Journal of Building Engineering 98 (2024) 111146

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Visual fatigue prediction using classification model based on physiological responses of occupants under office lightings

Dahyun Jung [a], Jongbaek An [a], Taehoon Hong [a,*], Minhyun Lee [b]

[a] Department of Architecture and Architectural Engineering, Yonsei University, Seoul, 03722, Republic of Korea
[b] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Visual fatigue
Classification model
Prediction
Physiological response
Lighting

ABSTRACT

Appropriate management of visual fatigue is crucial for eye health as it significantly impacts life quality throughout an individual's life cycle. The shift to digital tasks in modern office environments has increased occupants' exposure to visual fatigue, leading to various social problems. This study aimed to develop visual fatigue prediction models based on physiological responses and classification algorithms. Experiments were conducted to collect physiological responses and subjective visual fatigue under various lighting environments. Collected data was refined and reorganized as predictor variables by different time windows and target variables by scales. Then, visual fatigue prediction models were developed using supervised machine learning algorithms (i.e., artificial neural network, support vector machine, gradient boosting machine, and random forest). And improved by feature selection and adding subject-label variables. The resulting two-scale and three-scale visual fatigue prediction models demonstrated an average performance of 93.73 % and 94.64 % respectively. This research can contribute to reducing societal costs and enhancing productivity by proposing visual fatigue prediction models that help manage eye health in office environments.

1. Introduction

Eye health matters throughout a human's entire lifespan from infancy to old age [1]. The eyes are sensory organs that detect light and transmit visual information to the brain, and humans acquire 90 % of the information about the external environment through the eyes [2]. Humans identify faces and letters, recognize the surrounding area, and move to another place with the help of their eyes [3,4]. Furthermore, eyes play a significant role even in life-threatening moments, such as avoiding risk factors like falls and traffic accidents [5]. At the same time, the eyes are the fastest-aging organ, and older adults, who are vulnerable to eye diseases, such as cataracts and glaucoma, are highly likely to experience a decline in quality of life [6]. In the United States, the economic burden of eye diseases is expected to amount to approximately US$ 717 billion by 2050 due to increasing medical costs of the aging population [7]. Also, eye health affects work productivity greatly, and this is not just a problem limited to countries where vision and eye health care services are vulnerable due to a lack of medical facilities or welfare. The decline in productivity due to eye health and vision loss is known to be more evident in high-income countries [8]. In addition, research led by the Lancet Global Health Commission on Global Eye Health and the International Center for Eye Health demonstrates that improving eye health can contribute to achieving the UN Sustainable Development Goals, such as reducing poverty and improving work productivity, health, and equity [9]. As such, the

* Corresponding author. Yonsei University 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea.
E-mail addresses: danna23@yonsei.ac.kr (D. Jung), ajb2577@yonsei.ac.kr (J. An), hong7@yonsei.ac.kr (T. Hong).

https://doi.org/10.1016/j.jobe.2024.111146
Received 13 May 2024; Received in revised form 3 October 2024; Accepted 20 October 2024
Available online 22 October 2024
2352-7102/

importance and awareness of eye health tend to increase continuously.

One of the various factors that negatively impact eye health is an inappropriate indoor lighting environment [10]. From an architectural perspective, efforts have been made to create lighting environment that promote and maintain occupants' eye health. Firstly, countries and governments had suggested office illuminance standards to prevent excessive stimulus to eyes and ensure visibility [11]. However, the standards are mostly based on the results of previous experiments conducted on work involving paper documents. This means the standards are not suitable for occupants in modern offices that perform work mostly on visual display terminals like computers and smartphones [12]. Visual display terminals evoke visual display terminal syndrome, which is accompanied by eye fatigue and blurred vision, and thus, have emerged as a social issue [13]. They also work as an obstacle for exploring appropriate lighting environment for eye health since the brightness of the digital displays can be controlled by individual occupants. Thus, the stimulus of lighting to the eyes can vary for occupants that are under the same lighting environment. To create optimal lighting environment and enhance occupants' eye health, individual occupants' visual fatigue should be measured and considered in future smart buildings [14].

Many studies have been actively conducted regarding the development of models that predict occupants' subjective evaluation of indoor thermal environments, such as thermal comfort [16,17] as these models can contribute to control indoor environment intelligently [18]. However, there is still a lack of research on visual fatigue and even smaller number of studies have tried to develop visual fatigue prediction models. Firstly, there were two kinds of methods to measure or evaluate visual fatigue: (i) investigation of the physiological responses to measure objective visual fatigue, and (ii) analyzing survey responses to evaluate subjective visual fatigue. Dang et al. (2023) measured visual fatigue in various lighting environments through critical flicker-fusion frequency and subjective survey [15]. Liu et al. (2021) measured physiological responses, such as pupil diameter, fixation, and eye blinks, using eye-tracking glasses to measure visual fatigue [19] and evaluated various visual difficulties through a survey. However, most studies have been conducted only to the extent of the analyzing the relationship between lighting environments and visual fatigue. In this regard, only few models have been developed to predict subjective visual fatigue through physiological responses. Dang et al. (2023) developed a mathematical model to predict visual fatigue, but the variables only included environmental factors like directional luminance gradient, average luminance, and correlated color temperature (CCT) [15]. Similarly, most visual fatigue prediction models targeted visual fatigue caused by displays rather than lighting environments, and related studies did not set the physiological response of occupants or users as predictor variables but only focused on pixels or frames [20,21]. Some of the studies had developed visual comfort prediction models, but their interest was mostly on daylighting metrics without considering individual occupants' physiological responses [22,23]. One of the most relevant work for this study was the study conducted by Cen et al. (2019) who developed visual comfort prediction model using occupants' pupil data [24]. The median predictive accuracy recorded 0.65, but they discussed that limitations exist due to small number of subjects (i.e., 6 subjects).

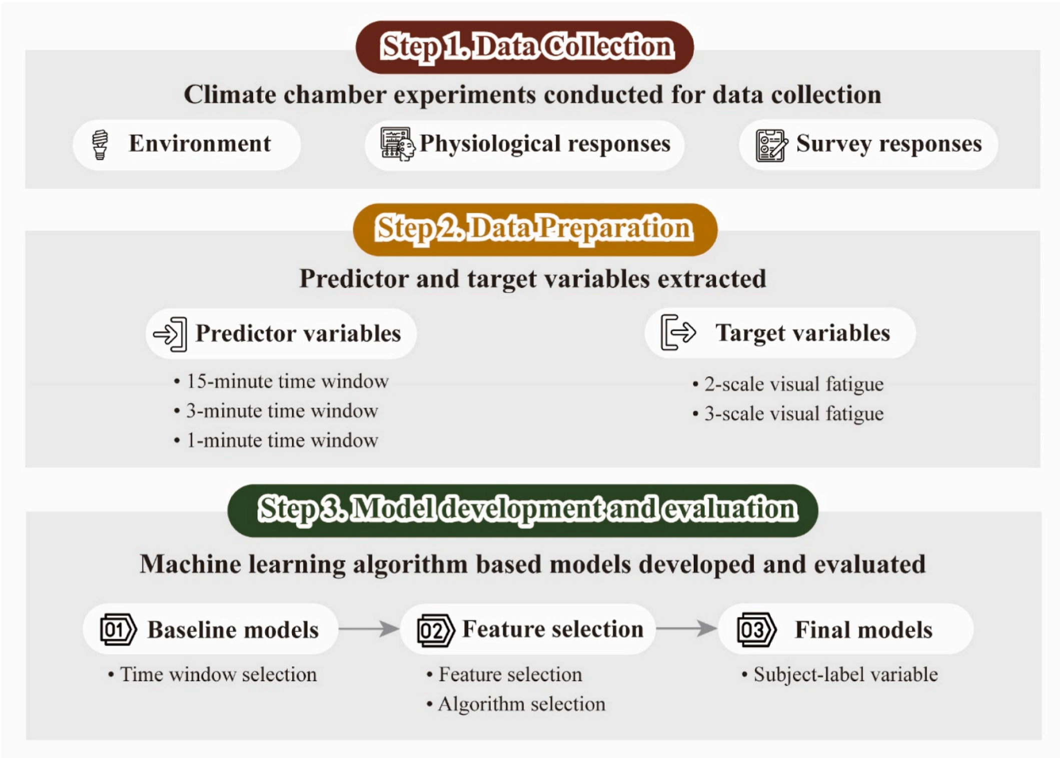

Fig. 1. Flowchart of visual fatigue prediction model development.

Despite emerging risks threatening occupants' eye health in modern offices, there is few studies that attempted to develop visual fatigue prediction model based on occupants' physiological responses and based on data-driven methods. To predict individual occupants' visual fatigue more accurately, a novel approach that considers occupants' physiological responses rather than environmental parameters has been applied in this study. For this purpose, three research questions have been stated as following.

(a) Can visual fatigue be effectively predicted using machine learning algorithms in combination with physiological responses?
(b) Which combination of datasets with different time windows and machine learning algorithms show the highest predictive performance?
(c) Which physiological responses can be used as predictor variables to predict occupants' visual fatigue?

To do this, data collection, such as physiological response and subjective visual fatigue, was performed via experiments in nine different lighting environments. Then, datasets with various predictor variables and target variables were created through the data preparation process. Four machine learning algorithms and various datasets were combined to develop prediction models, and the performance of the model was compared and improved through specific processes, such as adding feature selection and subject-label variables. To this end, 2-scale and 3-scale visual fatigue prediction models were proposed. A brief organization of the paper is as follows. The process of developing visual fatigue prediction model is illustrated in Section 2, and Section 3 presents the predictive performance of visual fatigue prediction model. Then, academic and practical implications of the study are discussed in Section 4. Lastly, Section 5 provides main findings and directions for future study.

2. Materials and methods

Fig. 1 shows the overall procedure of this study. To develop visual fatigue prediction models, this study (i) collected data, such as physiological responses and subjective visual fatigue via experiments in nine different lighting environments; (ii) preprocessed predictor variables and target variables; (iii) trained visual fatigue prediction model based on machine learning algorithms; and (iv) evaluated performance of prediction models. In this procedure, classification models developed by combining various datasets and algorithms were compared and evaluated to select a prediction model with the high predictive performance and high computational efficiency with a reduced number of predictor variables.

2.1. Data collection

In this study, physiological responses and survey data were collected through experiments with human subjects to develop models for predicting occupants' visual fatigue in the lighting environment of a general office.

2.1.1. Lighting environment and subject

The experimental environment was created so as not to deviate greatly from the lighting environment of a general office. However, to cause various visual fatigue levels, three levels of illuminance (i.e., 200lx, 500lx, and 800lx) and three types of CCT (i.e., 4,000K, 5,000K, and 6,500K) were combined to create nine types of lighting environments. In South Korea, recommendation for indoor illuminance range is presented as categories and there are three main categories for offices [14]. First category ranges from 150lx to 300lx, and it is recommended for offices have large digital screens such as meeting rooms. Second category ranges from 300lx to 600lx, and it is recommended for offices where occupants mainly work with computers. Third category ranges from 600lx to 1,500lx, and it is usually recommended for offices where designers work. So, the three levels of illuminance were selected from each of these categories to form environments similar to real offices. Meanwhile, there are no legal standards or recommendations for CCT of office lighting. However, previous studies have suggested various types of CCT for office lightings with different reasons [25–28]. Firstly, CCT ranging from 4,000K to 5,000K has been found out to be efficient and preferred office lightings by a number of previous studies [27,28], and thus, 4,000K and 5,000K were selected for this study. 6,500K was also selected as there were some studies that found out 6,500K of lighting can enhance cognitive performance [26,29], which means it is recommendable for office lightings. The experiment was conducted in an artificial climate chamber, and 32 lighting-emitting diode bulbs with illuminance and CCT control were installed uniformly on the ceiling to create nine types of lighting environments. In addition, through the artificial climate chamber, the thermal environment was maintained at 26 °C and relative humidity at 50 %.

The subjects of this experiment comprised 16 healthy adults (i.e., 8 males and 8 females). All the subjects were in their 20s or 30s, which is the period between adolescence and middle age when physical changes, including vision, are relatively stable. All subjects were undergraduate and graduate students who had no difficulty performing the computer-based cognitive test, and the average age was 24.38 $\pm$ 2.45 years. Before conducting the experiment, all subjects confirmed that they had no history of chronic diseases or disorders, such as color blindness or eye disease, high blood pressure, diabetes, stroke, or neurological disease. Subjects were asked to refrain from drinking 24 h before the experiment and sleep sufficiently. In particular, to prevent unnecessary brain arousal, caffeine intake was restricted on the day of the experiment. Lastly, as the mean clothing was set at 0.67clo for regular office workers, they were asked to wear long-sleeved shirts and long pants [30].

