The following publication Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. Automation in Construction, 109, 103000 is available at https://dx.doi.org/10.1016/j.autcon.2019.103000.

Identification and classification of construction equipment operators' multi-level mental fatigue using wearable eye-tracking technology

Abstract

In the construction industry, the operator's mental fatigue is one of the most important causes of construction equipment-related accidents. Mental fatigue can easily lead to poor performance of construction equipment operations and accidents in the worst case scenario. Hence, it is necessary to propose an objective method that can accurately detect multiple levels of mental fatigue of construction equipment operators. To address such issue, this paper develops a novel method to identify and classify operator's multi-level mental fatigue using wearable eye-tracking technology. For the purpose, six participants were recruited to perform a simulated excavator operation experiment to obtain relevant data. First, a Toeplitz Inverse Covariance-Based Clustering (TICC) method was used to determine the number of levels of mental fatigue using relevant subjective and objective data collected during the experiments. The results revealed the number of mental fatigue levels to be 3 using TICC-based method. Second, four eye movement feature-sets suitable for different construction scenarios were extracted and supervised learning algorithms were used to classify multi-level mental fatigue of the operator. The classification performance analysis of the supervised learning algorithms showed Support Vector Machine (SVM) was the most suitable algorithm to classify mental fatigue in the face of various construction scenarios and subject bias (accuracy between 79.5% and 85.0%). Overall, this study demonstrates the feasibility of applying wearable eye-tracking technology to identify and classify the mental fatigue of construction equipment operators.

Keywords: Mental fatigue identification and classification; Construction equipment operator; Eye-tracking; Machine learning; Toeplitz Inverse Covariance-Based Clustering

1. Introduction

Construction equipment operation often involves mentally demanding tasks during which operators are required to maintain sustained attention and keep vigilance on surrounding hazards simultaneously. Because of prolonged operating and vigilance activities, construction equipment operators are prone to mental fatigue [1,2]. Mental fatigue has been associated with serious hazard detection failures (i.e., change blindness, inattention, vigilance lapse, etc.) [1,3,4], which has been identified as one of the leading causes of construction fatal accidents [5-7]. For example, struck-by interactions between pedestrian workers and construction equipment accounted for almost one-fifth of fatalities in 2016 in the United States construction industry [8]. Similarly, Occupational Safety and Health Association (OSHA) found such struck-by accidents as one of the four major causes of fatalities in the construction industry. In addition to fatal accidents, short-term mental fatigue can easily lead to low productivity in the workplace [9], and long-term mental fatigue can result in serious physical and mental health problems [10]. Taking the above into account, detecting and managing mental fatigue are necessary which could help to reduce the risk of struck-by accidents that has become an important priority for researchers and practitioners in the construction industry.

Previous studies have endeavored to measure fatigue for construction safety [3,11-14]. However, many of these studies have mainly focused on physical fatigue which could be detected relatively easily using multiple physiological measures (i.e., heart rate, electromyography, skin temperature) [11]. Unlike physical fatigue, there are a few effective means of measuring mental fatigue on construction sites. Several studies have measured and assessed construction workers' mental fatigue by relying on workers' subjective assessment [15] or by asking the workers to be involved in additional task performance assessment [14]. As such, these methods interfere with regular work-tasks and are impractical for continuous real-time mental fatigue monitoring. Although a series of neurophysiological measurement techniques (e.g., electroencephalogram and electrocardiograph) have been developed which can provide objective, real-time mental fatigue detection, these techniques are based on the recording of the electrical activity of the human body, which is highly invasive to workers and the electrical signals are susceptible to harsh environmental conditions at the construction site, limiting their reliability and availability of application on construction sites. To address such issues, this study proposes a non-invasive, easy-to-install wearable eye-tracking technology to measure construction equipment operators' mental fatigue.

In order to enable automatic construction equipment operators' mental fatigue detection, this study applied supervised learning to extract eye-movement-related features and classify corresponding mental fatigue levels based on eye movement data. Moreover, interestingly, the current studies on supervised learning-based mental fatigue classification usually use individual subjective feelings or researchers' experience as a ground truth for determining mental fatigue levels. Most of these studies classify mental fatigue into a binary classification: fatigue and no fatigue, which is not sufficient to provide an adequate understanding of mental fatigue for researchers and practitioners to make safety management decisions [16,17]. The development of fatigue is a gradual process of accumulation in which multiple intermediate states of mental fatigue have different effects on operators' operating tasks. Therefore, it is necessary to find an objective and reliable method to classify mental fatigue at multiple levels to help construction practitioners identify the safety risks caused by each mental fatigue level and propose more effective safety interventions respectively.

This paper proposes a machine-learning-based automated mental fatigue identification and classification method for construction equipment operators using wearable eye-tracking technology and a Toeplitz Inverse Covariance-Based Clustering (TICC)-based multi-level mental fatigue identification and data labeling method. A laboratory experiment was conducted to collect relevant data from simulated equipment operating task. We collected mental fatigue, productivity, and safety related time series data from the experiment, used TICC method to identify mental fatigue levels and labels for the data. Different levels of mental fatigue are associated with different subjective feelings, task performance, and eye movement behaviors that are reflected in specific combination pattern of various eye movement features measured by wearable eye tracking technology. On this basis, classification and detection of mental fatigue were performed. Supervised learning algorithms were developed to classify levels of mental fatigue by using spatial and temporal features that reflect the specific eye movement data patterns. The detection performance evaluation was conducted based on the experiment data. According to the results, the feasibility of the proposed research approach, its future application prospect and corresponding improvement direction were discussed.

2. Related work

2.1. Mental fatigue measurement methods

Mental fatigue measurement methods can be broadly classified into two categories: subjective assessment and instrument-based measurement (Fig. 1). Early attempts at measuring mental fatigue involved various assessment scales that relied on subjective responses to a set of questions relating to physical and mental states [18,19]. Several construction-related studies have also utilized different self-assessment scales for assessing worker's mental fatigue [13-15]. However, an inherent problem of subjective assessment is that a person's true objective state (e.g. mental fatigue level) is different from his subjective feeling of his own state [20]. Furthermore, collecting worker's self-assessment is burdensome and not practical on construction sites, highlighting a need for methods that can continuously monitor mental fatigue with minimal interference to construction activities.

In order to address the problems of the subjective assessment methods, various instrument-based measurements have been proposed. As shown in Fig.1, a series of individual behavioral and physiological features that reflect the level of mental fatigue can be monitored through task performance testing, electrophysiological state monitoring, and physical activity monitoring. Task performance testing method assesses individual mental fatigue levels by recording task performance on standardized tasks reflecting mental fatigue, such as Psychomotor Vigilance Test (PVT) [11,21]. Electrophysiological state monitoring is a technique for measuring mental fatigue by monitoring the electrophysiological signals of human bodies. They are widely used and offer different electrophysiological signals for mental fatigue measurement, such as electroencephalogram (EEG) [22,23], electrooculogram (EOG) [24], and electrocardiograph (ECG) [25]. In recent years, physical activity monitoring method are applied to measure and evaluate car drivers' mental fatigue, which mainly involves facial expression recognition and eye-tracking [26]. Generally, these behavioral and physiological features contribute significantly to mental fatigue measuring because a person usually has little control over them unlike subjective assessment, which makes them reliable and objective source of information to determine person's mental fatigue [27].

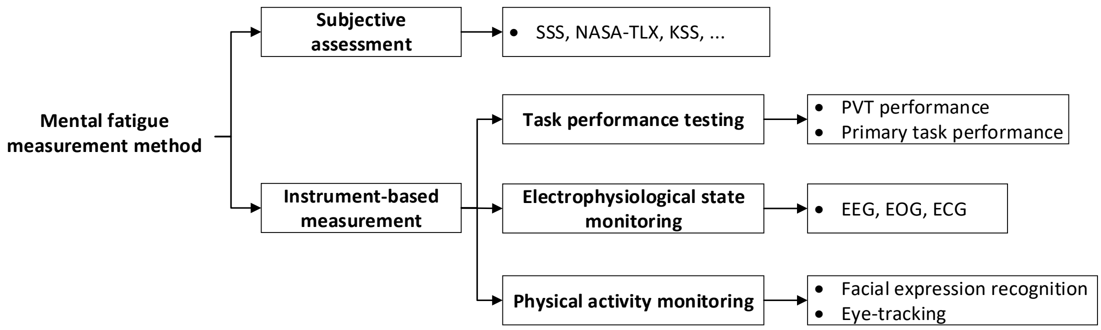

Fig. 1. Summary of mental fatigue measurement methods.

Note: SSS= Stanford Sleepiness Scale; NASA-TLX= NASA Task Load Index; KSS= Karolinska Sleepiness Scale.

In recent studies, EEG, ECG, and EOG have become common and effective methods that measure mental fatigue. However, these technologies are invasive in nature and require skin preparation for sensors adherence which might instigate irritation. Moreover, on-site equipment operation often requires operators to do a lot of physical and mental activities in a complex and harsh operating environment, leading to circumstances, such as equipment electromagnetic interference, high temperature, equipment vibration, body movement and sweat [28], which could hinder the application of these technologies on construction sites.

Compared to the mental fatigue measurement methods based on electrophysiological signal monitoring, eye-tracking technology is not susceptible to interference from those factors on the construction site. There have been few studies focusing on measuring mental fatigue or some cognitive activities related to mental fatigue using eye-tracking technology in the construction industry. Li, et al. [4] evaluated the impairing effect of mental fatigue on excavator operators' hazard detection performance through eye-tracking technology. From the statistical point of view, they demonstrated that the eye movement metrics measured by a wearable eye tracker can reflect operators' mental fatigue state. Other studies have applied eye-tracking technology to measure cognitive functions/activities that might be related to mental fatigue, such as situation awareness [29] and visual search patterns[30]. However, these studies did not directly investigate workers' mental fatigue, nor did they verify the feasibility of eye-tracking technology for mental fatigue measurement. In general, previous studies have not been able to directly measure and detect mental fatigue of workers or operators in the construction industry, which could greatly burden the management and intervention of mental fatigue on construction sites. In this study, a low-invasive, easy-to-install wearable eye tracker was used to monitor construction equipment operators' eye movement for measuring and detecting their mental fatigue.

2.2. Machine-learning-based mental fatigue detection

Previous studies have attempted to develop various methods for the detection of mental fatigue based on physiological signals and those studies mainly used statistical analysis [25,31], single-index-based assessment [32], and machine learning methods. Compared with the first two methods, machine-learning-based methods have been demonstrated in many studies that the classification of different individual physiological and psychological features through various supervised learning algorithms can support a quick, accurate and robust mental fatigue detection [26,33,34]. In the related research, the mental fatigue detection method with EEG signal as the main feature has been found the most widely used [21,35,36]. However, as mentioned earlier, it is difficult to apply EEG and other electrophysiological signal measurement methods at the construction site for a series of reasons aforementioned. Consequently, researchers have begun to use supervised machine learning to process individual eye movement data to detect mental fatigue [17,37]. This provides a potential solution for construction equipment operators' mental fatigue detection. Noteworthy, the eye movement features reflecting mental fatigue in a video watching situation [17,37] may not be applicable to the operator's mental fatigue test because of specific environment of the construction industry which needs to be evaluated. Accordingly, in this study, we explored suitable eye movement features that can effectively reflect the mental fatigue of the operators in construction sites.

