. <https://doi.org/10.1016/j.autcon.2014.04.013>.
71. Guo H, Yu Y, Xiang T, Li H, Zhang D. The availability of wearable-device-based physical data for the measurement of construction workers' psychological status on site: from the perspective of safety management. *Autom Constr*. 2017;82:207–17. <https://doi.org/10.1016/j.autcon.2017.06.001>.
72. Hwang S, Lee S. Wristband-type wearable health devices to measure construction workers' physical demands. *Autom Constr*. 2017;83:330–40. <https://doi.org/10.1016/j.autcon.2017.06.003>.
73. Ganaie M, Hu M, Malik AK, Tanveer M, Suganthan P. Ensemble deep learning: a review. *Eng Appl Artif Intell*. 2022;115: 105151. <https://doi.org/10.1016/j.engappai.2022.105151>.
74. Sagi O, Rokach L. Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2018;8: e1249. <https://doi.org/10.1002/widm.1249>.
75. Anwer S, Li H, Antwi-Afari M, Umer W, Wong A. Evaluation of physiological metrics as a real-time measurement of physical fatigue in construction workers: state-of-the-art reviews. *J Constr Eng Manage*. 2021;147. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002038](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002038).
76. Fang W, Wu D, Ding L, Luo H. Physiological computing for occupational health and safety in construction: review, challenges and implications for future research. *Adv Eng Inform*. 2022;54: 101729. <https://doi.org/10.1016/j.aei.2022.101729>.
77. Åhsberg E, Garnberale F, Kjellberg A. Perceived quality of fatigue during different occupational tasks development of a questionnaire. *Int J Ind Ergon*. 1997;20:121–35.
78. Smets EM, Garssen B, Bonke B, De Haes JC. The multidimensional fatigue inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res*. 1995;39(3):315–25. [https://doi.org/10.1016/0022-3999\(94\)00125-o](https://doi.org/10.1016/0022-3999(94)00125-o).
79. Flunkert V, Salinas D, Gasthaus J. DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int J Forecast*. 2017;36:1181. <https://doi.org/10.1016/j.ijforecast.2019.07.001>.
80. Lim B, Anik S, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast*. 2021;37:1748–64. <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
81. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–794. <https://doi.org/10.1145/2939672.2939785>.
82. Lin FC, Ko LW, Chuang CH, Su TP, Lin CT. Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system. *IEEE Trans Circuits Syst Regul Pap*. 2012;59(9):2044–55. <https://doi.org/10.1109/TCSI.2012.2185290>.
83. Huo XQ, Zheng WL, Lu BL. Driving fatigue detection with fusion of EEG and forehead EOG. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. 2016. p. 897–904. <https://doi.org/10.1109/IJCNN.2016.7727294>.
84. Yi W, Chan AP, Wang X, Wang J. Development of an early warning system for site work in hot and humid environments: a case study. *Autom Constr*. 2016;62:101–13. <https://doi.org/10.1016/j.autcon.2015.11.003>.
85. Du LH, Liu W, Zheng WL, Lu BL. Detecting driving fatigue with multimodal deep learning. In: *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*. 2017. p. 74–77. <https://doi.org/10.1109/NER.2017.8008295>.
86. Li H, Zheng WL, Lu BL. Multimodal vigilance estimation with adversarial domain adaptation networks. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. 2018. p. 1–6. <https://doi.org/10.1109/IJCNN.2018.8489212>.
87. Ma BQ, Li H, Zheng WL, Lu BL. Reducing the subject variability of EEG signals with adversarial domain generalization. In: *Intelligence Science and Big Data Engineering. Image and Video Data Engineering*. Springer; 2019. p. 16–25. https://doi.org/10.1007/978-3-030-36708-4_3.
88. Zhang G, Etemad A. Capsule attention for multimodal EEG-EOG representation learning with application to driver vigilance estimation. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:1686–96. <https://doi.org/10.1109/TNSRE.2021.3089594>.
89. Li X, Chi HL, Lu W, Xue F, Zeng J, Li C. Federated transfer learning enabled smart work packaging for preserving personal image information of construction worker. *Autom Constr*. 2021;128: 103738. <https://doi.org/10.1016/j.autcon.2021.103738>.
90. Wu W, Wu QMJ, Sun W, Yang Y, Yuan X, Zheng WL, Lu BL. A regression method with subnetwork neurons for vigilance estimation using EOG and EEG. *IEEE Trans Cogn Dev Syst*. 2019;11(4):583–93. <https://doi.org/10.1109/TCDS.2018.2889223>.
91. Senouci A. A model for predicting construction worker fatigue. *J Sustain Built Environ City Transit*. 2022;4(2):27–32. <https://doi.org/10.30564/jsbct.v4i2.35>.
92. Ghafouri M, Clevenger C, Abdallah M, Rens K. Heart rate modeling and prediction of construction workers based on physical activity using deep learning. *Autom Constr*. 2023;155: 105077. <https://doi.org/10.1016/j.autcon.2023.105077>.
93. Feng W, Zeng K, Zeng X, Chen J, Peng H, Hu B, Liu G. Predicting physical fatigue in athletes in rope skipping training using ECG signals. *Biomed Signal Process Control*. 2023;83: 104663. <https://doi.org/10.1016/j.bspc.2023.104663>.
94. Yuan D, Yue J, Xiong X, Jiang Y, Zan P, Li C. A regression method for EEG-based cross-dataset fatigue detection. *Front Physiol*. 2023;14:1196919. <https://doi.org/10.3389/fphys.2023.1196919>.

95. Shi J, Wang K. Fatigue driving detection method based on time-space-frequency features of multimodal signals. *Biomed Signal Process Control*. 2023;84: 104744. <https://doi.org/10.1016/j.bspc.2023.104744>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.