2.1.2. Measurement

As most of the previous studies have not suggested visual fatigue prediction model based on physiological data, there is lack of information about possible predictor for visual fatigue. Thus, various physiological responses were collected to figure out efficient

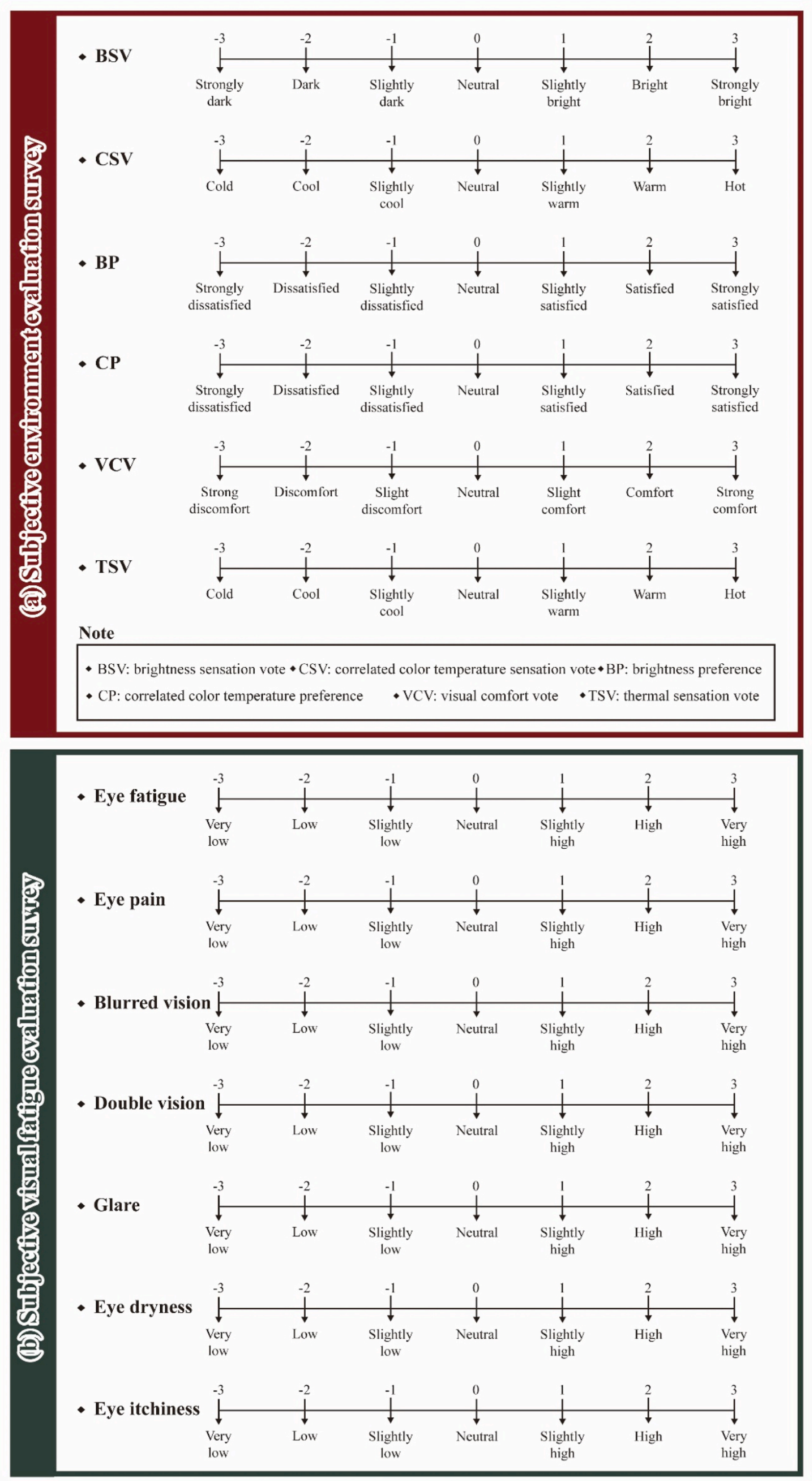

Fig. 2. Subjective environment and visual fatigue evaluation survey.

predictor for visual fatigue in this study. Four types of equipment were used to measure various physiological responses, and the measurable data for each equipment is as follows:

- ***Electroencephalography (EEG):*** EEG is a non-invasive method for measuring electrical activity in the brain [31]. The brain waves generated in the cerebral cortex were measured using a hat-shaped device with active electrodes placed on the scalp. The device is an EEG electrode hat manufactured by Biopac Systems in accordance with the international 10–20 system and can extract delta, theta, alpha, and beta waves in terms of frequency bands through the measured brain waves [32,33]. In this study, the brain waves were measured by attaching two electrodes to each of the four areas of the brain (i.e., frontal lobe, temporal lobe, parietal lobe, and occipital lobe). The sampling rate of data collection was 1000 Hz.
- ***Electrooculography (EOG):*** EOG is a non-invasive method for measuring the electrical activity of the eye with electrodes attached to the surface of the skin around the eye. As in the case of EEG, Biopac Systems' EOG measuring equipment was used, and the data was measured to remove artifacts from the EEG data [34,35]. The sampling rate of data collection was same as EEG.
- ***Electrocardiography (ECG):*** ECG is a non-invasive method for measuring the electrical activity of the heart with each beat. The mean heart rate (MHR), standard deviation of NN intervals (SDNN), and root mean square of successive differences (RMSSD) that are affected by autonomic nervous system can be extracted from raw ECG data [36]. The sampling rate of ECG data collection was also 1000 Hz.
- ***Eye tracking glasses:*** Tobii Glasses 3 manufactured by the Swedish company Tobii Technology was used as eye tracking glasses, which has a built-in camera that can collect eye gaze and eye movement data while capturing the wearer's viewing in any environment [37,38]. Data such as pupil diameter and eye validity were collected via the eye-tracking glasses. The sampling rate of Tobii Glasses 3 was 50 Hz.

In this experiment, methods to collect psychological responses were classified into two types: (i) subjective environment evaluation survey and (ii) subjective visual fatigue evaluation survey as shown in Fig. 2. These surveys were conducted and submitted by computers.

- ***Subjective environment evaluation survey:*** Subjective environment evaluation survey was conducted to confirm whether the experimental environment was created as intended at the planning stage. All six survey questions were designed on a seven-point Likert scale [39,40]. The specific questions are as follows: (i) brightness sensation vote (BSV), (ii) CCT sensation vote (CSV), (iii) brightness preference (BP), (iv) color temperature preference (CP), (v) visual comfort vote (VCV) and (vi) thermal sensation vote (TSV) to confirm that subjects conducted experiments in similar thermal environments [41].
- ***Subjective visual fatigue evaluation survey:*** This was conducted to measure the degrees of various eye-related symptoms perceived by the individual in a self-report format. Visual fatigue does not appear on single dimension, which means there are many types of visual fatigue. The survey has seven questions, which are measured on a seven-point Likert scale [19,42,43]. Specific questions to evaluate visual fatigue are as follows: (i) eye fatigue, (ii) eye pain, (iii) eye dryness, (iv) eye itchiness, (v) blurred vision, (vi) double vision, and (vii) glare.

Visual fatigue can be caused not only by the lighting environment but also by the visual tasks involved in work. Therefore, in this experiment, various cognitive tests were performed by subjects to form working situations in an office. In addition, the Freiburg visual acuity test (FrACT) was conducted in each experiment to secure visibility, the main purpose of lighting, and to confirm the subject's ability to identify letters on the chart.

- ***Primary cognitive test:*** The Diagnostic and Statistical Manual of Mental Disorders published by the American Psychiatric Association defines six primary cognition domains in humans [44]. In this study, a simple 1-min test was carried out for each cognition domain (i.e., attention, executive ability, memory, language, perceptual-motor ability, and social cognition).
- ***Complex cognitive test:*** The subjects were asked to complete complex cognitive tests that require various cognitive abilities in performing tasks in an actual office. First, the Alternative Uses Task was selected to quantitatively measure creativity [45–47]. Everyday objects, such as paper clips or bricks, were presented to the subjects, who had 3 min to come up with various ways to use the object. Second, a number pattern test was used to measure logical ability [48]. A series of numbers with certain rules were presented to the subjects, who were asked to identify the relationship between the given numbers and predict what the next number would be. Lastly, a comprehension test was performed to measure reading ability exerted mainly in a general office environment [49]. Passages with a similar degree of difficulty were extracted from open competitive examinations for the 9th-grade civil service in South Korea, and the subjects were asked to read three passages per experiment and solve multiple-choice questions.
- ***FrACT:*** FrACT is a computer-based test used to assess visual acuity and contrast sensitivity [50–52]. Among them, contrast sensitivity is the ability to identify minute differences in light and darkness and is important in detecting objects or letters. The tumbling E chart was used to measure contrast sensitivity. The letter 'E' with a gradual decrement of the contract appeared on the screen, rotated in a specific direction, up, down, left, or right. The subjects were instructed to identify E and press the direction key according to the rotated direction.

2.1.3. Experiment procedure

One subject participated in the experiment each time, and the order of the environment types was randomly assigned for each subject to minimize the influence that the previous lighting environment may have on the experiment. The subject's eligibility for participation in the experiment was confirmed first, and the method of the experiment and cognitive test were then explained to the subjects, who wore equipment to measure physiological responses after being fully aware of the rules for the cognitive test and making practices to obtain similar scores. Then, the following experiment sequence was repeated nine times, once for each lighting environment as Fig. 3 shows. First, the subjects closed their eyes and then waited to avoid unnecessary stimulation despite changes in the lighting environment. In the meantime, the experimenter created a lighting environment using smart LED lights. Second, the subjects opened their eyes as instructed and had time to adapt to the lighting environment while responding to the subjective environment evaluation survey. Third, the FrACT was performed to check the subjects' visual perception. After performing the test, data was measured using the equipment. Fourth, subjects conducted primary cognitive tests for 6 min. Fifth, the subjects completed the complex cognitive tests for 9 min. Sixth, the subjects responded to the self-report survey on visual fatigue after performing all tests. Lastly, they turned their heads away from the monitor and looked at a distant object to rest after completing all surveys and tests. Afterward, they closed their eyes and took a rest. After about 10 min of rest, the same process was repeated in the next lighting environment. To prevent lighting exposure history affecting the test results, subjects were instructed to close their eyes and the lights were turned off before the experiments. Each experiment took about 20 min, and all experiments were performed for 4 h. After a thorough safety and ethics assessment, the institutional review board approved this experiment, ensuring all procedures were conducted in compliance with the established and approved guidelines (IRB no. 7001988-202311-HR-2089-02).

2.2. Data preparation

Data processing was done to refine the raw data collected and transform it into predictor variables and target variables. To this end, three datasets of predictor variables and two sets of target variables containing various physiological responses were derived.

2.2.1. Predictor variables

To effectively train raw data of physiological responses collected via EEG, EOG, ECG, and eye-tracking glasses for machine learning, signal preprocessing, data cleaning, and scaling procedures were carried out.

First, EEG data was preprocessed using the AcqKnowledge software provided by Biopac Systems and the scipy.signal library in Python. EEG measures electrical activity in the cerebral cortex, and raw brain wave data includes various artifacts, such as eye movements, muscle movements, and heart rate [53]. A comb band-pass filter was used to remove signals of unnecessary frequencies included in brain waves, such as artifacts and noises, and artifacts caused by eye movements were eliminated using the measured EOG data. The EEG data was adjusted to 1–30 Hz, the range of the main four frequency bands (i.e., delta wave, theta wave, alpha wave and beta wave) related to brain activities. In addition, EEG in the time domain was converted into four frequency bands (i.e., delta wave, theta wave, alpha wave, beta wave) in the frontal lobe, temporal lobe, parietal lobe, and occipital lobe, respectively, through the power spectral density (PSD). Consequently, FR_δ, FR_θ, FR_α, and FR_β were derived as the relative strengths of the delta, theta, alpha, and beta waves in the frontal lobe. In the same manner, TE_δ, TE_θ, TE_α, and TE_β were derived as the relative strengths of each frequency band in the temporal lobe, PA_δ, PA_θ, PA_α, and PA_β as those in the parietal lobe, and OC_δ, OC_θ, OC_α, and OC_β as those in the occipital lobe. ECG is used to measure the electrical activity of the heart, and ECG signals are preprocessed using AcqKnowledge. The comb band-pass filter was used to remove outliers, adjusted to 0.5–100 Hz, and MHR, SDNN, and RMSSD were extracted using the Python hearty library based on the preprocessed data. Lastly, data collected via eye-tracking glasses was preprocessed using Tobii's 'Tobii Pro Lab software. This software is used to convert the pupil captured by the infrared camera embedded in the eye-tracking glasses into the pupil diameter

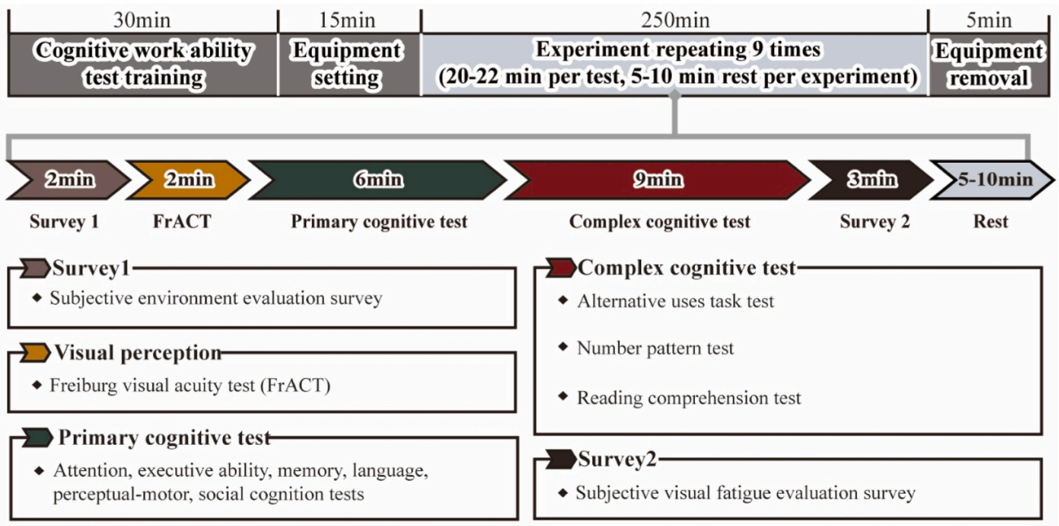

Fig. 3. Experiment procedure.

through an image processing algorithm. According to previous literature that demonstrates that raw pupil diameter is inappropriate as a predictor of visual fatigue [54], the pupillary unrest index (PUI), known as a potential visual fatigue predictor, was derived [19,55]. PUI is an index used to evaluate the eye's pupil response and confirm changes in the pupil diameter. In addition, blink amplitude (BA) was calculated using eye validity data [19]. BA is the ratio of time for eye blinks per unit of time.