A key process before mental fatigue detection based on machine learning is the identification of mental fatigue levels and the data labeling [16,17]. The labels are used to classify mental fatigue in training and evaluating the algorithms, which represent the ground truth of the mental fatigue levels being predicted [38]. Whether it is based on EEG, eye-tracking or other measurement methods, the identification and data labeling methods can be summarized as two main methods. In the first method, mental fatigue levels are often identified and manually labeled using the subjective assessment scale [39] or performance test [21]. The second method usually identifies mental fatigue levels and labels corresponding data according to time-on-task phases [17,34-36,40] or task types (e.g., working at heights *vs*. working on the ground) [41]. The above studies usually divide mental fatigue into binary states: fatigue and non-fatigue. However, identifying mental fatigue on a scalar or ordinal scale will be more useful for monitoring and managing operators' safety. Moreover, classifying mental fatigue from two levels through subjective evaluation or task time division is not accurate and objective enough [37,42], and most importantly, it is not enough to provide sufficient and explicable information and basis for safety management decision-making [16,17,34]. The development of mental fatigue is a process of gradual change and accumulation [22,43] in which there are multiple intermediate states, leading to different behaviors [44], task performance [45], and different levels of safety risks [11]. Determining these intermediate states/levels can help to deepen the understanding of construction equipment operators' mental fatigue and provide construction practitioners with a basis for developing mental fatigue intervention measures.

3. Methods

The research framework consists of four main steps, as shown in Fig. 2. In Step 1, an experiment was conducted to help acquire relevant data. An eye tracker was used to obtain the eye movement data of the operators, and data related to mental fatigue such as subjective feelings of mental fatigue (task load and sleepiness) and operation task performance (production performance and safety performance of the operating task) was collected. In Step 2, based on the TICC method, we analyzed the time series data that can intuitively reflect the operator's mental fatigue, identified multiple levels of operator's mental fatigue, and automatically labeled the eye movement data. In step 3, combining the labeled eye movement data, the automatic classification and detection of the multi-level mental fatigue of the operator were performed based on supervised learning, and the proposed classification algorithms were validated and evaluated. The details of the research method can be seen below.

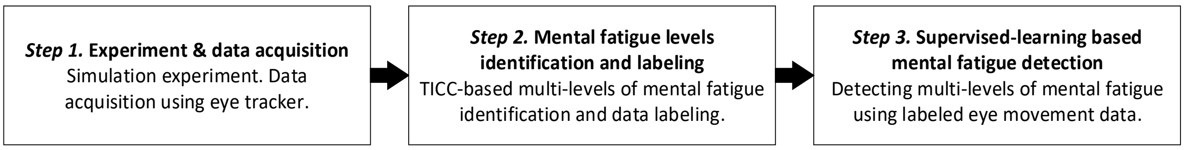

Fig. 2. Research framework.

3.1. Experiment

A laboratory experiment was conducted to help to collect relevant data, considering that it is difficult to control the uncertainties of the experimental tasks and ensure the experiment safety on the actual construction sites. The overview of the experiment is shown in Fig. 3.

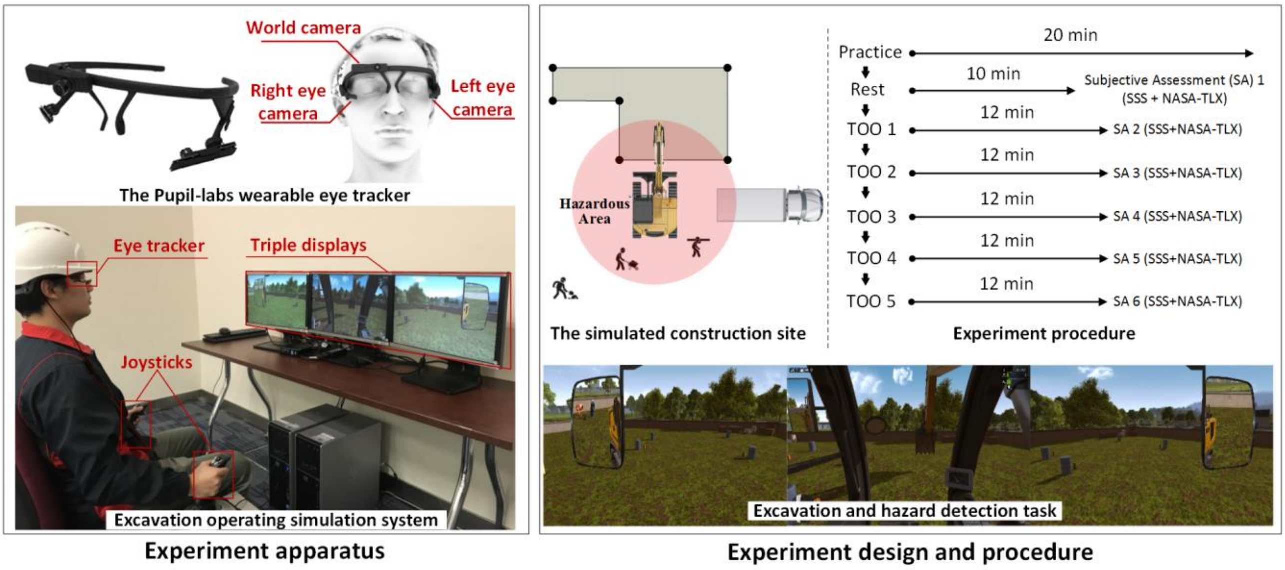

Fig. 3. Experiment overview.

3.1.1. Participants

Six participants (6 males), between the age of 26 and 31 (*Mean* 28.33 years, *SD* 1.86 years) took part in this study. The participants were all excavator operators from the industry with a mean experience of 3 years (range 1 to 8 years). All participants had normal vision and good health. They slept for at least seven hours and were required to quit alcohol or caffeine beverages 24 hours before the experiment. All participants were unaware of the purpose of the experiment and the expected results. Written informed consent was obtained before the experiment.

3.1.2. Experimental apparatus

The eye movement behaviors of the participants were recorded by the wearable Pupil-labs eye tracker [46], as shown in Fig. 3. The eye tracker consisted of a world camera and two eye cameras. The world camera was a 100 degrees diagonal camera directed toward the scene in front of the user with a 60 Hz sampling frequency and 1280×720 pixels resolution. The eye cameras can record the user's gaze point, blink and pupil behaviors with infrared illumination at a 200 Hz sampling frequency. Pupil-labs eye tracker is very light, and its frame is durable and can be adjusted according to the user's facial features, greatly reducing the interference with the user's task and increasing the reliability of the data acquisition. Moreover, this product has a lower price than other commercial eye trackers and provides open source software and alternative ordinary camera accessories. Its price can be as low as $100 under the premise of ensuring the reliability of eye tracker. Considering the cost control of the construction project and the practicality of the eye-tracking technology on the construction site, this product is very suitable for the construction industry.

Excavation operating simulation system consisted of three monitors with 1680×1050 pixels resolution, two joysticks with force feedback and an adjustable seat, as shown in Fig 3. The monitor in the middle was used to display the scene in front of the excavator cockpit, and the monitors on both sides were used to display key scenes in the two rearview mirrors of the excavator along with surrounding scenes. The distance between the participant's eye point and the front monitor was about 120 cm. This system can record the various task performances (e.g., excavation task and hazard detection task performance) of the participants in real time.

3.1.3. Experimental procedure and task design

All participants needed to complete a Time-On-Operating (TOO) experiment procedure as shown in Fig. 3. TOO is a common experimental design paradigm that helps induce mental fatigue [22,31,43]. The simulated construction site can also be seen in Fig. 3. Participants were instructed to perform the TOO task using the simulation system in a laboratory where temperature, illumination, and noise were controlled and constant. All light condition and noise changes are generated from simulation scenarios in the simulation system. The simulation scenario is a typical open construction site under daylight condition. There are no drastic, frequent lighting changes in this simulation scene. The entire experiment lasted about 100 min, including a practice task, a rest break, and five TOO phases. To avoid an end-spurt effect-reactivation that occurs when participants know they are approaching the end of an experiment session, participants were uninformed about the experiment duration [31].

A 60-min TOO procedure combined with dual experimental tasks setting was adopted to reproduce the characteristics of the real excavator operating situation. The dual experimental tasks included a primary task and a secondary task, these two tasks are carried out simultaneously according to the real construction operational task. This experimental task design is a common experiment paradigm in the field of experimental psychology [22,43], road transportation safety [31,47] and aviation safety [48]. The primary task was an excavation task that is the most common operation in earthworks. The secondary task was a hazard detection task (HDT). HDT required participants to manually respond to visual stimuli, which can be used to measure the hazard detecting performance of the participants [49,50]. This long-term continuous, repetitive dual experimental task setting is widely used in many fields such as psychology [51,52] and driving safety [25,50], enabling participants to feel sufficiently high level of mental fatigue in the experimental task that has a shorter duration than the actual task. It could make the experimental results reflective of the problems caused by mental fatigue in the real construction site.

In this experiment, the participant needed to respond to the situations reflected in the rearview mirrors comprising of occasional appearance of pedestrian worker. The instance of appearance of worker in the mirror and the distance between worker and excavator were random for all participants. Participants needed to determine whether the worker is in a hazardous area (within the radius of rotation of the excavator) by judging the distance between the worker and the excavator. Once the worker appearing in the rearview mirror enters the hazardous area, all participants needed to do their best to respond quickly by pressing a button on the joysticks and slow down the rotating speed of the excavator or temporarily stop the operation to avoid struck-by accident.

3.1.4 Data acquisition

During the experiment, the task performances and eye movements of the participants were recorded by the simulation system and the eye tracker respectively in real time. Although research in psychology [17,44,53], transportation safety [54] and other fields has concluded that there are several eye movement behaviors have significant correlation with the development of mental fatigue, the specific eye movement metrics to be selected in this study still need to comprehensively consider the function of the eye tracker, TICC algorithm, supervised learning algorithms, and construction practice. First, despite the Pupil-labs eye tracker is very suitable for the construction industry, its limited measurement accuracy and function cannot support the identification of some metrics that require higher image recognition capability, such as saccade velocity. Therefore we did not use these metrics. Second, to achieve TICC-based identification and supervised learning-based classification, on the one hand, a higher sampling frequency is required to record enough data samples; on the other hand, the metrics must be able to correctly reflect the dynamic features of operators' mental fatigue accumulation. Therefore, we abandon those metrics that only have sparse data or are prone to irregular changes due to the influences of uncertain external stimuli (e.g., pedestrian workers and surrounding moving equipment), such as fixation duration, fixation count, dwell time, etc.

As a result, this study selected blink rate, blink duration, pupil diameter, and gaze position for the identification and classification of mental fatigue. These metrics have been shown to be sensitive to the change of individual mental fatigue state. Among them, pupil diameter will be significantly reduced as mental fatigue increases [17,44,53]. Blink rate and blink duration increase significantly as mental fatigue increases [17,54-57]. The gaze position can be used to represent the operator's visual attention range, which is significantly related to mental fatigue [4]. These metrics are easily measured at the construction site and have been demonstrated to reflect the mental fatigue of construction equipment operators [4]. From the perspective of construction project cost control and eye tracker usability mentioned above, if we can use the data of these metrics to obtain satisfactory mental fatigue classification performance, this can provide strong evidence for the feasibility of applying eye-tracking technology in the construction industry. Therefore, these metrics were selected as the source of data for mental fatigue identification and classification in the study. All eye movement raw data in this study was recorded and generated by Pupil Capture and Pupil Player [46].