Second, data cleaning and data segmentation were performed. To remove outliers due to sudden movements of the subject or problems with electrode adhesion during the experiment, the interquartile range was applied to detect outliers and replace them with mean values. After data refinement, three time windows were applied to configure data segments. The time window was set to 15-min units in data collection and smaller units of 3-min and 1-min intervals. In the case of datasets applied with 3-min and 1-min time windows, a predictor variable called 'Time' indicating order was added to reflect the time series. In addition, Z-score standardization and min-max normalization were done to adjust the scale between predictor variables for each dataset with three-time windows applied. For some machine learning algorithms, normalization is an essential step because the scale of predictor variables can significantly impact the model's performance. Through normalization, the data was adjusted to a continuous range of values between 0 and 1.

Consequently, 21 predictor variables, and 'CCT,' 'LX,' and 'Time' predictor variables representing environmental factors (i.e., CCT and illuminance) and time series were derived via physiological responses. The 15-min time window dataset consisted of 23 predictor variables, while the 3-min and 1-min time window datasets comprised 24 predictor variables, including 'Time.' Separately, subject-label predictor variables representing the basic information of each occupant were created. While the predictor variables described earlier are features that vary with time, subject-label predictors are features that do not vary, such as the gender or eyewear of occupants. The subject-label predictors are 'ID' and 'Gender' that distinguish each occupant, 'Eyewear, and contrast sensitivity 'CS,' which is a result of FrACT that represents the occupant's visual ability. 'Eyewear' consists of '0,' which indicates wearing glasses, '0.5,' which indicates wearing contact lenses, and '1,' which indicates not wearing eyewear. 'CS' is configured by calculating the mean value of contrast sensitivity, the subject's FrACT result, and classifying it into 'low,' 'medium,' and 'high.' Table 1 summarizes the predictor variables derived from this study.

2.2.2. *Target variable*

The target variable of the visual fatigue prediction model is the subjective visual fatigue level, which was collected via the subjective visual fatigue evaluation survey during the experiment. The subjective visual fatigue evaluation survey questions include symptoms and problems with the eyes and uncomfortable phenomena related to visual activity, such as (i) eye fatigue, (ii) eye pain, (iii) eye dryness, (iv) eye itchiness, (v) blurred vision, (vi) double vision and (vii) glare. All seven questions in the questionnaires were designed to be answered using a seven-point Likert scale, where lower scores indicate lower levels of expression, and higher scores more severe levels of expression. Meanwhile, principal component analysis (PCA) was used to derive one target variable, visual fatigue, from seven items. PCA can reflect the degree of impact of each item on visual fatigue by maximizing the variability of the data rather than averaging the responses of each item and deriving them as variables. In addition, because the subject's sensitivity to changes in each item is different, a single visual fatigue level can be derived considering these factors. As a result of PCA, seven principal components were extracted, and the cumulative explained variance ratio was 76.62 %. The first principal component that best represents the volatility of the data was selected, and the loading of each item was as used as a weight (refer to Eq. (1)) [56–58]. Table 2 shows the component matrix extracted as a result of PCA.

$$Integrated\ index = \sum_{i=1}^{k} L_1 \bullet X_1 + L_2 \bullet X_2 + \cdots + L_k \bullet X_k \tag{1}$$

Where, L stands for the loading of principal component; X stands for survey response value.

The visual fatigue prediction model to be developed in this study is a classification model that produces results such as '0' and '1'. The reason for using classification models instead of a regression is that classification models produce results in classes which makes it possible for occupants to intuitively understand their visual fatigue level. In this manner, Deng et al. (2021) also developed classification model and one of the scale they used was 3-scale [59]. In addition to this, 2-scale was also proposed in this study as it can indicate the presence of visual fatigue. Consequently, two types of scales were proposed: (i) 2-scale and (ii) 3-scale. Firstly, a 2-scale of 1 and 0 indicates the presence or absence of visual fatigue. Secondly, a 3-scale of −1, 0, and 1 indicates the degrees of visual fatigue, respectively, (i) visual fatigue does not occur, (ii) relatively mild visual fatigue occurs, and (iii) relatively severe visual fatigue occurs. Visual fatigue calculated via PCA was divided into both 2-scale and 3-scale.

Table 1
Summary of predictor variables.

Classification		Predictor variables
Time variant	Physiological response	'FR_δ', 'FR_θ', 'FR_α', 'FR_β', 'TE_δ', 'TE_θ', 'TE_α', 'TE_β', 'PA_δ', 'PA_θ', 'PA_α', 'PA_β', 'OC_δ', 'OC_θ', 'OC_α', 'OC_β', 'MHR', 'SDNN', 'RMSSD', 'PUI', 'BA'
	Environment	'CCT', 'LX'
	Time series	'Time'
Time invariant	Subject-label	'ID', 'Gender', 'Eyewear', 'CS'

Table 2
Principal component matrix.

Principal component	Eye fatigue	Eye pain	Eye dryness	Eye itchiness	Blurred vision	Double vision	Glare
1	0.707	0.729	0.365	0.446	0.774	0.793	0.646
2	−0.603	−0.433	−0.136	0.602	0.052	0.238	0.456
3	0.170	0.052	0.865	0.349	−0.346	−0.323	0.076

2.3. Model development and evaluation

2.3.1. Machine learning algorithms

When compared to traditional statistical methods, machine learning can more efficiently process and analyze large amounts of data and is useful for solving complex and multidimensional problems as it identifies patterns or statistical structures [60,61]. To develop a visual fatigue-predicting classification model, the following four supervised learning algorithms were used. They are mainly used to solve classification problems by classifying new data into one of several predefined classes and making predictions by learning data patterns.

- ***Artificial neural network (ANN)***:It is an algorithm similar to the human brain's neural network structure and consists of multiple layers and neurons [62,63]. Each neuron processes input data through weights and displays the results via an activation function. In this study, hyperparameters such as 'learning rates,' 'the number of layers,' and 'the number of neurons' were determined in a grid search for performance optimization.
- ***Support vector machine (SVM)***:It finds the optimal decision boundary that maximizes the margin between classes to classify data. It can be used for addressing linear and nonlinear classification problems and effectively processes complex datasets through the tunnel trick [64,65]. In the grid search performed in this study, 'C,' 'kernels,' 'gammas,' and 'degrees' were among the hyperparameters.
- ***Gradient boosting machine (GBM)***:It is an ensemble method that improves performance while sequentially learning weak prediction learners, such as decision trees [66,67]. At each step, the model compensates for the errors of the previous model to improve the performance. In this study, hyperparameters such as several estimators,' 'maximum depth,' and 'subsamples' were determined using the grid search.
- ***Random forest (RF)***:RF is an ensemble learning method that combines multiple decision trees to create a powerful model [68,69]. Each tree is trained using a random subset of the dataset, and the final decision is made by combining the predictions of multiple trees. RF prevents overfitting with high accuracy level and can be applied to both classification and regression problems [70]. In this study, hyperparameters such as 'number of estimators,' 'maximum depth,' 'minimum sample splits,' and 'maximum feature options' were determined using the grid search.

2.3.2. Feature selection and model development

A total of 24 possible predictor variables were derived through data preparation. This number is larger compared to the research that developed thermal comfort prediction models. In this study, a prediction model was developed through the following steps to maintain or increase predictive performance above a certain level while reducing the number of predictor variables to improve computational efficiency.

Firstly, a prediction model was developed using all 24 time-variant predictor variables. Because of the scant research related to physiological responses based visual fatigue prediction model, information on the predictor of visual fatigue is insufficient. Accordingly, three predictor variable datasets (i.e., 15-min, 3-min, and 1-min) and two target variable datasets (i.e., 2-scale and 3-scale) according to time window, and four machine learning algorithms were combined to develop visual fatigue prediction models, and their performance was compared and evaluated. Then, the time window of the most advantageous predictor variable was selected in each of the 2-scale models and 3-scale prediction models.

Secondly, a prediction model with the number of predictor variables reduced to more than 10 through feature selection was developed to increase the computational efficiency of the prediction model. Based on the results confirmed in the previous step, feature selection was performed on predictor variables datasets applied with the most advantageous time window for 2-scale and 3-scale, respectively. Also, for ANN and SVM that are not a decision tree-based algorithm, predictor variables were filtered based on Pearson correlation analysis. GBM and RF, decision tree-based algorithms, mainly selected predictor variables with high feature importance using recursive feature elimination (RFE) among wrapper methods. The objective of feature selection was to reduce the number of predictor variables and at the same time to maintain average performance above a certain level. Consequently, the most appropriate predictor variables and algorithm were selected via a comparative evaluation between the predictor variables through feature selection at 2-scale and 3-scale, respectively, and the models developed with the corresponding algorithm.

Lastly, the prediction performance of the model was improved by adding subject-label variables to the best prediction model among the models developed through feature selection, when sequential feature selection (SFS) was used to compare and evaluate model performance when four subject-label variables were added respectively. Unlike time-variant variables, the subject-label variable was collected at a single point in time, so its computational efficiency was low. Ultimately, the most appropriate predictor variables and algorithms were proposed as 2-scale and 3-scale visual fatigue prediction models.

2.3.3. Model training and grid search

The model training process, shown as Fig. 4, described earlier is as follows. The datasets consisting of predictor variables and target variables were split into training datasets used to train the model and test sets used to evaluate the performance. According to the previous literature, because the optimal split ratio of the training set and test set is the $\sqrt{p}$:1 (p: number of predictor variables) [71], the ratio is equal to 8:2 or 7:3, depending on the number of predictor variables in each model. K-fold cross-validation and grid search were performed on the training dataset to train the model and explore hyperparameter combinations to derive the optimal performance. K-fold cross-validation is a statistical method used to evaluate the performance of a model and validate its generalization ability. The dataset is divided into K number of subsets, one of which is used as validation data, and the remaining K-1 dataset is used as training data. The process is repeated K number of times until each data is used once as validation data. This can prevent overfitting, evaluate generalization performance, and most importantly, reduce bias in performance evaluation via random partitioning. It is common to select 10-fold cross-validation when the number of data is relatively small (hundreds to thousands), and five-fold cross-validation when the number of data is relatively large (thousands to tens of thousands). Therefore, 10-fold cross-validation was selected for cross-validation of the visual fatigue prediction model developed in this study. Grid search, one of the methods for optimizing the hyperparameters of a machine learning model, explores all possible combinations of hyperparameters and presents the optimal combination to maximize model performance. The performance indicator is accuracy, which is the ratio of the number of correct predictions to the total number of predictions and is the most intuitive performance indicator.