In addition, subjective feeling of mental fatigue was assessed through the Stanford Sleepiness Scale (SSS) and the NASA Task Load Index (NASA-TLX) before the start of the TOO procedure and after each TOO phase (a total of 6 times) in order to measure their temporal development (see Fig. 3).

3.2. Multi-level mental fatigue identification and data labeling based on Toeplitz Inverse Covariance-Based Clustering

The aforementioned measures resulted in a large amount of time series data. Determining multiple levels of mental fatigue from time series data obtained through different sensors/measurement methods is challenging. Moreover, manual labeling of a large amount of mental fatigue time series data is labor-intensive and tedious. To this end, a highly efficient subsequence clustering method, the Toeplitz Inverse Covariance-Based Clustering method (TICC) [58] was used to identify and classify multiple levels of operator mental fatigue and automatically label relevant time series data.

3.2.1 Toeplitz Inverse Covariance-Based Clustering

TICC is a model-based multivariate time series subsequence clustering method for discovering repeated patterns in temporal data [58]. TICC defines each cluster as a dependency network showing the relationships between different sensors in a short subsequence, which is able to find accurate and interpretable structure in the data. Considering that the data related to mental fatigue (e.g., eye movement metrics, task performance metrics and subjective assessment) obtained in the experiment is a multi-dimensional time series and mental fatigue is gradually accumulated over time, TICC is suitable for the problem of multi-level classification of mental fatigue in this study.

As defined by TICC, we first consider a time series of *T* sequential observations,

$$x_{orig} = \begin{bmatrix} | & | & | & & | \\ x_1 & x_2 & x_3 & \cdots & x_T \\ | & | & | & & | \end{bmatrix}, \tag{1}$$

where $x_t \in \boldsymbol{R}^n$ is the *t*-th multivariate observation. In brief, the goal is to cluster these *T* observations into *K* clusters. TICC focus on clustering a short subsequence of size $w \ll T$ that ends at t. The observations $x_{t-w+1}, \ldots, x_t$ is constructed into an nw-dimensional vector X_t. As a result, a new sequence from X_1 to X_t is generated, which is a helpful medium for providing proper context for each of the *T* observations. Therefore, the TICC method does not directly cluster the observations, but clusters these subsequences $X_1, \dots, X_t$.

TICC defines each cluster as a Markov Random Field (MRF) to describe the interrelationships between various observations in a representative subsequence of the cluster [58]. Specifically, a Gaussian inverse covariance $\Theta_i \in \boldsymbol{R}^{nw\times nw}$ is used as the mathematical expression of the cluster i's MRF in the form of block Toeplitz:

$$\Theta_i = \begin{bmatrix} A^{(0)} & (A^{(1)})^T & (A^{(2)})^T & \cdots & \cdots & (A^{(w-1)})^T \\ A^{(1)} & A^{(0)} & (A^{(1)})^T & \ddots & & \vdots \\ A^{(2)} & A^{(1)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & (A^{(1)})^T & (A^{(2)})^T \\ \vdots & & \ddots & A^{(1)} & A^{(0)} & (A^{(1)})^T \\ A^{(w-1)} & \cdots & \cdots & A^{(2)} & A^{(1)} & A^{(0)} \end{bmatrix}, \tag{2}$$

where $A^{(0)}, A^{(1)}, \dots, A^{(w-1)} \in \boldsymbol{R}^{n\times n}$. $A^{(0)}$ describes the intra-time partial interdependencies, so that $A_{ij}^{(0)}$ defines the interrelationship between concurrent values of sensor i and sensor j (e.g., the hazard detection performance recording system and the eye tracker).

The overall goal of TICC is to solve the K inverse covariances $\boldsymbol{\Theta} = \{\Theta_1, \dots, \Theta_K\}$ and get the corresponding point assignment sets $\boldsymbol{P} = \{P_1, \dots, P_K\}$ ($P_i \subset \{1,2,\dots,T\}$), which is an optimization problem and can be expressed as follows:

$$\underset{\boldsymbol{\Theta}\in\boldsymbol{\tau},\boldsymbol{P}}{\operatorname{argmin}} \sum_{i=1}^{K} \left[\overbrace{\|\lambda \circ \Theta_i\|_1}^{sparsity} + \sum_{X_t\in P_i} \left(\overbrace{-\ell\ell(X_t, \Theta_i)}^{log\ likelihod} + \overbrace{\beta 1\{X_{t-1} \notin P_i\}}^{temporal\ consistency} \right) \right], \tag{3}$$

where, $\boldsymbol{\tau}$ is the set of symmetric block Toeplitz $nw \times nw$ matrices. $\|\lambda \circ \Theta_i\|_1$ is an ℓ_1-norm penalty of the Hadamard product to incentivize a sparse inverse covariance, where $\lambda \in \boldsymbol{R}^{nw\times nw}$ is a regularization parameter. $\ell\ell(X_t, \Theta_i)$ is the log likelihood that X_t came from cluster i. The penalty parameter β is used to enforce temporal consistency, and $1\{X_{t-1} \notin P_i\}$ is an indicator function judging whether adjacent points are assigned to the same cluster.

TICC solves the optimization problem expressed as Eq. (3) by two main steps: Assigning Points to Clusters and Toeplitz Graphical Lasso. A variation of the expectation maximization (EM) algorithm is used to alternate between assigning points to clusters and then updating the cluster parameters.

Step 1. Assigning Points to Clusters: All observation points are assigned to clusters by determining $\boldsymbol{\Theta}$ and solving the combinatorial optimization problem for $\boldsymbol{P} = \{P_1, \ldots, P_K\}$, which can be expressed as follows:

$$\min \sum_{i=1}^{K} \sum_{X_t \in P_i} -\ell\ell(X_t, \Theta_i) + \beta 1\{X_{t-1} \notin P_i\}. \tag{4}$$

This step assigns each *T* subsequences to one of the *K* clusters to collectively maximize the log likelihood and the time consistency, and trade off between the two objectives adjusted by the regularization parameter *β*. TICC use a dynamic programming algorithm to assign each X_t into a cluster.

Step 2. Updating Cluster Parameters (Toeplitz Graphical Lasso): Updating the cluster parameters $\Theta_1, \ldots, \Theta_K$ by solving the optimization problem expressed as Eq. (3) while holding the given $\boldsymbol{P}$ constant. Each Θ_i is solved in parallel, which can be expressed as

$$\min -\log\det \Theta_i + tr(S_i \Theta_i) + \frac{1}{|P_i|} \|\lambda \circ \Theta_i\|_1, \text{ subject to } \Theta_i \in \boldsymbol{\tau}, \tag{5}$$

where $|P_i|$ is the number of points assigned to cluster *i*, S_i is the empirical covariance of these points. This is a variation on the graphical lasso problem. Alternating direction method of multipliers (ADMM) is used to solve a separate Toeplitz graphical lasso for each cluster at every iteration of TICC algorithm to solve the overall optimization problem expressed in Eq. (3). A message passing algorithm is used to iteratively converge to the global optimal solution. The detailed TICC solution process can be referred to [58].

3.2.2 TICC-based mental fatigue multi-level identification and data labeling

The framework of TICC-based mental fatigue multi-level identification and data labeling is shown in Fig. 4. We first selected typical variables related to mental fatigue as a time series data source for TICC-based analysis (see Table 1). These multidimensional temporal variables can intuitively describe the characteristics of operators' subjective feelings, task performance and objective eye movement behaviors under the state of mental fatigue, making it easier to explain and understand the effects of different levels of mental fatigue on various facets of construction equipment operation. The subjective feelings include the NASA-TLX score reflecting the task mental workload, and the SSS score reflecting the degree of drowsiness. The task performance includes the performance of the excavation task reflecting the productivity and the HDT performance reflecting the operation safety. The eye movement behaviors are related to the three main aspects of blink, pupil, and gaze.

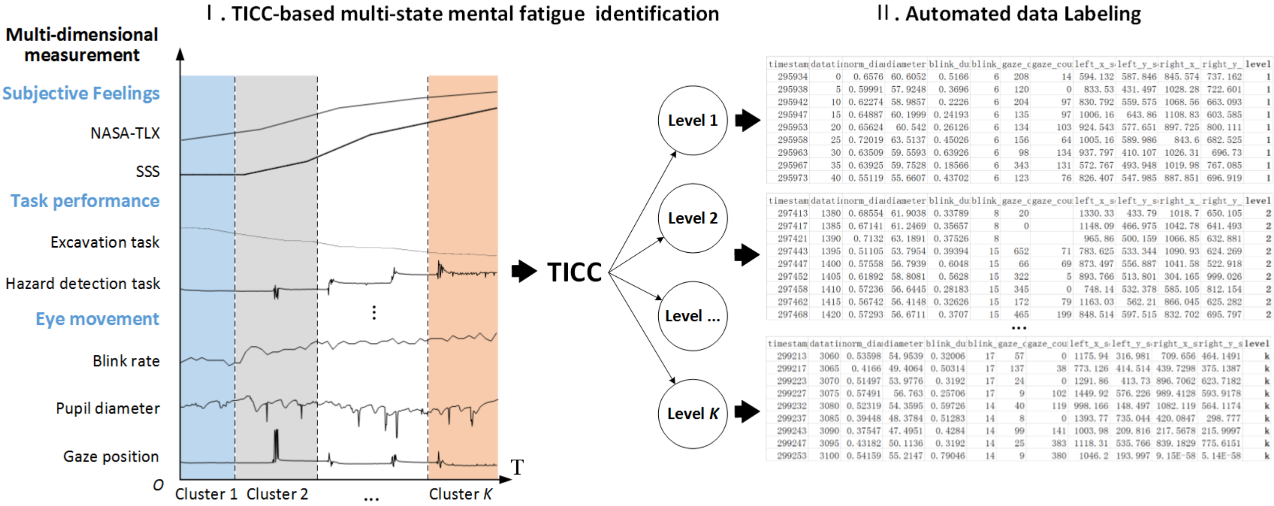

Fig. 4. TICC-based mental fatigue multi-level identification and data labeling.

Specifically, the eye movement behavior-related variables mentioned above include blink rate, pupil diameter, and gaze position. The blink rate and the pupil diameter represent typical blink activity and pupil state that reflect the level of mental fatigue, respectively. The gaze position can visually reflect the operator's visual attention allocation range to the surrounding hazards under mental fatigue. The visual attention allocation range can be calculated from the spatial positions of the gaze points of participants. As shown in Fig. 5, each monitor was assigned a standardized coordinate system, and the position of the gaze point on each monitor can be quantized by the coordinates of the corresponding coordinate system. The visual attention allocation range was measured by the distances from the gaze points on the left/right monitor to the middle monitor, which are $d_{l \to m}$ and $d_{r \to m}$ respectively. In order to make it more intuitive to quantify, the average value of $d_{l \to m}$ and $d_{r \to m}$ was directly used to represent the visual attention allocation range of the participant.

Table 1 The multidimensional time series data for TICC-based mental fatigue identification.

Data source	**Description**
Subjective feelings	NASA score and SSS score of each TOO phase.
Excavation task performance	The productivity of excavation task per minute (number of excavator operating cycles/minute).
HDT performance	Hit rate (times/minute). A hit is a correct response to a hazardous situation.
	False alarm (FA) rate (times/minute); A FA is an erroneous response to a safety situation.
Objective eye movement behaviors	Pupil diameter (pixel), blink rate (times/minute), visual attention allocation range = $(d_{l\to m} + d_{r\to m})/2$

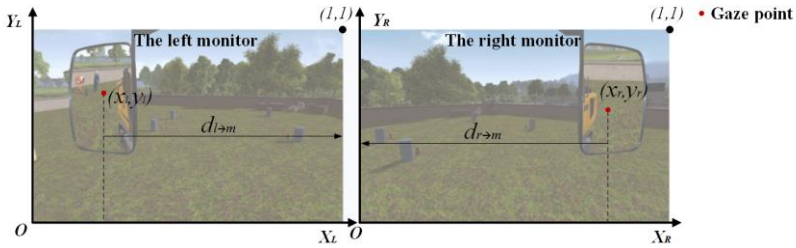

Fig. 5. Normalized coordinate systems of the left and right monitors.

Note: to illustrate the above settings clearly, the middle monitor is omitted.

To make the relevant time series data meet the requirements of the TICC algorithm, data interpolation and resampling technique were adopted for data preprocessing. Considering that the amount of missing data is small and most of the missing data is in a small time interval in which the data values are basically linear, a linear interpolation is used to complete the missing data. Furthermore, considering that the time series data were obtained at various sampling frequencies and the time scale for distinguishing changes in eye movement variables is less than 1 second, we set the data sampling frequency to 10 Hz. For variables with insufficient sampling frequency, such as subjective evaluation scores and task performance, upsampling are used to fill the data; for eye movement data with sampling frequency exceeding 10 Hz, down sampling was performed. In addition, the above data was normalized to eliminate individual differences (such as differences in pupil size) and to eliminate differences in characteristic properties between variables. After the pre-processing, a $8 \times T \times 10$ multivariate time series was obtained.

TICC was used to discover repeated patterns in the above time series. Once these patterns are recognized, the large set of time series data can be described by a limited number of mental fatigue states. As shown in Fig. 4, the TICC performs simultaneous segmentation and clustering on the time series of all participants. In this study, without considering the performance of the TICC algorithm, we set the window value as the minimum value 0.1s (1 sample). Refer to the preferred empirical values of the parameters in the TICC cases, the penalty factor β was selected from 10, 40, 100, 400 and regularization parameter λ was selected from 0.01, 0.1, and 0.5. Bayesian information criterion (BIC) is used as an objective criterion for evaluating and determining the optimization model and the appropriate clustering results. All data preprocessing, TICC-based analysis, and data visualization of the results were implemented using Python.

3.3. Mental fatigue classification using supervised learning

The relevant data required for the automated classification of mental fatigue was derived from the operator's eye movement data only. Based on the wearable eye-tracking technology, a variety of eye movement data-sets were collected, including pupil-related, blink-related, and gaze-related data. We performed basic data preprocessing on these data, including outlier deletion, noise data elimination, missing data interpolation and data resampling. The basic data used for supervised learning was a data set consisting of the above six participants' eye movement data, which had a sampling frequency of 0.2 Hz and a total of 4320 samples. The sliding window technique was used for data segmentation to convert sequential supervised learning problems into traditional supervised learning problems. The length of each feature extraction window was 10 s. And a 50% overlap of the adjacent windows was used. Based on the clustering results obtained in the previous step, preprocessed eye movement data were labeled by the corresponding level of mental fatigue.

As has been explained in Section 3.1.4, this study identified and selected three main eye movement feature sets that have been demonstrated to be related to mental fatigue, which are blink behavior, pupil state, and gaze point distribution, respectively. As shown in Table 2, blink behavior-related features include blink rate and blink duration. The pupil diameter was used as the primary pupil state feature in this paper, which has been widely used to assess individual mental fatigue. Gaze point distribution was divided into six features related to gaze point count and gaze point location, where the gaze point location was quantized by the coordinate values of the coordinate system shown in Fig. 5. In addition, task duration was also identified as a key feature that is directly related to the accumulation of mental fatigue.

Table 2 Definition and descriptions for eye movement-related features for supervised learning.

Feature set	Feature	Definition and description
Blink behavior	Blink rate (times/minute)	The number of blink events detected.
	Blink duration (ms)	Calculated from the start position of blink to the end value of blink.
Pupil measure	Pupil diameter (pixel)	Pupil size measured in pixels.
Gaze point	Horizontal gaze point location	The abscissa values of the gaze points in the left and right monitors.
	Vertical gaze point location	The ordinate values of the gaze points in the left and right monitors.
	Gaze count	Number of gaze points in the left and right monitors.

In order to construct the mental fatigue classification model, some features that reflect the characteristics of the above-mentioned data needed to be extracted. First, the data of the feature sets described above was normalized using the Z score normalization method. The features extracted from these standardized data include the mean value, variance, standard deviation, maximum, minimum, data range, and kurtosis. As a result, 70 features were extracted, including 14 blink behavior features, 7 pupil diameter features, 42 gaze point distribution features, and 7 task duration features. These extracted features were then used in the training and testing of supervised learning algorithms.

Different well-established classification algorithms were tested to select the classification algorithm that provides the highest accuracy, including Support Vector Machines (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), Boosted Tree (BT), and Linear Discriminant Analysis (LDA) [11]. Due to the limitation of the length of the paper, a detailed introduction to these algorithms is not provided. For more information, please refer to the machine learning related literature [59].

To determine the model parameters (e.g., features and algorithms) to achieve the best performance for classifying the multiple levels of mental fatigue, the performance of each algorithm was evaluated by cross-validation. Considering the practical application of the proposed method, the study used leave-one-subject-out (LOSO) cross-validation approach to evaluate the accuracy and validity of the classification algorithms. LOSO partitions the data of one subject as the test dataset, and the data of remaining subjects as the training dataset, which enables it to make full use of each subject's sample to reduce the subject bias caused by individual differences[60]. Therefore, it is very practical that LOSO requires training data from only one operator and does not require collecting new training data and retraining the classifier, before applying to a new operator. This advantage makes LOSO has been gradually applied to construction safety management, such as workers' activity recognition [61,62]. In this paper, the proposed classification algorithms and the classification evaluation were implemented in Python 3.7.

4. Results

4.1 The identified multiple levels of operators' mental fatigue

A total of 216,000 samples of data from 6 participants was used for TICC-based analysis. According to the BIC value of the results corresponding to the combination of different model parameters, we obtained an optimal combination of model parameters, where the penalty factor β was 400 and the regularization parameter λ was 0.01. As shown in Table 3, it was inferred that 3 with the smallest BIC value represents the appropriate number of clusters. Moreover, taking into account the actual needs of construction safety management, three levels might be a reasonable selection. The corresponding three-level mental fatigue intervention measures could be both effective and easy to implement on the construction site. Therefore, the levels of mental fatigue of the excavator operator in this study were determined to be 3.

Table 3 The BIC values corresponding to each K values (window size = 1, β = 400, λ = 0.01).

K	2	3	4	5	6
BIC	2162.63	**1610.46**	2046.54	2679.96	2527.65

In order to explain the three-level of operator mental fatigue obtained from the TICC-based analysis, the mean values of the points in each cluster are used to directly reflect the effects of different levels of mental fatigue on various aspects, such as subjective feelings, task performance, and eye movement behaviors. As shown in Table 4, although the data varies monotonically among three levels, the average trend and the difference between the levels was not consistent, reflecting the characteristics of different levels of mental fatigue. In Level 1 of mental fatigue, the operator had unapparent subjective feelings. In this level, operators performed both the excavation task and the hazard detection task well; the eye movement metrics also indicated that the operator was in a relaxed and energetic status. In the Level 2, the subjective feelings of the operator's mental fatigue increased significantly, and the objective eye movement behaviors also confirmed this; in this level, the task performances were reduced, especially the excavation performance dropped by more than 10%. The operator's subjective feelings of mental fatigue at Level 3 were further affected by fatigue. Their extremely high blink frequency and significantly reduced pupil diameter and visual attention range all showed great impact from mental fatigue. Meanwhile, both the performances of excavation task and hazard detection task had been significantly impaired from Level 2 to Level 3.

Table 4 The mean value of features of each level of mental fatigue.

Level	Subjective feelings		ET performance	HDT performance		Eye movement behaviors		
	NASA-TLX	SSS		Hit	FA	GD	PD	BR
1	0.150	0.180	0.834	0.903	0.015	0.598	0.568	0.451
2	0.327	0.417	0.736	0.825	0.036	0.537	0.463	0.664
3	0.515	0.693	0.707	0.614	0.057	0.448	0.424	0.859

Note: NASA-TLX= NASA Task Load Index; SSS= Stanford Sleepiness Scale; ET= Excavation task; HDT= Hazard detection task; FA=false alarm rate; GD= gaze distribution; BR= blink rate; PD= pupil diameter.

Furthermore, we can explain the above results from a perspective that more reflects the characteristics of construction practices. This study analyzes the characteristics of different levels of mental fatigue and its impact on operators from three dimensions: *Safety*, *Productivity*, and *Mental Fatigue Index*. The *Safety* dimension consists of HDT performance and gaze points distribution. Gaze points distribution indicates the operator's visual attention range, which has a significant impact on operational safety. The *Mental Fatigue Index* is determined by both the subjective feelings and objective symptom. *Productivity* is determined by excavation performance. The data values of the three dimensions varies from 0 to 1. The values are calculated from the average of the features constituting this dimension.

In the *Safety* dimension, 0 means unsafe and 1 refers to safe approach. 0 in the *Productivity* dimension indicates low excavation performance while 1 indicates high performance. For the *Mental Fatigue Index*, 0 indicates no mental fatigue, and 1 represents the highest value of mental fatigue.

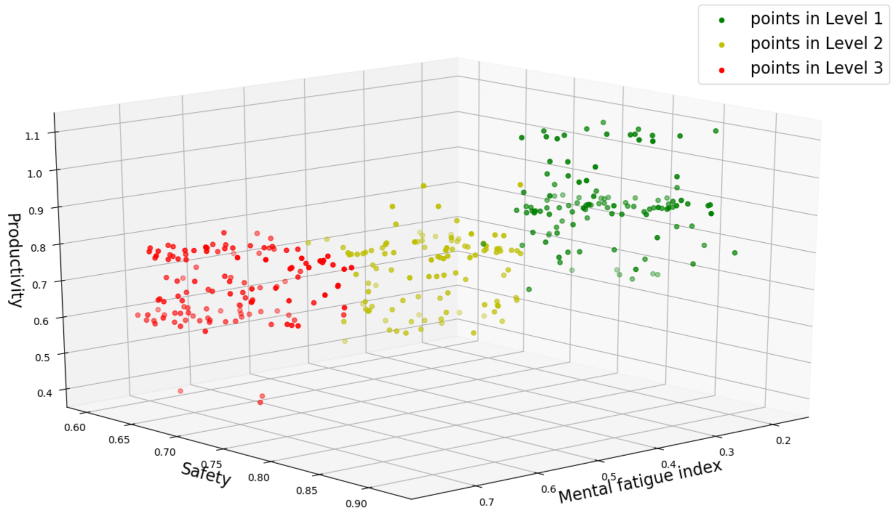

Fig. 6. Scatter plot of each mental fatigue level data in dimensions of *Safety*, *Productivity* and *Mental Fatigue Index*.

A three-dimensional scatterplot was drawn to intuitively demonstrate the characteristics of mental fatigue in the above three dimensions. As shown in Fig. 6, the operator in Level 1 is in a very low mental fatigue level, and its operational safety and productivity are significantly higher than the performance of Level 2 and Level 3. In Level 2, the operator's productivity was significantly reduced, operational safety was further reduced, and the mental fatigue index increased. This level shows that the operator is in a transition phase from low to high mental fatigue. At Level 3, the operator was in a very mentally exhausted phase, and the productivity is barely equal to Level 2, but operational safety is at a minimum level.