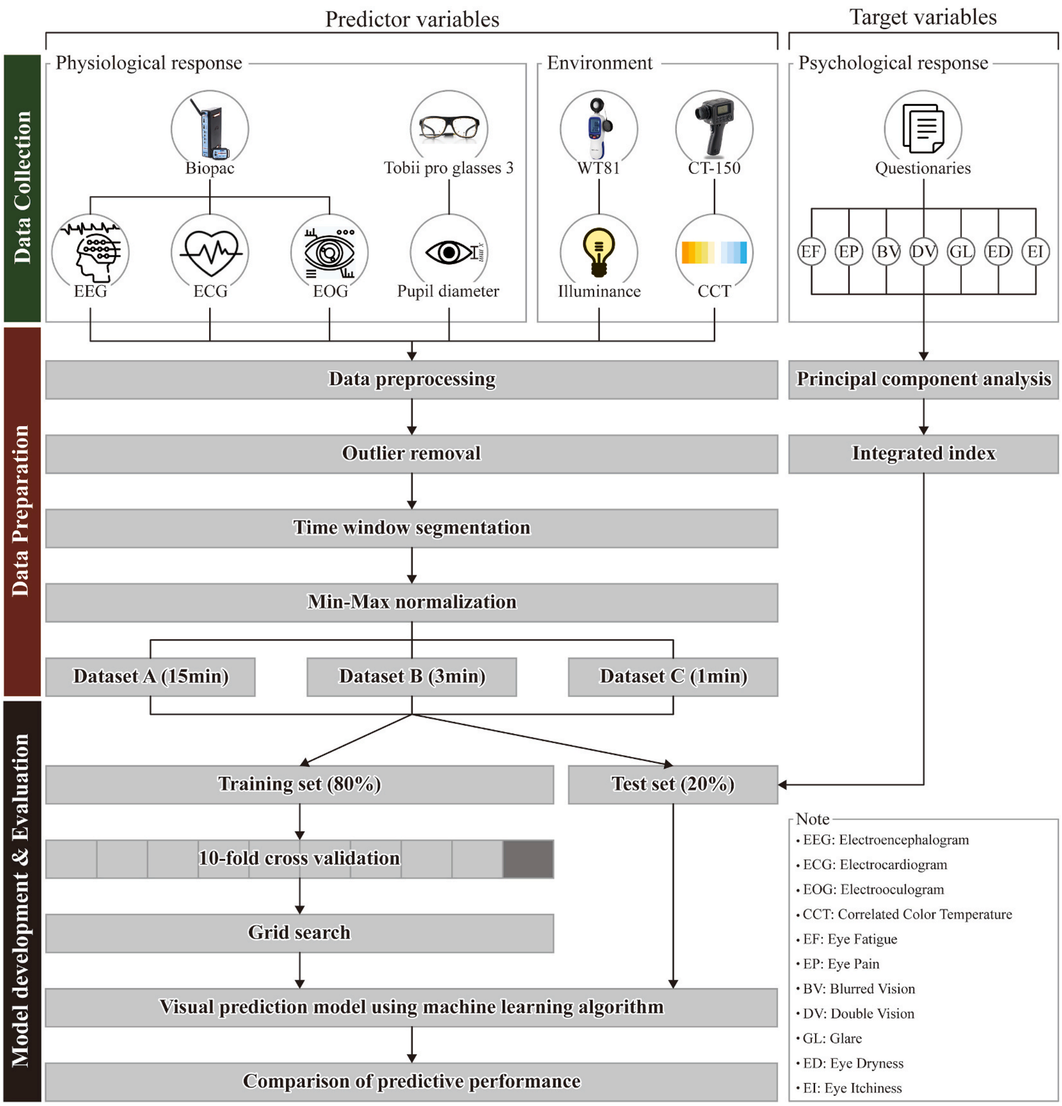

Fig. 4. Prediction model development.

2.3.4. Model evaluation

After training the visual fatigue prediction model and exploring the optimal hyperparameter combination, the visual fatigue prediction model was evaluated using various performance indicators. The evaluation was performed using test dataset that had not been used during model training. If only a single performance indicator is used for evaluating the predictive power of a machine learning model, it fails to consider various performance aspects of the model, leading to a biased result. Therefore, in this study, the model was evaluated using five performance indicators. These performance indices are derived from the confusion matrix and stem from four key factors to evaluate the performance of a classification model: (i) true positive is an instance in which the model correctly predicts a positive class, (ii) false positive is an instance in which the model incorrectly predicts a positive class when it should have been negative, (iii) false negative is an instance in which the model incorrectly predicts a negative class when it should have been positive, and (iv) true negative is an instance in which the model correctly predicts a negative class. The performance indices used to evaluate the model in this study are as follows.

- ***Accuracy:***It is the ratio of the number of correct predictions to the total number of predictions and is the most intuitive and basic performance indicator. However, it may show biased values when datasets are imbalanced. In this study, it is the ratio of correct predictions for the occurrence or visual fatigue level (refer to Eq. (2)).
- ***Precision:***It is the ratio of correct positive predictions to all positive predictions. This indicator is mainly used when it is important to minimize false positives. In the 2-scale model, it is the ratio of cases where visual fatigue was predicted to have occurred. In the 3-scale model, it is the ratio calculated for each class (refer to Eq. (3)).
- ***Recall:***It is the ratio of correctly predicted positive cases to the total number of actual positive cases. This indicator is mainly used when it is relevant to minimize false negatives (refer to Eq. (4)). In the case of the 2-scale of the visual fatigue prediction model, the value of this indicator is highlighted because the key issue is to detect the visual fatigue status.
- ***F1 score:***It is the harmonic mean of precision and recall (refer to Eq. (5)). This indicator is useful when evaluating whether or not performance is biased toward one class because it considers both indicators simultaneously.
- ***Area under the receiver operating characteristics curve(AUC-ROC)***:ROC is a curve that exhibits how well a binary classification model detects positive classes, with the ratio of true positives to false positives as the axis. AUC-ROC is the area under the ROC curve and is represented as a value between 0 and 1. The closer it is to 1, the better the performance of the model. In this study, it was represented as a percentage to unify the units with other indices.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{5}$$

Where, TP stands for the true positive; TN stands for the true negative; FP stands for the false positive; FN stands for the false negative.

Accuracy is a performance indicator that can be used equally in binary classification or multiple classification because it is the ratio of correct classifications to the total number of classifications. However, since other indicators are performance indicators for binary classification, the final index was calculated using the weighted mean, which is an average computed by calculating these indicators like binary classification for each class and applying weights according to the number of samples in each class to apply these performance indicators to the 3-scale in multiclass classification.

Table 3
Mean value and standard deviation of subjects' subjective environment evaluation.

Survey	Lighting environment								
	4,000K			5,000K			6,500K		
	200lx	500lx	800lx	200lx	500lx	800lx	200lx	500lx	800lx
BSV[a]	−1.63 (0.72)[g]	0.88 (1.45)	1.56 (1.41)	−0.94 (1.44)	0.50 (1.32)	1.88 (1.02)	−1.25 (0.86)	0.31 (1.20)	2.13 (0.72)
CSV[b]	1.13 (1.20)	1.13 (1.02)	1.56 (0.81)	−0.75 (0.93)	0.25 (0.86)	0.38 (1.36)	−1.00 (0.82)	−1.06 (0.93)	0.19 (1.42)
BP[c]	−0.50 (1.67)	1.06 (1.00)	0.56 (1.36)	−0.06 (1.34)	0.56 (1.26)	0.56 (1.67)	−0.56 (1.09)	0.75 (1.57)	1.69 (1.14)
CP[d]	−0.06 (1.57)	0.94 (1.24)	0.75 (1.29)	0.44 (1.55)	0.81 (0.98)	1.06 (1.34)	−0.31 (1.35)	0.19 (1.52)	1.25 (1.24)
VCV[e]	−0.56 (1.41)	1.19 (1.05)	0.56 (1.31)	0.38 (1.45)	0.75 (1.34)	0.75 (1.57)	0.06 (1.29)	0.63 (1.54)	1.25 (1.18)
TSV[f]	0.13 (0.72)	−0.13 (0.72)	0.25 (0.58)	−0.13 (0.34)	−0.19 (0.66)	−0.13 (0.62)	−0.44 (0.63)	−0.19 (0.54)	0.00 (0.52)

Note: *BSV[a]* stands for brightness sensation vote; *CSV[b]* stands for corelated color sensation vote; *BP[c]* stands for brightness preference; *CP[d]* stands for correlated color temperature sensation vote; *VCV[e]* stands for visual comfort vote; *TSV[f]* stands for thermal sensation vote; *-1.63(0.72)[g]* stands for mean value (standard deviation).

3. Results

3.1. Experimental results

3.1.1. Visual fatigue under different lightings

Nine types of lighting environments were formed in the experiments and the results of the subjective environment evaluation survey are shown in Table 3.

Fig. 5 specifies the results of a survey with seven questions on visual fatigue that occurs in nine different lighting environments. Among the seven questions, the questions that elicited the subject's most sensitive responses were eye fatigue, eye pain, and eye dryness. Also, it was found that most of the subjects responded insensitively to eye itchiness and double vision. Among the nine lighting

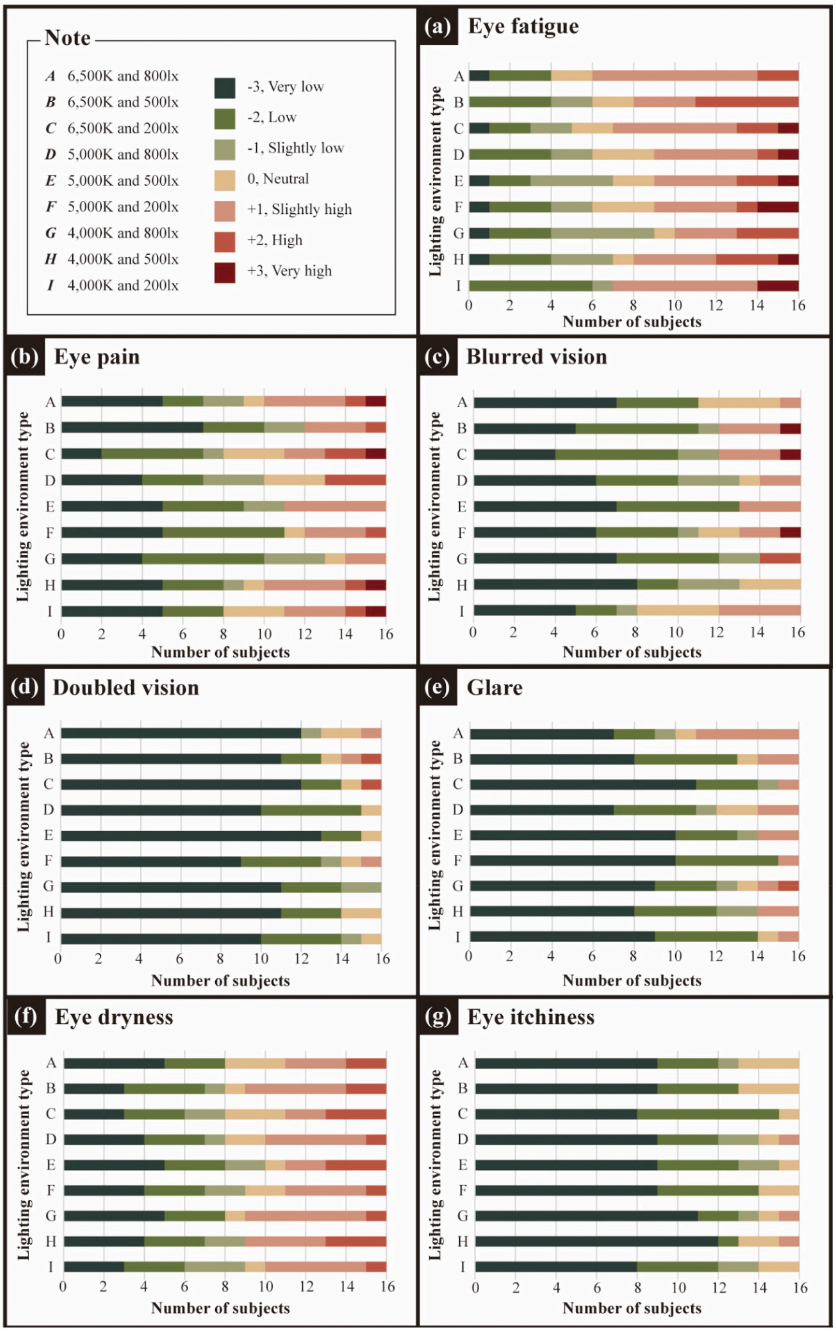

Fig. 5. Subjective visual fatigue survey.

environments, 4000 K and 6500 K mostly evoked visual fatigue, and visual fatigue tended to change depending on illuminance rather than CCT.

BSV shows higher values as illuminance increases in all three CCTs, indicating that subjects are aware of the brightness difference. When CSV is compared by illuminance level, 4000 K shows the highest value, whereas 6500 K shows the lowest value, which indicates that the subjects are aware of CCT appropriately. BP records a negative value at 200 lx, the lowest illuminance level in all CCTs, implying that most of the subjects do not prefer a darker environment. However, the degree of preference was different for each CCT in the 500 lx and 800 lx environments. The consistency of results was relatively lower in CP than in BP. First of all, the preference for 5000 K was the highest, and at 6500 K, a significant difference occurred in preference depending on illuminance. The environment with the highest VCV was the one with a combination of 6500 K and 800 lx, whereas the environment with the lowest VCV was the one with a combined 4000 K and 200 lx. In addition, the average TSV value was ± 0.5 in all environments. Therefore, it can be confirmed that all environments were created as neutral environments, where the temperature is neither cold nor hot for the subjects.

3.1.2. *Effect of lighting exposure history and cumulative impact of repeated experiments*

The order of the nine lighting environments were randomly assigned for each subject and a mixed effects model was used to analyze whether lighting exposure history, specifically the previous lighting conditions of each experiment, affected visual fatigue. The previous lighting condition was set as the fixed effect and the subjects were set as a random effect. Table 4 shows the results of the mixed effect model analysis. As a result, none of coefficients for the previous lighting conditions had *p*-values below 0.05, indicating that none of them significantly influenced visual fatigue in subsequent conditions.