4.2 Evaluation of operators' mental fatigue classification algorithms

The eye movement data was labeled according to 3 levels of mental fatigue and the data set (with a total of 70 features that has been mentioned above) was used for the classification model. To acquire a more comprehensive understanding of the classification results, the classification performance of the proposed algorithms was evaluated by overall mean accuracy, macro average precision (macro-P), macro average recall (macro-R), and macro average F1 score (macro-F1). Additionally, four different combinations of features were selected to demonstrate the feasibility of wearable eye-tracking technology in various construction scenarios, which are summarized as follows.

(1) Feature set 1: Only pupil-based and blink-based features (21 features). In situations where the operator's operating position is not fixed (such as truck drivers who often get on and off the truck), gaze-based data is prone to loss. Therefore, the combination of features of pupil and blink is considered.

(2) Feature set 2: Only blink-based and gaze-based features (56 features). In extreme situations where there are frequent and intense illumination changes, since the pupil diameter is sensitive to such light changes, the pupil diameter-related features are likely to be invalid for mental fatigue classification. Therefore, the combination of blink and gaze features is considered.

(3) Feature set 3: All eye movement features (63 features). In situations where the external environment is relatively stable and the operator's operating position is relatively fixed (such as a tower crane operator), it is possible that all eye movement features are useful for mental fatigue detecting.

(4) Feature set 4: All eye movement and task duration features (70 features). Task duration is an important feature that reflects the accumulation of mental fatigue. Therefore, under the assumption that the task duration can be measured, we consider the model performance under the combination of all eye movement features and task duration features.

The classification performances of each supervised learning algorithm with these four feature sets are listed in Table 5. From the comparative evaluation based on LOSO cross-validation, it can be observed that the overall better-performing algorithms are SVM, LDA, and DT, and the classification performance of BT and KNN is lower than the first three algorithms. SVM had the most balanced classification performance and also achieved high performance across all four feature sets (accuracy is from 79.5% to 85.0%). LDA achieved the best classification performance in feature set 1 (86.0% accuracy), 2 (81.2% accuracy), and 3 (80.5% accuracy) compared with other four algorithms, but the classification performance with the feature set 4 was poor (78.8% accuracy). The performances of DT with the feature set 1, 2, and 3 were not as good as those of SVM and LDA. However, when using feature set 4, DT produced the best classification results compared with other algorithms (87.1% accuracy, 88.4% macro-P, 87.1% macro-R, and 86.6% macro-F1). In general, SVM has the best adaptability for all four feature sets while ensuring high mental fatigue detection performance. LDA has the best capability of mental fatigue detection for the feature set 1, 2, and 3. In this paper, confusion matrix and receiver operating characteristic (ROC) curve were used to further evaluate the performance of SVM and LDA with all four feature sets.

Table 5 Classification performance of the proposed classification algorithms with different feature sets.

Algorithm	Feature set	Classification performance (%)			
		Accuracy	macro-P	macro-R	macro-F1
SVM	1	85.0	86.6	85.9	85.1
	2	81.3	82.0	81.4	80.9
	3	79.5	80.6	79.6	79.0
	4	84.1	85.7	84.5	83.6
Decision Tree	1	78.4	80.3	80.4	78.8
	2	79.4	80.1	79.9	79.2
	3	79.7	80.1	80.2	79.3
	4	87.1	88.4	87.1	86.6
Boosted Tree	1	81.0	81.2	82.6	79.8
	2	71.5	73.6	74.2	69.9
	3	73.6	75.3	80.0	72.8
	4	75.4	75.6	72.4	71.0
KNN	1	76.5	78.1	76.8	76.3
	2	63.4	63.8	63.6	62.6
	3	63.9	64.8	64.3	63.6
	4	73.9	74.5	73.7	73.1
LDA	1	86.0	87.7	86.7	86.1
	2	81.2	82.4	81.3	80.8
	3	80.5	82.0	80.8	80.2
	4	78.8	80.0	81.1	75.8

Confusion matrix is a specific visualization of the performance of each algorithm, which allows easy identification of confusion between different classes. In this study, normalized confusion matrix was used to achieve a more visual interpretation of which level is being misclassified (Fig. 7 and Fig. 8). Each column of the matrix represents the instances in a predicted mental fatigue level while each row represents the instances in a true level. The values of the diagonal elements represent the proportion of correctly predicted levels.

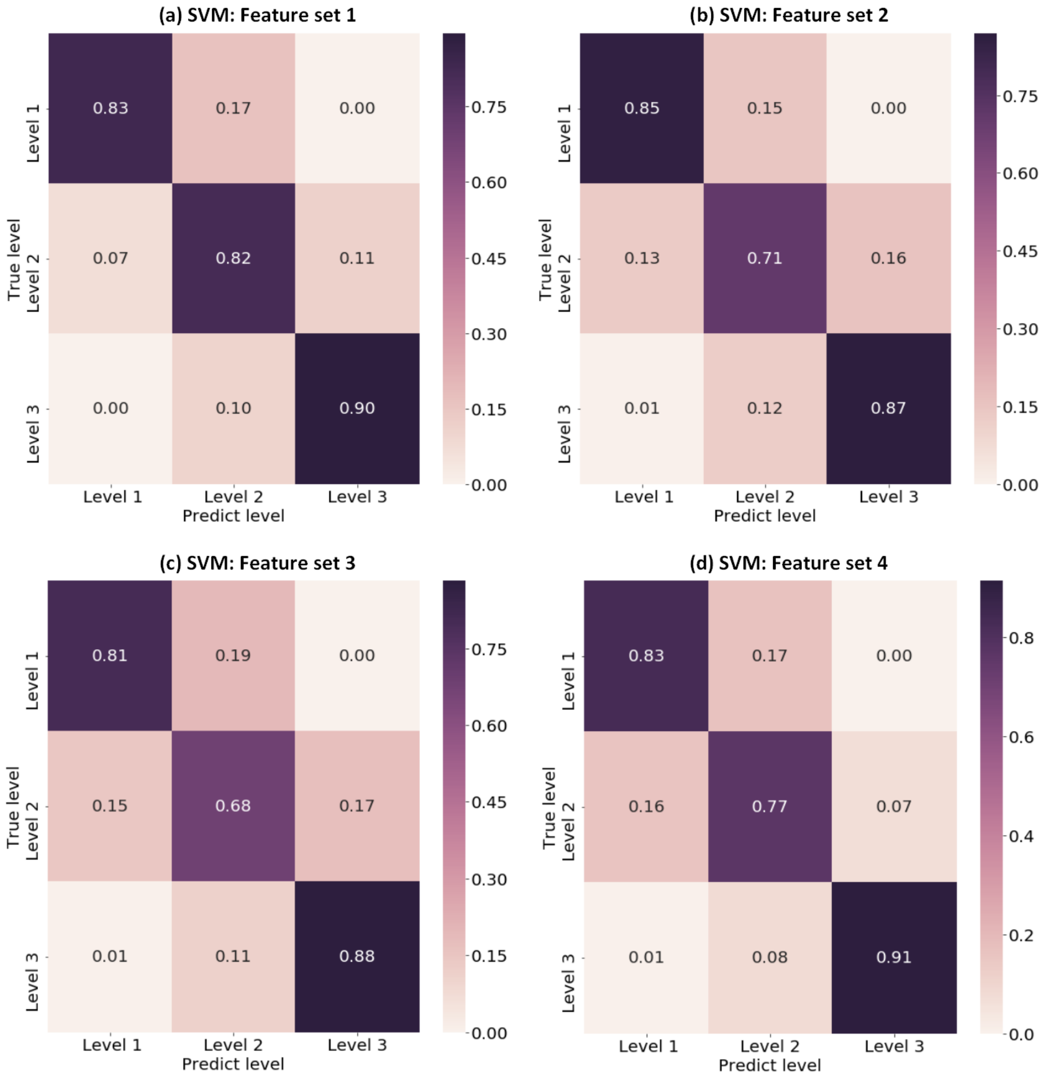

Fig. 7. Normalized confusion matrixes of SVM.

Fig. 7 (a-d) show the accuracy of SVM predicting three levels of mental fatigue under the four feature sets. Considering the classification results in all feature sets together, SVM yielded the average accuracy of 83%, 75%, and 89% for mental fatigue level 1, 2, and 3, respectively. Fig.8 (a-d) present the accuracy of LDA predicting each mental fatigue level with the four feature sets. The average accuracies of LDA for mental fatigue level 1, 2, and 3 under all four feature sets were 85%, 70%, and 89%, respectively. The above results suggested that although both algorithms showed excellent mental fatigue classification performance, their correct prediction rates for the intermediate-level mental fatigue (level 2) were lower than those for other levels. For example, SVM achieved the classification accuracy of 68% for mental fatigue level 2 with feature set 3 (Fig. 7(c)), and LDA had a lower accuracy of 56 % for mental fatigue level 2 with feature set 4 (56%, as shown in Fig.8 (d)), which means that the probability of misclassification was high when detecting mental fatigue level 2. This might be attributed that mental fatigue is a continuous cumulative process, and the corresponding data is also continuous, resulting in a decrease in the accuracy of predicting some intermediate states adjacent to mental fatigue level 1 or level 3.

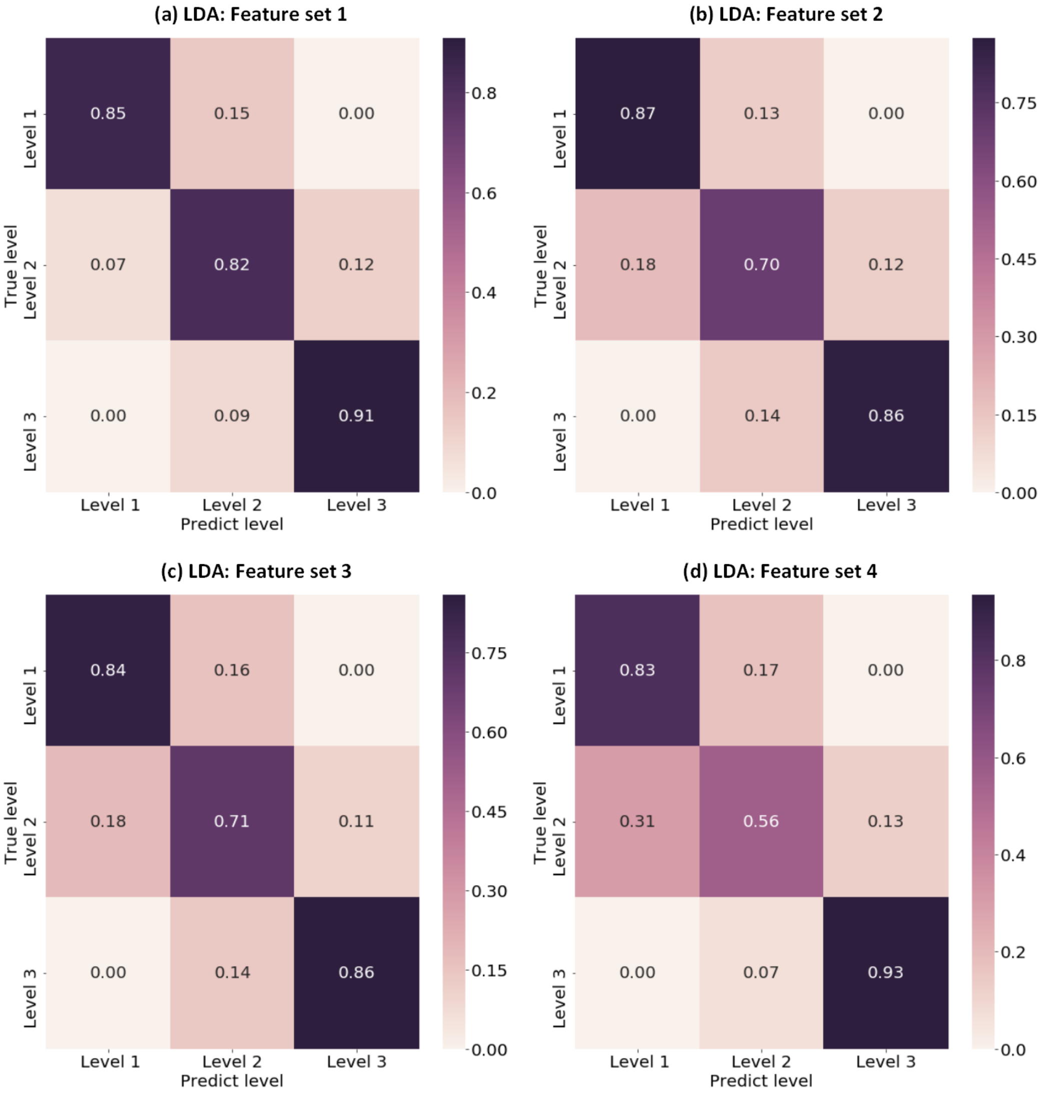

Fig. 8. Normalized confusion matrixes of LDA.

ROC curve is usually applied as a visual tool to illustrate the relationship between true positive rate (TPR, or sensitivity) and false positive rate (FPR, or 1-specificity) along with the change of a threshold parameter [63]. The area under the ROC curve (AUC) is a robust measure to compare the performance of different features or classification algorithms because it is not sensitive to the proportion of true cases. Generally, a classification model is considered to be effective when the AUC is greater than 0.9, a larger AUC indicates a better classification performance. Since six participants were corresponding to six LOSO cross-validation folds, the ROC curves of each fold of the SVM and LDA algorithms with all four feature sets are illustrated in Fig. 9 and Fig. 10, respectively.

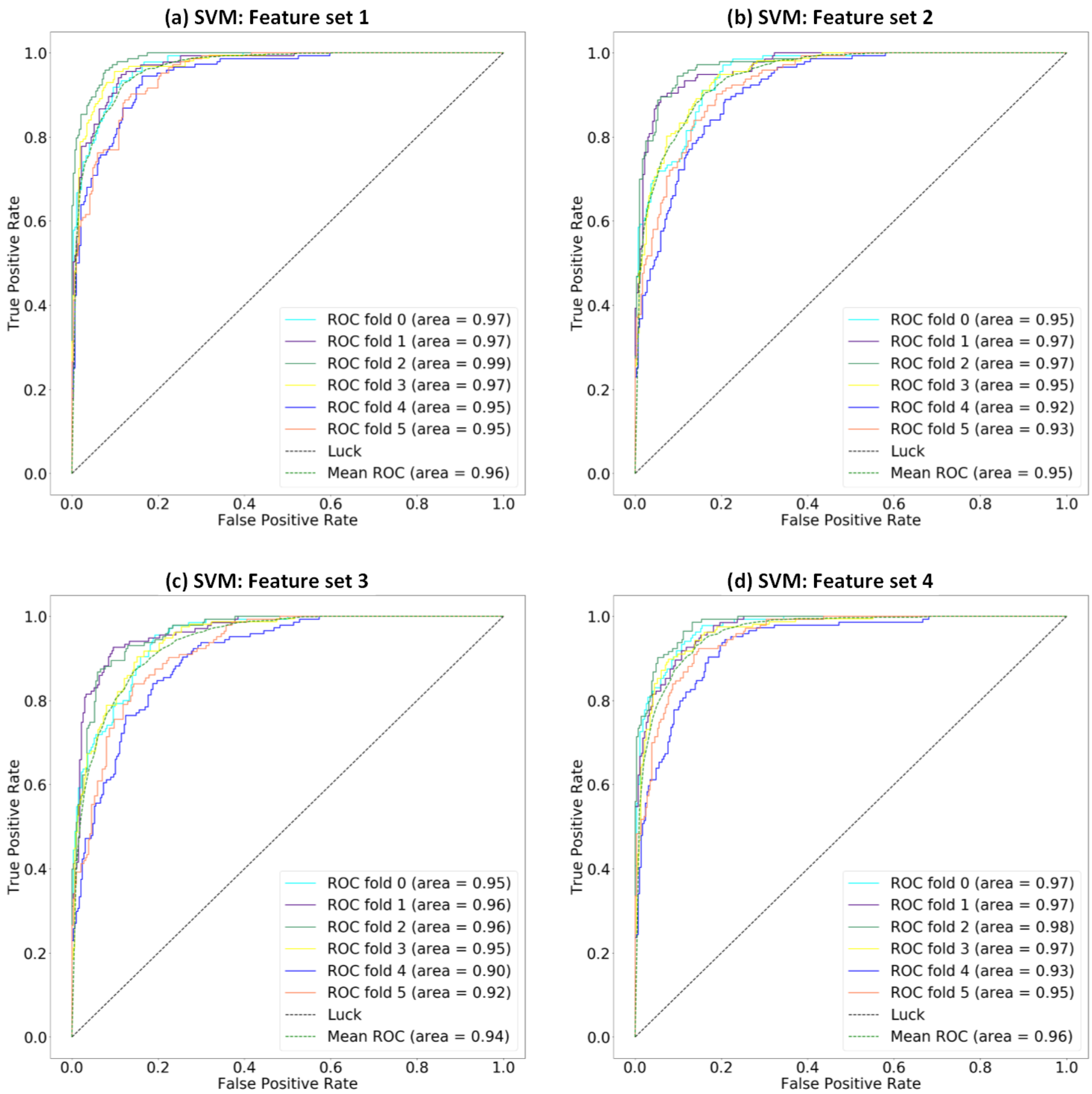

Fig. 9. ROC curves of SVM algorithm under all four feature sets.

Note: "ROC fold 0" indicates that the ROC curve corresponds to the case where the test data is from subject 1 while the training data is from the remaining five subjects, and so on.

Fig. 9 (a-d) and Fig.10 (a-d) illustrate ROC curves of SVM and LDA for the proposed four feature sets respectively. The characteristics of each LOSO cross-validation fold are compared using their corresponding ROC curves, which could help us to identify the difference in mental fatigue detection performance of the classification algorithm in the face of different individuals. From the comparison of the ROC curves corresponding to each fold, SVM had a better ability to suppress the subject bias generated by individual differences than LDA. By using the feature set 4, LDA algorithm achieved a minimum AUC of 0.88 in LOSO cross-validation fold 1 (Fig 10 (d)). Hence, although the fold 2 of LDA with the feature set 4 yielded a high AUC of 0.97, the mean AUC of the six folds was only 0.92, which indicates that LDA is less adaptable to subject bias than SVM. Furthermore, to evaluate the overall classification performance of each algorithm, their average ROC curves in various feature sets were illustrated and the corresponding average AUCs were computed. It can be observed that compared to LDA, SVM can achieve higher AUC for all four feature sets (0.96 for “pupil+blink”, 0.95 for “blink+gaze”, 0.94 for “all eye movement”, and 0.96 for “eye movement+task duration”), which once again supports the statement that SVM has a better adaptability to various construction scenarios mentioned above.

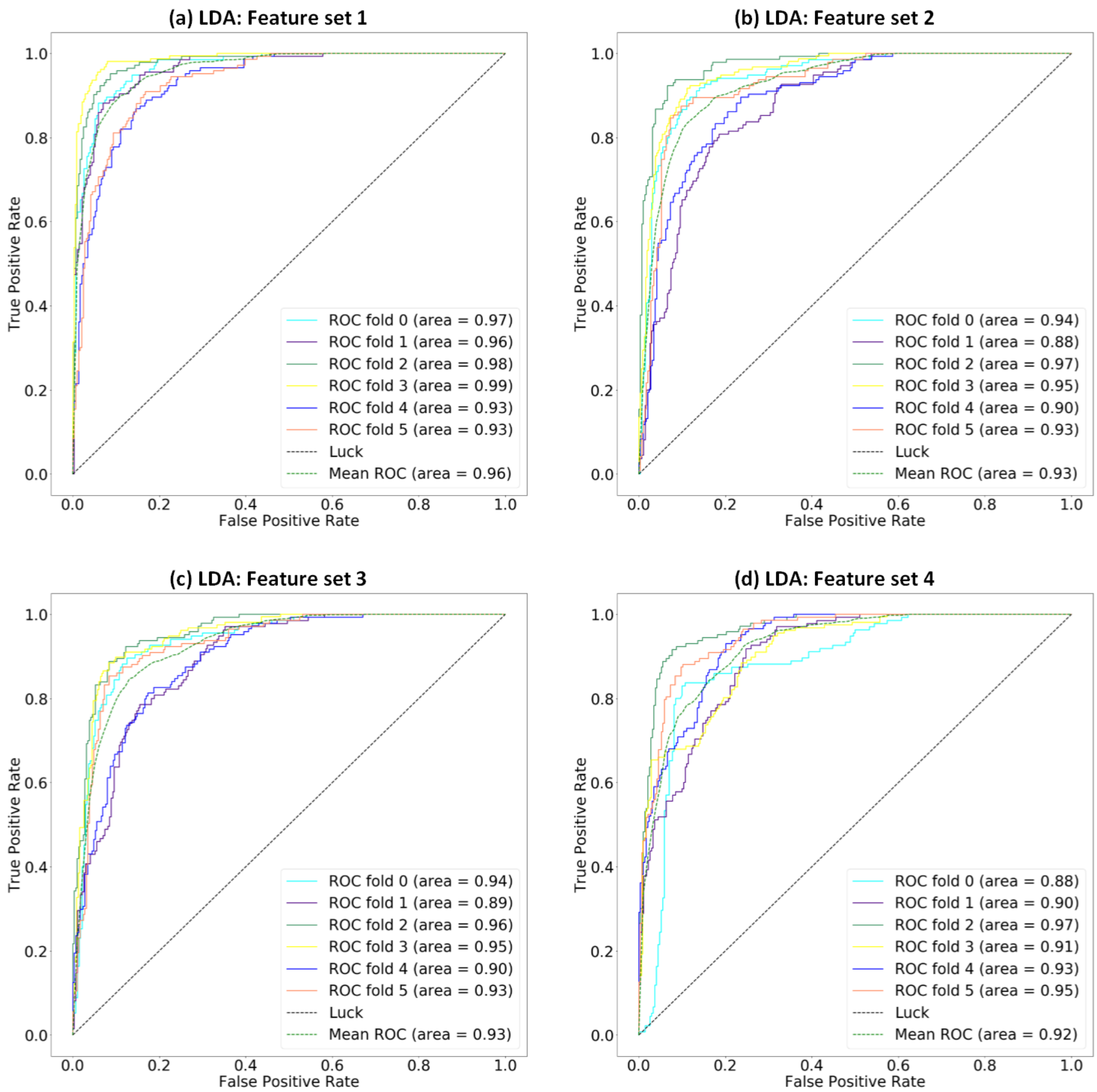

Fig. 10. ROC curves of LDA algorithm under all four feature sets.