As subjects repeated nine experiments on one day, there could be cumulative impact of visual fatigue over time. To verify that visual fatigue was not affected by cumulative impact, statistical approach was used. Shapiro-Wilk test was conducted to check the normality of the visual fatigue data over time. The test statistic was 0.929 and the *p*-value was 1.22×10^{-6} which indicates that the visual fatigue data over time does not follow the normal distribution. As the data does not meet the assumption of normality, the Friedman test was employed. The Friedman test yielded a statistic of approximately 9.92 with a *p*-value of about 0.271. This indicates that the visual fatigue experienced by subjects did not significantly vary with the experiment order and also that the time for resting in between was sufficient for the subjects to recover from visual fatigue. SM v. 1 and 2 show the brief results of cognition tests under different lightings.

3.2. *Performance evaluation of visual fatigue prediction model*

3.2.1. *Visual fatigue prediction model without feature selection*

Firstly, visual fatigue prediction models without a feature selection process were developed, where the optimal time window of the dataset was decided. Predictor variables were all applied without being filtered out, and there were three types of dataset depending on the time window (i.e., 15-min, 3-min, and 1-min). The number of predictor variables was 23 for the 15-min dataset and 24 for the 3-min and 1-min datasets. As a result, three types of predictor datasets were applied to four different algorithms to make 2-scale and 3-scale prediction models. Table 5 enumerates the results of the grid search of these models, and Table 6 reveals the performance of visual fatigue prediction models without feature selection.

In general, the performance of the 2-scale model was higher than that of the 3-scale model, and the average performance was about 7 % higher. Overall, multiclass classification models have a higher complexity level in setting decision boundaries between multiple categories than binary classification models; therefore, the higher the classification scale, the lower the accuracy level. However, the average performance of both models (i.e., 2-scale model and 3-scale model) was 85.85 % and 78.90 %, respectively, which was higher when compared to the performance of the existing thermal comfort predictive model or visual comfort predictive model. For both the 2-scale and 3-scale models, the performance difference between the 15-min time window and the 3-min time window was greater than the performance difference between the 3-min time window and the 1-min time window. In the case of the 2-scale model, the average performance was the highest at 92.42 % when the 3-min time window dataset was used. On the other hand, in the case of the 3-scale model, the average performance was the highest at 86.76 % when the 1-min time window dataset was used. Specifically, the 2-scale model showed the best performance at 95.5 % when the 3-min time window dataset and ANN were combined, while the 3-scale model showed the highest performance at 89.41 % when the 1-min time window dataset and GBM were combined. Fig. 6 highlights the average predictive performance by time window and algorithms and shows the overall predictive performance.

3.2.2. *Visual fatigue prediction model with feature selection*

The results of the evaluation of the visual fatigue prediction models without feature selection showed that the 2-scale model had the best performance with the 3-min time window dataset applied, while the 3-scale model had the best performance with the 1-min time window dataset applied. The average performance of both models was high at 92.42 % and 86.76 %, respectively. However, to

Table 4
Mixed effect model analysis to investigate effect of lighting exposure history.

Variable	Coefficient	Standard error	*p*-value
Intercept	0.489	0.088	0.000
Previous lighting	0.011	0.006	0.062
Group variance	0.106	0.227	

Table 5
Optimal combinations of hyperparameters derived by grid search.

Scale	Time window	Algorithms	Hyperparameter
2-scale	15-min	ANN[b]	'learning rate': 0.01, 'number of layers': 1, 'number of neurons': 64
		SVM[c]	'C': 10, 'kernel': rbf, 'gamma': scale, 'degree': 3
		GBM[d]	'number of estimators': 200, 'maximum depth': 5, 'subsample': 0.8
		RF[e]	'number of estimators': 100, 'maximum depth': 30, 'minimum samples split': 5, 'maximum features': sqrt
	3-min	ANN	'learning rate': 0.001, 'number of layers': 2, 'number of neurons': 128
		SVM	'C': 10, 'kernel': rbf, 'gamma': 0.1, 'degree': 3
		GBM	'number of estimators': 100, 'maximum depth': 7, 'subsample': 0.8
		RF	'number of estimators': 100, 'maximum depth': 30, 'minimum samples split': 5, 'maximum features': sqrt
	1-min	ANN	'learning rate': 0.01, 'number of layers': 2, 'number of neurons': 128
		SVM	'C': 10, 'kernel': rbf, 'gamma': 0.1, 'degree': 3
		GBM	'number of estimators': 300, 'maximum depth': 7, 'subsample': 0.8
		RF	'number of estimators': 400, 'maximum depth': 30, 'minimum samples split': 2, 'maximum features': sqrt
3-scale	15-min	ANN	'learning rate': 0.001, 'number of layers': 3, 'number of neurons': 128
		SVM	'C': 10, 'kernel': rbf, 'gamma': 0.1, 'degree': 3
		GBM	'number of estimators': 300, 'maximum depth': 3, 'subsample': 0.8
		RF	'number of estimators': 200, 'maximum depth': 30, 'minimum samples split': 2, 'maximum features': sqrt
	3-min	ANN	'learning rate': 0.01, 'number of layers': 2, 'number of 'neurons': 128
		SVM	'C': 10, 'kernel': rbf, 'gamma': 0.1, 'degree': 3
		GBM	'number of estimators': 300, 'maximum depth': 7, 'subsample': 0.8
		RF	'number of estimators': 100, 'maximum depth': 20, 'minimum samples split': 5, 'maximum features': sqrt
	1-min	ANN	'learning rate': 0.01, 'number of layers': 2, 'number of neurons': 128
		SVM	'C': 10, 'kernel': rbf, 'gamma': 0.1, 'degree': 3
		GBM	'number of estimators': 200, 'maximum depth': 7, 'subsample': 1.0
		RF	'number of estimators': 300, 'maximum depth': 20, 'minimum samples split': 5, 'maximum features': log2

Note: *ANN*[a] stands for artificial neural network; *SVM*[b] stands for support vector machine; *GBM*[c] stands for gradient boosting machine; *RF*[d] stands for random forest.

increase the computational efficiency of the model, an attempt was made to reduce the number of predictor variables through feature selection. Firstly, feature selection of the model using ANN and SVM which are not decision tree-based algorithms was performed via the Pearson correlation analysis. Features were selected based on the correlation analysis results with subjective visual fatigue, including only those that were significant at the 0.01 level as shown in Table 7. After that, when the number of variables was greater than the standard, those with the smallest absolute value of the Pearson correlation coefficient were excluded from among the variables with a correlation at 0.01 level. As a result, 11 out of 24 predictor variables remained in the 2-scale model, and specifically, the predictor variables were SDNN, RMSSD, BA, TE_δ, PA_θ, OC_θ, TE_α, PA_α, FR_β, PA_β, and OC_β. In the meantime, 13 out of 24 predictor variables remained in the 3-scale model. Specifically, the predictor variables were MHR, SDNN, RMSSD, BA, TE_δ, PA_δ, FR_θ, TE_θ, PA_θ, TE_α, PA_α, PA_β, and OC_β.

Secondly, feature selection of the model using GBM and RF, decision tree-based algorithms, was performed via RFE. Fig. 7 shows the feature importance of GBM and RF based models for 2-scale and 3-scale targets. Among the previously developed 2-scale models without feature selection, the feature importance of the model developed with GBM, RF, and a 3-min time window dataset was confirmed. Then, variables with lower rank in feature importance were removed, and 11 predictor variables were selected. Specifically, the predictor variables of the 2-scale model using GBM were PA_θ, OC_θ, OC_β, PA_α, FR_α, PA_β, MHR, FR_β, OC_δ, PUI, and TE_β. In addition, the predictor variables of the 2-scale model using RF were PA_θ, PA_α, TE_α, OC_β, OC_θ, BA, TE_β, PA_β, FR_α, FR_β, and RMSSD. Among the 3-scale models without feature importance, the feature importance of the model developed with GBM, RF, and 1-min time window dataset was confirmed, and 13 predictor variables were then selected by removing low-rank variables. To be specific, the predictor variables of the 3-scale model using GBM were FR_β, PA_θ, OC_β, PA_β, TE_β, SDNN, TE_θ, PA_α, OC_α, CCT, OC_δ, PA_δ, and LX. In addition, the predictor variables of the 3-scale model using RF were PA_θ, OC_β, FR_β, PA_β, TE_β, OC_α, TE_θ, PA_α, SDNN, TE_α, FR_α, OC_θ, and OC_δ.

Table 8 summarizes the performance of the visual fatigue prediction model developed via feature selection. The performance of the 2-scale model decreased by 5.87 % on average when the number of predictor variables was reduced from 24 to 11. The greatest decrease was apparent in the models using ANN and SVM rather than decision tree-based algorithms, where the performance decreased by 14.29 % and 7.11 %, respectively. However, models using GBM and RF, decision tree-based algorithms, showed similar or higher levels of performance despite the more than half reduction in the number of predictor variables. The average performance of the GBM-based 2-scale model without feature selection was 89.27 %, while that of the model with feature selection was 86.49 %, showing a 2.77 % decrease. In addition, the average performance of the RF-based 2-scale model was 89.44 % before feature selection, but it increased by 0.71 %–90.15 % after feature selection. Therefore, it can be concluded that the overall performance has significantly improved compared to the previous model as the computational efficiency has largely increased with a higher average through feature selection. Consequently, the 2-scale model with the best performance after feature selection was the one using RF (i.e., 90.15 %). The performance of the 3-scale model decreased by 6.87 % when the number of predictor variables was reduced from 24 to 13. Even in this case, when an algorithm other than decision tree-based algorithms (i.e., ANN and SVM) was used, the performance decreased by more

Table 6
Predictive performance of visual fatigue prediction models without feature selection.

Time window of predictor variables	Algorithms	2-scale prediction model						3-scale prediction model					
		Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC-ROC[a](%)	Average (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC-ROC (%)	Average (%)
15-min	ANN[b]	58.62	73.33	57.89	64.71	66.32	64.17	55.17	58.29	55.17	51.39	76.11	59.23
	SVM[c]	65.52	76.47	68.42	72.22	68.95	70.32	58.62	58.78	58.62	57.94	78.19	62.43
	GBM[d]	79.31	88.24	78.95	83.33	79.47	81.86	75.86	77.22	75.86	74.62	87.52	78.22
	RF[e]	79.31	88.24	78.95	83.33	77.37	81.44	75.86	78.15	75.86	75.70	83.17	77.75
	Average	70.69	81.57	71.05	75.89	73.03	74.45	66.37	68.11	66.38	64.91	81.25	69.41
3-min	ANN	94.44	90.79	98.57	94.52	99.19	95.5	77.08	77.42	77.08	76.98	92.96	80.30
	SVM	94.44	91.89	97.14	94.44	99.52	95.49	82.64	82.69	82.64	82.59	96.26	85.36
	GBM	89.58	91.04	87.14	89.05	89.52	89.27	73.61	73.77	73.61	73.34	91.35	77.14
	RF	87.50	83.33	92.86	87.84	95.68	89.44	76.39	77.59	76.39	76.01	90.12	79.30
	Average	91.49	89.26	93.93	91.46	95.98	92.42	77.43	77.87	77.43	77.23	92.67	80.53
1-min	ANN	89.81	91.40	88.99	90.18	96.43	91.36	83.33	83.44	83.33	83.36	95.49	85.79
	SVM	91.90	93.64	90.75	92.17	97.11	93.11	87.50	87.50	87.50	87.48	97.07	89.41
	GBM	90.05	90.71	90.31	90.51	90.03	90.32	87.50	87.48	87.50	87.47	97.25	89.44
	RF	86.11	87.44	85.90	86.67	93.54	87.93	79.40	79.35	79.40	79.34	94.60	82.42
	Average	89.47	90.80	88.99	89.88	94.28	90.68	84.43	84.44	84.43	84.41	96.10	86.76
Average		83.88	87.21	84.66	85.75	87.76	85.85	76.08	76.81	76.08	75.52	90.01	78.90

Note: *AUC-ROC*[a] stands for the area under the receiver operating characteristics curve; *ANN*[b] stands for artificial neural network; *SVM*[c] stands for support vector machine; *GBM*[d] stands for gradient boosting machine; *RF*[e] stands for random forest; Shaded areas represent average values.