Note: "ROC fold 0" indicates that the ROC curve corresponds to the case where the test data is from subject 1 while the training data is from the remaining five subjects, and so on.

5. Discussion

5.1. Contributions and implications

The results highlight that wearable eye-tracking technology has achieved good mental fatigue detection performance in several widely used classification algorithms. Especially for the SVM and LDA, high classification accuracy was obtained. Regardless of the various feature set used, the accuracies of mental fatigue classification with the SVM algorithm was at least 80%. Additionally, the results of LOSO cross-validation further indicate that SVM can better alleviate the low classification accuracy caused by subject bias, which is very important for applying the proposed method to the actual construction site. With regards to the feature sets, the findings indicate the effectiveness of four feature sets for predicting multi-level mental fatigue. Interestingly, "pupil + blink" feature combination revealed a high classification accuracy (85.0% for SVM and 86.0% for LDA) with only 21 features. This emphasizes that for some budget-constrained construction projects, even low-cost webcams can be used to capture pupil and blink data for automatic mental fatigue detection. For the feature sets, the results highlight that by combining different features, the adverse site conditions which could hinder the usage of certain features, can be mitigated.

In addition to highlighting the usefulness of eye-tracking technology for operator's mental fatigue, the experimental results also reveal some interesting observations. First, the different levels of mental fatigue had varying effecting on the operator's productivity and safety performance, as well as the operator's coping strategy to mental fatigue. While mental fatigue increases from Level 1 to Level 3, operators were found to be compromising on task safety and productivity (Fig. 6). Second, the decline rates of the safety and productivity dimensions are very different, the results showed that HDT performance declined rapidly, while excavation performance depicted a slight decline. The operator seems to put more energy into the excavation task and pay less attention to the hazard detection. Many relevant studies have reached similar conclusions [64,65]. At the real construction site, although ensuring construction safety is the primary goal, when an operator is suffering from mental fatigue along with a schedule pressure, he/she is more likely to devote more energy to complete construction tasks. Under such macro conditions, the operator's ability to detect hazards is often greatly impaired, resulting in a variety of safety risks.

The above results and discussion can serve as the basis for designing multi-level intervention strategies for three levels of operator mental fatigue. Since different levels of mental fatigue have different characteristics and different influences on operator's task performance, their corresponding intervention strategies might be very different, such as auditory attentional cueing [66], meditation [67], neuropsychological recovery [68], etc. As such, combining engineering psychology, ergonomics and other disciplines to conduct targeted grading and adapting interventions for construction equipment operators' mental fatigue could be a meaningful research direction in the future.

5.2. Limitations and future works

First, the data sources used for TICC-based analysis were diverse. Eye movement data had a high sampling frequency. However, subjective evaluation data had a low sampling frequency. Low sampling frequency results in sparse data samples. Hence data upsampling technique was used to fill the sample, but such an approach might affect the data quality and the corresponding results. Moreover, more eye movement metrics associated with mental fatigue need to be discovered and applied to further improve the detection performance of multi-level mental fatigue. Second, the experimental results from the 60-min simulation experiment task may not fully reflect all levels of mental fatigue of operators who work more than 8 hours a day in real construction tasks, such as extreme fatigue or even the state of drowsiness. In the future, we will consider collecting the mental fatigue data of operators who perform a long-term operational task in the real construction site. Furthermore, the real construction site environment is more dynamic and complex than the laboratory experiment environment, which may lead to some different behavioral characteristics and task performance of operators under the influence of mental fatigue. At the same time, operators in the real-time construction site may feel more physical fatigue, and the interaction between physical fatigue and mental fatigue may make the equipment operation safety problem more complicated. In the case that ICT technology is widely used in current construction projects [69-71], the future work will focus on obtaining high sampling frequency multi-modal data through other effective ICT technologies, and using the "TICC + supervised learning" method for not only the mental fatigue detection but also a wider range of pattern discovery, identification and detection problems at the construction site.

6. Conclusions

Detecting construction equipment operator's mental fatigue has great potential in reducing the risk of caught-in-between and struck-by hazards. The present research proposed an automated mental fatigue identification and classification method for operators' mental fatigue using wearable eye-tracking technology. Toeplitz Inverse Covariance-Based Clustering (TICC) method was successfully used to determine multiple levels of mental fatigue and to complete the labeling of relevant eye movement data. Considering that different types of eye movement data may be invalidated in different construction scenarios, four corresponding types of eye movement feature sets were extracted to examine the feasibility of the wearable eye-tracking technology. The results depicted that SVM and LDA algorithms had better detection performance when compared to the other three commonly used supervised learning algorithms. The high classification accuracies corresponding to the four feature sets demonstrated that eye-tracking technology had a great potential to be used for various construction site conditions. Additionally, the results also revealed that the operator's operational safety performance and productivity have experienced different degrees of decline under the influence of different levels of mental fatigue. Considering the limitations of simulation-based experiment and the practicality of the proposed method, future research will focus on the verification and application of the proposed method in the real construction site and the development of multi-level intervention strategies for construction equipment operators' mental fatigue.

Acknowledgment

This research study was partially supported by The Hong Kong Polytechnic University, the General Research Fund (GRF) Grant (BRE/PolyU 152099/18E) entitled “Proactive Monitoring of Work-Related MSD Risk Factors and Fall Risks of Construction Workers Using Wearable Insoles” and the project (PolyU 152047/19E) entitled “In search of a suitable tool for proactive physical fatigue assessment: an invasive to non-invasive approach. Besides, this work was partially supported by the National Natural Science Foundation of China (Grants 71390524, 71821001).

Reference

[1] A.S. Wagstaff, J.A. Sigstad Lie, Shift and night work and long working hours--a systematic review of safety implications, Scandinavian Journal of Work, Environment & Health 37 (3) (2011) 173-185. doi: 10.5271/sjweh.3146.

[2] T.W. Wong, Occupational injuries among construction workers in Hong Kong, Occupational Medicine 44 (5) (1994) 247-252. doi: 10.1093/occmed/44.5.247.

[3] M. Zhang, L.A. Murphy, D. Fang, A.J. Caban-Martinez, Influence of fatigue on construction workers' physical and cognitive function, Occupational Medicine 65 (3) (2015) 245-250. doi: 10.1093/occmed/kqu215.

[4] J. Li, H. Li, H. Wang, W. Umer, H. Fu, X. Xing, Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology, Automation in Construction 105 (2019). doi: 10.1016/j.autcon.2019.102835.

[5] Y.H. Fang, Y.K. Cho, Effectiveness analysis from a cognitive perspective for a real-time safety assistance system for mobile crane lifting operations, Journal of Construction Engineering and Management 143 (4) (2017). doi: 10.1061/(Asce)Co.1943-7862.0001258.

[6] J.W. Hinze, J. Teizer, Visibility-related fatalities related to construction equipment, Safety Science 49 (5) (2011) 709-718. doi: 10.1016/j.ssci.2011.01.007.

[7] A. Shapira, Y. Rosenfeld, I. Mizrah, Vision system for tower cranes, Journal of Construction Engineering and Management 134 (5) (2008) 320-332. doi: 10.1061/(Asce)0733-9364(2008)134:5(320).

[8] United States Department of Labor, Commonly used statistics, Available from: https://www.osha.gov/oshstats/commonstats.html, 2016, Accessed date: November 2018.

[9] A. Murataa, A. Uetakeb, Y. Takasawab, Evaluation of mental fatigue using feature parameter extracted from event-related potential, International Journal of Industrial Ergonomics 35 (8) (2005) 761-770. doi: 10.1016/j.ergon.2004.12.003.

[10] T. Akerstedt, J. Axelsson, M. Lekander, N. Orsini, G. Kecklund, Do sleep, stress, and illness explain daily variations in fatigue? A prospective study, Journal of Psychosomatic Research 76 (4) (2014) 280-285. doi: 10.1016/j.jpsychores.2014.01.005.
[11] A. Aryal, A. Ghahramani, B. Becerik-Gerber, Monitoring fatigue in construction workers using physiological measurements, Automation in Construction 82 (2017) 154-165. doi: 10.1016/j.autcon.2017.03.003.
[12] M. Chan, Fatigue: the most critical accident risk in oil and gas construction, Construction Management and Economics 29 (4) (2011) 341-353. doi: 10.1080/01446193.2010.545993.
[13] D. Fang, Z. Jiang, M. Zhang, H. Wang, An experimental method to study the effect of fatigue on construction workers' safety performance, Safety Science 73 (2015) 80-91. doi: 10.1016/j.ssci.2014.11.019.
[14] U. Techera, M. Hallowell, R. Littlejohn, S. Rajendran, Measuring and predicting fatigue in construction: Empirical field study, Journal of Construction Engineering and Management 144 (8) (2018). doi: 10.1061/(asce)co.1943-7862.0001513.
[15] M. Zhang, E.H. Sparer, L.A. Murphy, J.T. Dennerlein, D. Fang, J.N. Katz, A.J. Caban-Martinez, Development and validation of a fatigue assessment scale for U.S. construction workers, American Journal of Industrial Medicine 58 (2) (2015) 220-228. doi: 10.1002/ajim.22411.
[16] X. Wang, C. Xu, Driver drowsiness detection based on non-intrusive metrics considering individual specifics, Accident Analysis Prevention 95 (Pt B) (2016) 350-357. doi: 10.1016/j.aap.2015.09.002.
[17] Y. Yamada, M. Kobayashi, Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults, Artificial Intelligence Medicine (2018). doi: 10.1016/j.artmed.2018.06.005.
[18] S.G. Hart, NASA-task load index (NASA-TLX); 20 years later, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 50, Sage Publications Sage CA: Los Angeles, CA, 2006, pp. 904-908. doi: 10.1177/154193120605000909.
[19] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, W. Dement, Quantification of sleepiness: a new approach, Psychophysiology 10 (4) (1973) 431-436. doi: 10.1111/j.1469-8986.1973.tb00801.x.
[20] L.S. Aaronson, C.S. Teel, V. Cassmeyer, G.B. Neuberger, L. Pallikkathayil, J. Pierce, A.N. Press, P.D. Williams, A. Wingate, Defining and measuring fatigue, Image: Journal of Nursing Schofarship 31 (1) (1999) 45-50. doi: 10.1111/j.1547-5069.1999.tb00420.x.
[21] K.Q. Shen, X.P. Li, C.J. Ong, S.Y. Shao, E.P.V. Wilder-Smith, EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate, Clinical Neurophysiology 119 (7) (2008) 1524-1533. doi: 10.1016/j.clinph.2008.03.012.
[22] M.A. Boksem, T.F. Meijman, M.M. Lorist, Effects of mental fatigue on attention: an ERP study, Cognitive Brain Research 25 (1) (2005) 107-116. doi: 10.1016/j.cogbrainres.2005.04.011.
[23] D. Wang, J. Chen, D. Zhao, F. Dai, C. Zheng, X. Wu, Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system, Automation in Construction 82 (2017) 122-137. doi: 10.1016/j.autcon.2017.02.001.
[24] T.C. Chieh, M.M. Mustafa, A. Hussain, S.F. Hendi, B.Y. Majlis, Development of vehicle driver drowsiness detection system using electrooculogram (EOG), 2005 1st International Conference on Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering (CCSP), IEEE, Kuala Lumpur, Malaysia, 2005, pp. 165-168. doi: 10.1109/CCSP.2005.4977181.
[25] C. Zhao, M. Zhao, J. Liu, C. Zheng, Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator, Accident Analysis and Prevention 45 (2012) 83-90. doi: 10.1016/j.aap.2011.11.019.