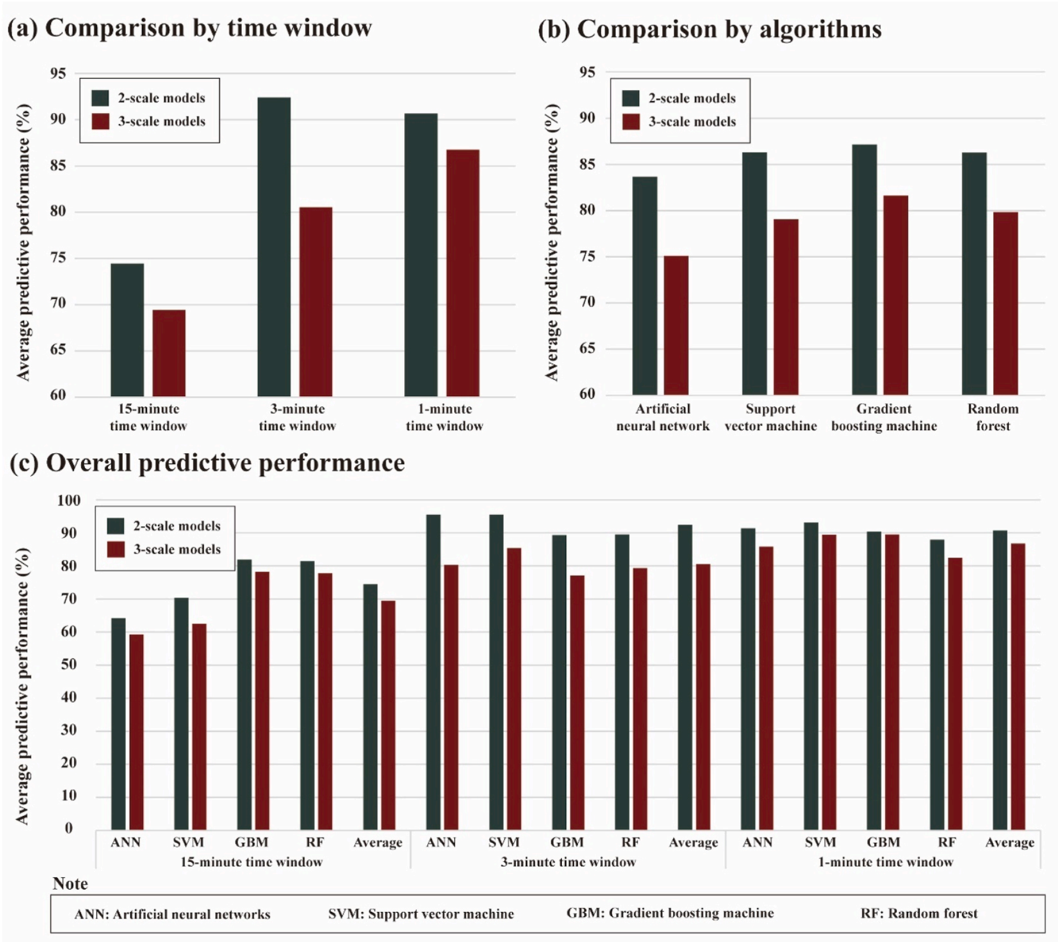

Fig. 6. Average predictive performance of models without feature selection.

than 10 %, showing a large difference in performance. In the case of decision tree-based algorithms (i.e., GBM and RF), the performance decreased by less than 3 %, showing a relatively small difference. Ultimately, the 3-scale model with the best performance after feature selection was the one using GBM, and the average performance was 88.47 %, showing a 0.97 decrease. Fig. 8 highlights the predictive performance of the 2-scale and 3-scale models without and with feature selections.

3.2.3. Subject-label visual fatigue prediction model

Among the models that reduced the number of predictor variables through feature selection, the 2-scale model and 3-scale model with the best performance showed average performance levels of 90.15 % and 88.47 %, respectively. Meanwhile, an attempt was made to investigate whether (i) prediction performance improves and (ii) which subject-label variable is most effective at improving performance when the subject-label was added based on these two models. The performance was compared by adding 'ID,' 'Gender,' 'Eyewear,' and 'CS' once to the dataset of the 2-scale model based on RF and the 3-scale model based on GBM. Also, by applying the SFS method, the performance was compared by adding two subject-label variables with the model performance that improved the most. As a result, 'ID' and 'CS' contributed the most to performance improvement in both the 2-scale model and 3-scale model, as depicted in Table 9. In the 2-scale model, the average performance improved the most by 3.57 % when CS was added, followed by 2.08 % when ID was added. When gender and eyewear were added, there was a 2 % improvement. When CS and ID were added at the same time, the average performance was 92.33 %, which was lower than when only CS was added. On the other hand, in the 3-scale model, the average performance improved the most by 4.6 % when ID was added, followed by 3.89 % when CS was added. The degree of performance improvement due to the subject-label variable was greater in the 3-scale model than in the 2-scale model. In addition, when both ID and CS were added, the average performance was 94.64 %, which was higher than that of the 2-scale model. Fig. 9 shows the average performance of the 2-scale and 3-scale models with different subject-label variables.

Accordingly, the final performance comparison of the 2-scale model and 3-scale model is as follows (refer to Table 10): The 2-scale model is a dataset applied with a 3-min time window, and the predictor variables were RMSSD, BA, PA_θ, OC_θ, FR_α, TE_α, PA_α, FR_β, TE_β, PA_β, and OC_β. In addition, when 'CS' was added as a subject-label variable, the total number of predictor variables was 12, and the algorithm was RF. The 3-scale model is a dataset applied with a 1-min time window, and the predictor variables were CCT, LX, SDNN,

Table 7
Results of Pearson correlation analysis.

Features	2-scale			3-scale		
	Pearson correlation coefficient	*p*-value	Selected	Pearson correlation coefficient	*p*-value	Selected
Time	0.000	1.000		0.000	1.000	
CCT	0.054	0.147		0.011	0.610	
LX	0.034	0.362		0.021	0.336	
MHR	−0.056	0.134		−0.131	0.000[b]	O
SDNN	−0.183	0.000[b]	O	−0.216	0.000[b]	O
RMSSD	−0.193	0.000[b]	O	−0.111	0.000[b]	O
PUI	−0.073	0.049[a]		0.006	0.789	
BA	−0.207	0.000[b]	O	−0.129	0.000[b]	O
FR_δ	−0.032	0.395		−0.057	0.008[b]	
TE_δ	−0.119	0.001[b]	O	−0.113	0.000[b]	O
PA_δ	−0.006	0.872		0.073	0.001[b]	O
OC_δ	0.016	0.668		−0.033	0.125	
FR_θ	−0.092	0.013[a]		−0.110	0.000[b]	O
TE_θ	0.044	0.233		−0.074	0.001[b]	O
PA_θ	−0.272	0.000[b]	O	−0.0309	0.000[b]	O
OC_θ	0.181	0.000[b]	O	0.000	0.986	
FR_α	−0.005	0.903		0.037	0.087	
TE_α	−0.262	0.000[b]	O	−0.087	0.000[b]	O
PA_α	−0.258	0.000[b]	O	−0.075	0.001[b]	O
OC_α	−0.031	0.399		−0.012	0.566	
FR_β	−0.148	0.000[b]	O	−0.006	0.772	
TE_β	−0.015	0.678		−0.069	0.001[b]	
PA_β	−0.177	0.000[b]	O	−0.200	0.000[b]	O
OC_β	−0.166	0.000[b]	O	−0.205	0.000[b]	O

Note.
[a] Stands for coefficient significant at 0.05 level.
[b] Stands for coefficient significant at 0.01 level.

PA_δ, OC_δ, TE_θ, PA_θ, PA_α, OC_α, FR_β, TE_β, PA_β, and OC_β. When 'CS' and 'ID' were added as subject-label variables, the total number of predictor variables was 15, and the algorithm was GBM. The final average performance rates of the 2-scale and 3-scale visual fatigue models were 93.73 % and 94.64 %, respectively.

4. Discussion

4.1. Proposal of visual fatigue prediction model

This study developed classification models that predict visual fatigue based on the occupants' physiological responses. In the process of addressing this objective, the following answers were presented to the three research questions outlined in the introduction.

- *Can visual fatigue be effectively predicted using machine learning algorithms in combination with physiological responses?*

While models for predicting occupants' subjective thermal comfort through physiological responses have been developed by various studies, research on models for predicting visual fatigue felt by occupants is still in their early stages. A few studies have attempted to predict visual fatigue, but their models mostly considered environmental parameters and investigated methods such as mathematical model development rather than data-driven approaches [15,20]. Therefore, this study sought to develop a visual fatigue prediction model based on various physiological responses and machine learning algorithms. Consequently, two visual fatigue models were proposed in this study; (i) a 2-scale model based on RF with 12 predictor variables at a 3-min time window and (ii) a 3-scale model based on GBM with 15 predictor variables at a 1-min time window. The accuracy of the proposed 2-scale model was 92.59 % and the 3-scale model was 93.52 %. The average performance rates exhibited, considering all performance indices, were 93.73 % and 94.64 %, respectively. The results highlight the possibility of predicting visual fatigue in data-driven method using occupants' physiological responses. The performance of prediction models developed in this study can be evaluated as reasonably high compared to some of the recent thermal comfort prediction models with similar approaches (i.e., approximately 90 % of average performance [72] and 89.2 % of average accuracy [73]). Additionally, the previous model that used pupil data to predict visual fatigue has shown median predictive accuracy of 0.65 [24] highlighting the potential of physiological responses in predicting visual fatigue. Moreover, this also implies that findings of this study demonstrate an improvement in performance compared to the previous model.

- *Which combination of datasets with different time windows and machine learning algorithms show the highest predictive performance?*

To develop a visual fatigue prediction model, models were developed using all predictor variables without feature selection, and the most appropriate time windows for 2-scale and 3-scale were then selected. Smaller time window may provide more granular

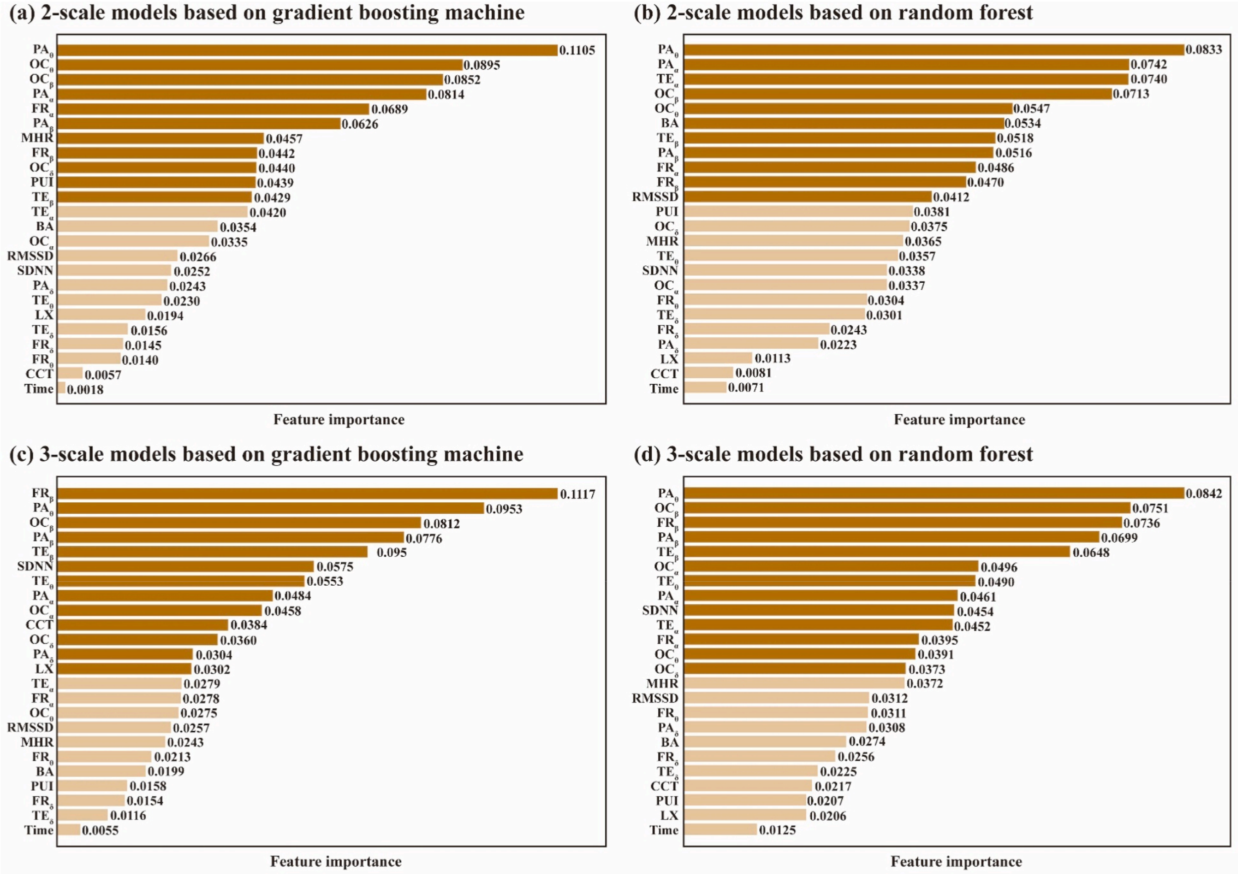

Fig. 7. Feature importance of the 2-scale and 3-scale models.

Table 8
Comparison of predictive performance of visual prediction model without feature selection and with feature selection.