[26] D. Dawson, A.K. Searle, J.L. Paterson, Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry, Sleep Medicine Reviews 18 (2) (2014) 141-152. doi: 10.1016/j.smrv.2013.03.003.

[27] C. Conati, Probabilistic assessment of user's emotions in educational games, Applied Artificial Intelligence 16 (7-8) (2002) 555-575. doi: 10.1080/08839510290030390.

[28] B. Somers, T. Francart, A. Bertrand, A generic EEG artifact removal algorithm based on the multi-channel Wiener filter, Journal of Neural Engineering 15 (3) (2018) 036007. doi: 10.1088/1741-2552/aaac92.

[29] S. Hasanzadeh, B. Esmaeili, M.D. Dodd, Examining the relationship between construction workers' visual attention and situation awareness under fall and tripping hazard conditions: Using mobile eye tracking, Journal of Construction Engineering and Management 144 (7) (2018). doi: 10.1061/(asce)co.1943-7862.0001516.

[30] I. Jeelani, A. Albert, K. Han, R. Azevedo, Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology, Journal of Construction Engineering and Management 145 (1) (2019) 04018115. doi: 10.1061/(asce)co.1943-7862.0001589.

[31] J.M. Morales, C. Diaz-Piedra, H. Rieiro, J. Roca-Gonzalez, S. Romero, A. Catena, L.J. Fuentes, L.L. Di Stasi, Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data, Accident Analysis and Prevention 109 (2017) 62-69. doi: 10.1016/j.aap.2017.09.025.

[32] C. Chen, K. Li, Q. Wu, H. Wang, Z. Qian, G. Sudlow, EEG-based detection and evaluation of fatigue caused by watching 3DTV, Displays 34 (2) (2013) 81-88. doi: 10.1016/j.displa.2013.01.002.

[33] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, F. Babiloni, Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness, Neuroscience and Biobehavioral Reviews 44 (2014) 58-75. doi: 10.1016/j.neubiorev.2012.10.003.

[34] Z. Yin, J. Zhang, Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine, Neurocomputing 283 (2018) 266-281. doi: 10.1016/j.neucom.2017.12.062.

[35] J. Chen, H. Wang, C. Hua, Electroencephalography based fatigue detection using a novel feature fusion and extreme learning machine, Cognitive Systems Research (2018). doi: 10.1016/j.cogsys.2018.08.018.

[36] C. Zhao, C. Zheng, M. Zhao, Y. Tu, J. Liu, Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic, Expert Systems with Applications 38 (3) (2011) 1859-1865. doi: 10.1016/j.eswa.2010.07.115.

[37] Y. Yamada, M. Kobayashi, Fatigue detection model for older adults using eye-tracking data gathered while watching video: Evaluation against diverse fatiguing tasks, 2017 International Conference on Healthcare Informatics (ICHI), IEEE, Park City, UT, USA, 2017, pp. 275-284. doi: 10.1109/ICHI.2017.74.

[38] A.D. McDonald, J.D. Lee, C. Schwarz, T.L. Brown, Steering in a random forest: ensemble learning for detecting drowsiness-related lane departures, Human Factors 56 (5) (2014) 986-998. doi: 10.1177/0018720813515272.

[39] A.D. McDonald, J.D. Lee, C. Schwarz, T.L. Brown, A contextual and temporal algorithm for driver drowsiness detection, Accident Analysis Prevention 113 (2018) 25-37. doi: 10.1016/j.aap.2018.01.005.

[40] F. Laurent, M. Valderrama, M. Besserve, M. Guillard, J.-P. Lachaux, J. Martinerie, G. Florence, Multimodal information improves the rapid detection of mental fatigue, Biomedical Signal Processing and Control 8 (4) (2013) 400-408. doi: 10.1016/j.bspc.2013.01.007.

[41] H. Jebelli, M.M. Khalili, S. Hwang, S. Lee, A supervised learning-based construction workers' stress recognition using a wearable electroencephalography (EEG) device, Construction Research Congress 2018, 2018, pp. 40-50. doi: 10.1061/9780784481288.005.

[42] C. Jacobe de Naurois, C. Bourdin, A. Stratulat, E. Diaz, J.L. Vercher, Detection and prediction of driver drowsiness using artificial neural network models, Accident Analysis Prevention (2017). doi: 10.1016/j.aap.2017.11.038.

[43] L.G. Faber, N.M. Maurits, M.M. Lorist, Mental fatigue affects visual selective attention, PLoS One 7 (10) (2012) e48073. doi: 10.1371/journal.pone.0048073.

[44] J.F. Hopstaken, D. van der Linden, A.B. Bakker, M.A.J. Kompier, Y.K. Leung, Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data, Journal of Experimental Psychology: Human Perception Performance 42 (6) (2016) 878-889. doi: 10.1037/xhp0000189.

[45] M.A. Boksem, M. Tops, Mental fatigue: costs and benefits, Brain Research Reviews 59 (1) (2008) 125-139. doi: 10.1016/j.brainresrev.2008.07.001.

[46] M. Kassner, W. Patera, A. Bulling, Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct, 2014, pp. 1151-1160. doi: 10.1145/2638728.2641695.

[47] W. Song, F.L. Woon, A. Doong, C. Persad, L. Tijerina, P. Pandit, C. Cline, B. Giordani, Fatigue in younger and older drivers: Effectiveness of an alertness-maintaining task, Human Factors 59 (6) (2017) 995-1008. doi: 10.1177/0018720817706811.

[48] X. Wanyan, D. Zhuang, Y. Lin, X. Xiao, J.-W. Song, Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation, International Journal of Industrial Ergonomics 64 (2018) 1-7. doi: 10.1016/j.ergon.2017.08.004.

[49] V. Faure, R. Lobjois, N. Benguigui, The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior, Transportation Research Part F: Traffic Psychology and Behaviour 40 (2016) 78-90. doi: 10.1016/j.trf.2016.04.007.

[50] E.T. Greenlee, P.R. DeLucia, D.C. Newton, Driver vigilance in automated vehicles: Hazard detection failures are a matter of time, Human Factors 60 (4) (2018) 465-476. doi: 10.1177/0018720818761711.

[51] Y. Xiao, F. Ma, Y. Lv, G. Cai, P. Teng, F. Xu, S. Chen, Sustained attention is associated with error processing impairment: evidence from mental fatigue study in four-choice reaction time task, PLoS One 10 (3) (2015) e0117837. doi: 10.1371/journal.pone.0117837.

[52] Z. Guo, R. Chen, K. Zhang, Y. Pan, J. Wu, The impairing effect of mental fatigue on visual sustained attention under monotonous multi-object visual attention task in long durations: An event-related potential based study, PLoS One 11 (9) (2016) e0163360. doi: 10.1371/journal.pone.0163360.

[53] J.F. Hopstaken, D. van der Linden, A.B. Bakker, M.A.J. Kompier, A multifaceted investigation of the link between mental fatigue and task disengagement, Psychophysiology 52 (3) (2015) 305-315. doi: 10.1111/psyp.12339.

[54] J. He, W. Choi, Y. Yang, J. Lu, X. Wu, K. Peng, Detection of driver drowsiness using wearable devices: A feasibility study of the proximity sensor, Applied Ergonomics 65 (2017) 473-480. doi: 10.1016/j.apergo.2017.02.016.

[55] J.A. Stern, D. Boyer, D. Schroeder, Blink rate: a possible measure of fatigue, Human Factors 36 (2) (1994) 285-297. doi: 10.1177/001872089403600209.

[56] R. Schleicher, N. Galley, S. Briest, L. Galley, Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?, Ergonomics 51 (7) (2008) 982-1010. doi: 10.1080/00140130701817062.

[57] R. Zargari Marandi, P. Madeleine, O. Omland, N. Vuillerme, A. Samani, Eye movement characteristics reflected fatigue development in both young and elderly individuals, Scientific Reports 8 (1) (2018) 13148.

doi: 10.1038/s41598-018-31577-1.
[58] D. Hallac, S. Vare, S. Boyd, J. Leskovec, Toeplitz Inverse Covariance-Based Clustering of multivariate time series data, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Vol. 2017, Halifax, NS, Canada, 2017, pp. 215-223. doi: 10.1145/3097983.3098060.
[59] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data mining: Practical machine learning tools and techniques, Fourth Edition ed., Morgan Kaufmann, Cambridge, MA 02139, United States, 2017. isbn: 9780128042915. doi: https://doi.org/10.1016/C2015-0-02071-8.
[60] M. Esterman, B.J. Tamber-Rosenau, Y.C. Chiu, S. Yantis, Avoiding non-independence in fMRI data analysis: leave one subject out, Neuroimage 50 (2) (2010) 572-576. doi: 10.1016/j.neuroimage.2009.10.092.
[61] R. Akhavian, A.H. Behzadan, Smartphone-based construction workers' activity recognition and classification, Automation in Construction 71 (2016) 198-209. doi: 10.1016/j.autcon.2016.08.015.
[62] N.D. Nath, T. Chaspari, A.H. Behzadan, Automated ergonomic risk monitoring using body-mounted sensors and machine learning, Advanced Engineering Informatics 38 (2018) 514-526. doi: 10.1016/j.aei.2018.08.020.
[63] L.L. Chen, Y. Zhao, P.F. Ye, J. Zhang, J.Z. Zou, Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers, Expert Systems with Applications 85 (2017) 279-291. doi: 10.1016/j.eswa.2017.01.040.
[64] S. Folkard, P. Tucker, Shift work, safety and productivity, Occupational Medicine 53 (2) (2003) 95-101. doi: doi.org/10.1093/occmed/kqg047.
[65] A. Williamson, D.A. Lombardi, S. Folkard, J. Stutts, T.K. Courtney, J.L. Connor, The link between fatigue and safety, Accident Analysis and Prevention 43 (2) (2011) 498-515. doi: 10.1016/j.aap.2009.11.011.
[66] C.D. Wickens, J.G. Hollands, S. Banbury, R. Parasuraman, Engineering psychology & human performance, Psychology Press, 2015. isbn: 9780205021987. doi: 10.4324/9781315665177.
[67] S. Kaplan, Meditation, restoration, and the management of mental fatigue, Environment and Behavior 33 (4) (2001) 480-506. doi: Doi 10.1177/00139160121973106.
[68] X. Xing, H. Li, J. Li, B. Zhong, H. Luo, M. Skitmore, A multicomponent and neurophysiological intervention for the emotional and mental states of high-altitude construction workers, Automation in Construction 105 (2019). doi: 10.1016/j.autcon.2019.102836.
[69] H.T. Zhou, H.W. Wang, W. Zeng, Smart construction site in mega construction projects: A case study on island tunneling project of Hong Kong-Zhuhai-Macao Bridge, Frontiers of Engineering Management 5 (1) (2018) 78-87. doi: 10.15302/J-Fem-2018075.
[70] J.J.S. Shi, S.X. Zeng, X.H. Meng, Intelligent data analytics is here to change engineering management, Frontiers of Engineering Management 4 (1) (2017) 41-48. doi: 10.15302/J-Fem-2017003.
[71] R. Edirisinghe, Digital skin of the construction site Smart sensor technologies towards the future smart construction site, Engineering Construction and Architectural Management 26 (2) (2019) 184-223. doi: 10.1108/Ecam-04-2017-0066.