Scale	Algorithm	Number of predictors	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC-ROC[a](%)	Average (%)
2-scale	ANN[b]	24	94.44	90.79	98.57	94.52	99.19	95.50
		11	78.7	87.23	70.69	78.1	91.36	81.22
	SVM[c]	24	94.44	91.89	97.14	94.44	99.52	95.49
		11	86.11	85.25	89.66	87.39	93.45	88.37
	GBM[d]	24	89.58	91.04	87.14	89.05	89.52	89**0.27
		11	86.11	89.81	83.62	86.61	86.31	86.49
	RF[e]	24	87.50	83.33	92.86	87.84	95.68	89.44
		11	88.43	89.57	88.79	89.18	94.80	90.15
3-scale	ANN	24	83.33	83.44	83.33	83.36	95.49	85.79
		13	72.45	72.55	72.45	72.17	87.87	75.50
	SVM	24	87.50	87.50	87.50	87.48	97.07	89.41
		13	72.99	72.97	72.99	72.91	87.31	75.83
	GBM	24	87.50	87.48	87.50	87.47	97.25	89.44
		13	86.42	86.47	86.42	86.41	96.62	88.47
	RF	24	79.40	79.35	79.40	79.34	94.60	82.42
		13	76.7	76.60	76.70	76.63	92.25	79.78

Note: *AUC-ROC*[a] stands for the area under the receiver operating characteristics curve; *ANN*[b] stands for artificial neural network; *SVM*[c] stands for support vector machine; *GBM*[d] stands for gradient boosting machine; *RF*[e] stands for random forest.

information, but it can also miss important patterns and increase noise. Therefore, the appropriate time window for prediction can vary by characteristics of predictors and target variables. For 2-scale models, 3-min physiological dataset showed the highest predictive performance while 1-min dataset was the most appropriate time window for 3-scale models. Specifically, the average predictive performance of 3-min time window for 2-scale models was 92.42 % and that of 1-min time window was 90.68 %. Also, the average predictive performance of 3-min time window for 3-scale model was 80.53 % while that of 1-min time window was 86.76 %. Compared

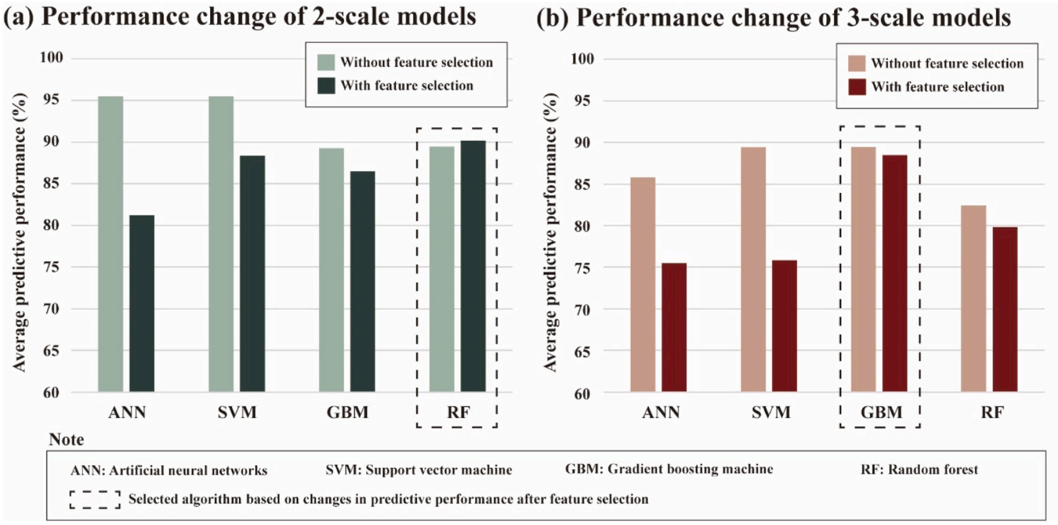

Fig. 8. Average predictive performance of models without feature selection.

Table 9
Predictive performance of visual fatigue models adding subject-label variables.

Scale	Algorithm	Subject-label variable	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC-ROC[a](%)	Average (%)
2-scale	RF[a]	None	88.43	89.57	88.79	89.18	94.80	90.15
		ID	90.74	91.38	91.38	91.38	96.30	92.24
		Gender	89.35	90.43	89.66	90.04	95.06	90.91
		Eyewear	90.74	93.64	88.79	91.15	95.34	91.93
		CS	92.59	93.86	92.24	93.04	96.91	93.73
		ID and CS	90.74	91.38	91.38	91.38	96.79	92.33
3-scale	GBM[b]	None	86.42	86.47	86.42	86.41	96.62	88.47
		ID	91.67	91.65	91.67	91.65	98.72	93.07
		Gender	87.81	87.78	87.81	87.77	97.53	89.74
		Eyewear	87.04	87.02	87.04	87.03	97.06	89.04
		CS	90.90	90.88	90.90	90.87	98.26	92.36
		ID and CS	93.52	93.56	93.52	93.51	99.09	94.64

Note: *RF*[a] stands for random forest; *GBM*[b] stands for gradient boosting machine.

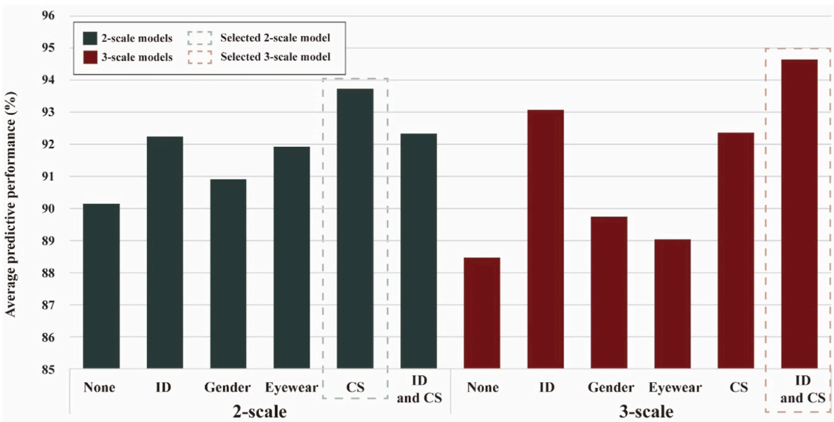

Fig. 9. Average predictive performance of models with different subject-label variables.

to these performances, 15-min time window of 2-scale and 3-scale showed large differences in average predictive performance, specifically 74.45 % and 69.41 % respectively. Future studies can refer to this time window to set appropriate time interval of physiological data collection for visual fatigue prediction.

Table 10
Predictive performance of visual fatigue models adding subject-label variables.

Scale	Algorithm	Time window	Predictor variables	Average (%)
2-scale	RF[a]	3-min	RMSSD, BA, PA_θ, OC_θ, FR_α, TE_α, PA_α, FR_β, TE_β, PA_β, OC_β, CS	93.73
3-scale	GBM[b]	1-min	CCT, LX, SDNN, PA_δ, OC_δ, TE_θ, PA_θ, PA_α, OC_α, FR_β, TE_β, PA_β, OC_β, CS, ID	94.64

Note: *RF*[a] stands for random forest; *GBM*[b] stands for gradient boosting machine.

After that, feature selection was conducted for each algorithm to reduce the number of predictor variables, and the algorithm with the least performance degradation was selected via comparative evaluation for each model. For 2-scale, RF was selected as the predictive performance improved (i.e., improved from 89.44 % to 90.15 %) instead of declining even after the number of predictors was reduced more than half. Previous studies on RF has demonstrated that RF can benefit from feature selection by eliminating irrelevant predictors which can cause noise or overfitting [74]. Meanwhile, GBM was selected for 3-scale model as there was the smallest reduction in average predictive performance (i.e., reduced by 0.97 %) after feature selection compared to other algorithms. GBM is known for its strong generalization ability in situations with complex features as it tends to effectively manage noise [67]. This characteristic might have contributed to the smallest performance degradation after feature selection.

- *Which physiological responses can be used as predictor variables to predict occupants' visual fatigue?*

Due to lack of information on potential predictors for visual fatigue prediction from previous studies, various physiological responses have been used for visual fatigue prediction and feature selection was conducted to investigate the feature importance of each variable. Lastly, the final performance of the visual fatigue prediction model was optimized by adding subject-label variables which are time-invariant. The results revealed that in addition to BA and PUI, heart rate variability data, such as RMSSD and SDNN, and brain wave data can be used as predictive indices of visual fatigue. However, feature importance investigated in this study showed each physiological response has varying impact on visual fatigue prediction. Firstly, while studies such as Liu et al.(2021) have suggested BA and PUI as the indices for visual fatigue, the feature importances of BA and PUI were not relatively high compared to other physiological responses [19]. Also, among the brain wave data, brain waves in the occipital lobe, which were found to be closely related to visual fatigue in the previous research [75,76], were included in the 2-scale and 3-scale models. However, it was found that brain waves in the parietal lobe are more important predictor variables than those in the occipital lobe in this study. Among different frequencies, alpha and beta waves were predictors with high feature importance in most cases. This finding aligns with the result of Hsu et al. (2013) that also revealed alpha and beta waves were the most effective indices for visual fatigue [77]. Lastly, heart rate variability data (i.e., MHR, RMSSD and SDNN) were also found out to be significant predictors as at least one of them was included in range after feature selection. Heart rate variability data has not been investigated as the predictor for visual fatigue in the past.

Among the subject labels, which are personal information, ID identifies an individual, and CS, the result of FrACT, were found to be the most effective predictor variables. The 3-scale model showed a 7 % improvement in accuracy when ID and CS were added as predictor variables. This result suggests that the visual fatigue model should be developed as a personalized model, rather than a general model like the current thermal comfort model. In addition, while the existing thermal comfort model is used to improve performance based on gender difference [78], contrast sensitivity, the individual's letter identification ability, seems to be a more effective predictor variable than gender difference in the visual fatigue model [17].

4.2. Implications and limitations of the study

The development of visual fatigue prediction model in this study holds several implications from academical and practical perspective. From academical perspective, this study has a novelty for being one of the initial studies that attempted to predict visual fatigue based on physiological responses. Most of the previous studies that predicted visual fatigue only considered environmental parameters. These studies have a limitation that their models cannot predict visual fatigue of individual occupant who uses visual display terminals. However, this study developed visual fatigue prediction model that uses physiological responses as predictor variables and thus, can predict individual occupants' visual fatigue more accurately. The most relevant model of this study is the model developed by Cen et al. (2019) which used occupants' pupil data for prediction and recorded the median predictive accuracy of 0.65 [24]. Compared to the previous model, this study has high average predictive performance of 93.73 % and 94.64 % for 2-scale and 3-scale model respectively. Also, visual fatigue prediction models in this study present various predictor variables to visual fatigue prediction-related studies in their early stages. In addition, it was suggested that in addition to pupillary data such as PUI and BA, heart rate variability data such as SDNN and RMSSD, and various brain waves can be effective visual fatigue predictor variables.

The visual fatigue prediction models developed in this study have practical implications as well. First of all, predictions on 2-scale and 3-scale visual fatigue level of occupants can be integrated into the future smart building systems to provide better lighting environment including task lightings. Also, visual fatigue data can be collected continuously based on physiological responses and the cumulatively collected data can be used for appropriate lighting design or control in specific office rooms. These models can contribute to individual occupants as well. The developed models are capable of intuitively giving information on the occurrence and degree of visual fatigue to a group of occupants. International organizations, such as the National Eye Institute in the United States and the World Health Organization, recommend the '20-20-20 rule', which requires occupants to look at something at least 20 feet away for 20 s after looking at a screen on a computer for 20 min to reduce visual fatigue. However, it is difficult for modern people to keep this rule in reality. Particularly, occupants who are insensitive to visual fatigue or who underestimate the risk of visual fatigue do not feel the need to take actions to reduce visual fatigue, such as taking a break, despite excessive visual perception activities, when excessive visual fatigue may lead to eye health deterioration among occupants. In this context, using the visual fatigue prediction model makes it possible for occupants to predict visual fatigue based on physiological responses and inform occupants with their visual fatigue level that they may not have noticed, and then take actions such as taking a break or adjusting lighting and monitor brightness appropriately.

However, this study has a few limitations that should be considered in future studies. First, most of the predictor variables were physiological responses that are difficult to measure with the current wearable devices, so the practical applicability of the model is low. Therefore, exploring visual fatigue predictor variables for physiological responses with relatively high levels of accessibility is recommended for future research. Second, because the data used to develop the visual fatigue prediction model was collected via the experiment performed on 16 subjects in their 20s and 30s, there is a lack of diversity in data. The model this study proposed can be applied to a group of people with similar sample size, but it cannot be proposed as a general model yet. In addition, visual fatigue may occur differently depending on age; therefore, it is necessary to verify whether the model can be applied to other age groups or occupants with dye diseases.

5. Conclusion

Eye health has a significant impact on the quality of life throughout the individual lifecycle and is directly related to social problems caused by the economic burden of eye diseases and productivity loss among the elderly population. Occupants in modern offices mostly work using various displays such as computers, unlike in the past when they performed paper-based work, and are frequently exposed to visual fatigue, a threat to eye health. Therefore, this study developed visual fatigue prediction models using classification algorithms and the physiological responses of occupants. First, experiments to collect subjects' physiological responses and subjective visual fatigue were conducted. Next, data preparation was performed to refine raw data and make predictor variables and target variables. Lastly, visual fatigue prediction models were developed using classification algorithms and their performances were evaluated to decide optimal 2-scale and 3-scale visual fatigue prediction models.

The main findings of this study are as following:

- Occupants' subjective visual fatigue can be effectively predicted using their physiological responses in data-driven way (i.e., machine learning algorithms)).
- 3-min and 1-min time window dataset showed the highest performance for 2-scale and 3-scale models, respectively. For both models, 15-min time window showed significantly low performance.
- Random forest (RF) and gradient boosting machine (GBM) demonstrated the highest predictive performance for 2-scale and 3-scale models, respectively.
- Consequently, the 2-scale model based on RF and 3-min predictor variables recorded an average performance of 93.73 %. Also, the 3-scale model based on GBM and 1-min predictor variables recorded an average performance of 94.64 %.
- Alpha and beta brain waves are significant predictor variables for visual fatigue. Heart rate variability data can be potential predictor variables as well.
- Subject-label variables, such as identification number and contrast sensitivity showed improvements in predictive performance.

The visual fatigue prediction models developed in this study are expected to contribute to helping occupants perceive their visual fatigue as they predict the visual fatigue of individual occupants via physiological responses and derive results as intuitively understandable classes. In future research, it will be necessary to develop visual fatigue prediction models targeting physiological responses that can be measured using wearable devices for enhanced usability and applicability. Also, as this study had collected data from a limited number of subjects, future visual fatigue prediction models should be more generalized by collecting data from subjects of various age groups and eye health conditions. Lastly, field experiments are required to validate the visual fatigue prediction models and to additionally consider variables such as daylighting or exposure time to visual tasks.

CRediT authorship contribution statement

Dahyun Jung: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jongbaek An:** Writing – review & editing, Visualization, Validation, Formal analysis, Data curation. **Taehoon Hong:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Minhyun Lee:** Writing – review & editing, Visualization, Validation.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; Ministry of Science and ICT) (NRF-2021R1A3B1076769).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jobe.2024.111146.

Nomenclature

ANN	Artificial neural network
AUC-ROC	Area under the receiver operating characteristics curve
BA	Blink amplitude
BP	Brightness preference
BSV	Brightness sensation vote
CCT	Correlated color temperature
CP	Color temperature preference
CS	Contrast sensitivity
CSV	Correlated color temperature sensation vote
ECG	Electrocardiography
EEG	Electroencephalography
EOG	Electrooculography
FrACT	Freiburg visual acuity test
GBM	Gradient boosting machine
MHR	Mean heart rate
PCA	Principal component analysis
PSD	Power spectral density
PUI	Pupillary unrest index
RF	Random forest
RFE	Recursive feature elimination
RMSSD	Root mean square of successive differences
SDNN	Standard deviation of NN intervals
SFS	Sequential feature selection
SVM	Support vector machine
TSV	Thermal sensation vote
VCV	Visual comfort vote

Data availability

Data will be made available on request.

References

[1] M.J. Burton, J. Ramke, A.P. Marques, R.R.A. Bourne, N. Congdon, I. Jones, B.A.M. Ah Tong, S. Arunga, D. Bachani, C. Bascaran, A. Bastawrous, K. Blanchet, T. Braithwaite, J.C. Buchan, J. Cairns, A. Cama, M. Chagunda, C. Chuluunkhuu, A. Cooper, J. Crofts-Lawrence, W.H. Dean, A.K. Denniston, J.R. Ehrlich, P. M. Emerson, J.R. Evans, K.D. Frick, D.S. Friedman, J.M. Furtado, M.M. Gichangi, S. Gichuhi, S.S. Gilbert, R. Gurung, E. Habtamu, P. Holland, J.B. Jonas, P. A. Keane, L. Keay, R.C. Khanna, P.T. Khaw, H. Kuper, F. Kyari, V.C. Lansingh, I. Mactaggart, M.M. Mafwiri, W. Mathenge, I. McCormick, P. Morjaria, L. Mowatt, D. Muirhead, G.V.S. Murthy, N. Mwangi, D.B. Patel, T. Peto, B.M. Qureshi, S.R. Salomão, V. Sarah, B.R. Shilio, A.W. Solomon, B.K. Swenor, H.R. Taylor, N. Wang, A. Webson, S.K. West, T.Y. Wong, R. Wormald, S. Yasmin, M. Yusufu, J.C. Silva, S. Resnikoff, T. Ravilla, C.E. Gilbert, A. Foster, H.B. Faal, The Lancet global health commission on global eye health: vision beyond 2020, Lancet Global Health 9 (2021) e489–e551, https://doi.org/10.1016/S2214-109X(20)30488-5.
[2] D. Man, R. Olchawa, The possibilities of using BCI technology in biomedical engineering, Adv. Intell. Syst. Comput. (2018), https://doi.org/10.1007/978-3-319-75025-5_4.
[3] S.A. Haymes, A.W. Johnston, A.D. Heyes, Relationship between vision impairment and ability to perform activities of daily living, Ophthalmic Physiol. Opt. 22 (2002) 79–91, https://doi.org/10.1046/j.1475-1313.2002.00016.x.
[4] N. Jones, H.E. Bartlett, R. Cooke, An analysis of the impact of visual impairment on activities of daily living and vision-related quality of life in a visually impaired adult population, Br. J. Vis. Impair. (2019), https://doi.org/10.1177/0264619618814071.
[5] B.N. Ekpenyong, D. Echendu, E. Ekanem, Visual health status and its relationship with road traffic accidents amongst Nigerian vehicle drivers: a publication of the Nigerian Optometric Association, African Vis, Eye Heal. 79 (2020) 1–8, https://doi.org/10.4102/AVEH.V79I1.577.
[6] B.K. Swenor, J.R. Ehrlich, Ageing and vision loss: looking to the future, Lancet Global Health 9 (2021) e385–e386, https://doi.org/10.1016/S2214-109X(21)00031-0.
[7] A.W. Scott, N.M. Bressler, S. Ffolkes, J.S. Wittenborn, J. Jorkasky, Public attitudes about eye and vision health, JAMA Ophthalmol 134 (2016) 1111–1118, https://doi.org/10.1001/jamaophthalmol.2016.2627.
[8] K.A. Eckert, M.J. Carter, V.C. Lansingh, D.A. Wilson, J.M. Furtado, K.D. Frick, S. Resnikoff, A simple method for estimating the economic cost of productivity loss due to blindness and moderate to severe visual impairment, Ophthalmic Epidemiol. 22 (2015) 349–355, https://doi.org/10.3109/09286586.2015.1066394.

[9] J.H. Zhang, J. Ramke, C. Jan, C. Bascaran, N. Mwangi, J.M. Furtado, S. Yasmin, C. Ogundo, M. Yoshizaki, A.P. Marques, J. Buchan, P. Holland, B.A.M. Ah Tong, J.R. Evans, N. Congdon, A. Webson, M.J. Burton, Advancing the Sustainable Development Goals through improving eye health: a scoping review, Lancet Planet. Health 6 (2022) e270–e280, https://doi.org/10.1016/S2542-5196(21)00351-X.
[10] P.R. Boyce, Light, lighting and human health, Light. Res. Technol. (2022), https://doi.org/10.1177/14771535211010267.
[11] P. Reviewed, L. Berkeley, B. Cancer, Lawrence Berkeley National Laboratory Lawrence Berkeley National Laboratory, 2010, pp. 35–43.
[12] Y. Lin, C.C. Chen, Y. Ashraf Gandomi, Strategies on visual display terminal lighting in office space under energy-saving environment, Energies (2023), https://doi.org/10.3390/en16031317.
[13] A. Kuwahara, K. Nishikawa, R. Hirakawa, H. Kawano, Y. Nakatoh, Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping(EARM), Cogn. Robot. 2 (2022) 50–59, https://doi.org/10.1016/j.cogr.2022.01.003.
[14] D. Jung, J. An, T. Hong, Exploring the relationship between office lighting, cognitive performance, and psychophysiological responses: A multidimensional approach, Building and Environment 263 (2024) 111863.
[15] R. Dang, Y. Liu, S. Chang, The impact patterns of classroom lighting parameters on visual fatigue and a mathematical model, Build. Environ. 234 (2023) 110193, https://doi.org/10.1016/j.buildenv.2023.110193.
[16] J. Jeoung, S. Jung, T. Hong, M. Lee, C. Koo, Thermal comfort prediction based on automated extraction of skin temperature of face component on thermal image, Energy Build. 298 (2023) 113495, https://doi.org/10.1016/j.enbuild.2023.113495.
[17] H. Kim, D. Jung, H. Choi, T. Hong, Advanced prediction model for individual thermal comfort considering blood glucose and salivary cortisol, Build. Environ. (2022), https://doi.org/10.1016/j.buildenv.2022.109551.
[18] H. Kim, H. Kang, H. Choi, D. Jung, T. Hong, Human-building interaction for indoor environmental control: evolution of technology and future prospects, Autom. ConStruct. 152 (2023) 104938, https://doi.org/10.1016/j.autcon.2023.104938.
[19] Y. Liu, S. Zhang, Y. Wu, D. Yang, Studies on visual health features of luminous environment in college classrooms, Build. Environ. (2021), https://doi.org/10.1016/j.buildenv.2021.108184.
[20] J. Choi, Visual fatigue modeling and analysis for stereoscopic video, Opt. Eng. 51 (2012) 017206, https://doi.org/10.1117/1.oe.51.1.017206.
[21] Y. Shi, Y. Tu, L. Wang, N. Zhu, AtLSMMs network: an attentional-biLSTM based multi-model prediction for smartphone visual fatigue, Displays 84 (2024) 102754, https://doi.org/10.1016/j.displa.2024.102754.
[22] K. Van Den Wymelenberg, M. Inanici, A critical investigation of common lighting design metrics for predicting human visual comfort in offices with daylight, LEUKOS - J. Illum. Eng. Soc. North Am. (2014), https://doi.org/10.1080/15502724.2014.881720.
[23] A. Ghourkhanehchi Zirak, A. Saeedian, Z.S. Zomorodian, M. Tahsildoost, Enhancing visual comfort prediction in space layouts using deep learning models, Intell. Build. Int. 15 (2023) 266–284, https://doi.org/10.1080/17508975.2024.2351800.
[24] L. Cen, J.H. Choi, X. Yao, Y. Gil, S. Narayanan, M. Pentz, A personal visual comfort model: predict individual's visual comfort using occupant eye pupil size and machine learning, IOP Conf. Ser. Mater. Sci. Eng. 609 (2019), https://doi.org/10.1088/1757-899X/609/4/042097.
[25] R.R. Baniya, E. Tetri, L. Halonen, A study of preferred illuminance and correlated colour temperature for LED office lighting, Light Eng. 23 (3) (2015) 39.
[26] C. Liu, L. Sun, X. Jing, Y. Zhang, X. Meng, C. Jia, W. Gao, How correlated color temperature (CCT) affects undergraduates: a psychological and physiological evaluation, J. Build. Eng. (2022), https://doi.org/10.1016/j.jobe.2021.103573.
[27] E. Kürkçü, The Effects of Color Temperature on Performance and Mood of Users: A Color Task Implementation Model, vol. 11, Master's Thesis, Bilkent Univ., 2017, pp. 92–105.
[28] C. Dai, N. Li, A.H. Shah, Y. Lin, Human perception and preferences of evening office lighting in terms of the LED spectrum, in: 2021 18th China Int. Forum Solid State Light. 2021 7th Int. Forum Wide Bandgap Semicond. SSLChina IFWS 2021, 2021, https://doi.org/10.1109/SSLChinaIFWS54608.2021.9675244.
[29] M.L. Wang, M.R. Luo, Effects of LED lighting on office work performance, in: Conf. Proc. - 2016 13th China Int. Forum Solid State Light, SSLCHINA, 2016, 2017, https://doi.org/10.1109/SSLCHINA.2016.7804366.
[30] H. Choi, H. Kim, T. Hong, J. An, Examining the indirect effects of indoor environmental quality on task performance: the mediating roles of physiological response and emotion, Build. Environ. (2023), https://doi.org/10.1016/j.buildenv.2023.110298.
[31] H. Kim, T. Hong, J. Kim, S. Yeom, A psychophysiological effect of indoor thermal condition on college students' learning performance through EEG measurement, Build. Environ. (2020), https://doi.org/10.1016/j.buildenv.2020.107223.
[32] M. Islam, T. Ahmed, M.S.U. Yusuf, M. Ahmad, Cognitive state estimation by effective feature extraction and proper channel selection of EEG signal, J. Circ. Syst. Comput. (2015), https://doi.org/10.1142/S0218126615400058.
[33] D. Choi, S. Seo, H. Park, T. Hong, C. Koo, Forecasting personal learning performance in virtual reality-based construction safety training using biometric responses, Autom. ConStruct. 156 (2023) 105115, https://doi.org/10.1016/j.autcon.2023.105115.
[34] T. Elbert, W. Lutzenberger, B. Rockstroh, N. Birbaumer, Removal of ocular artifacts from the EEG - a biophysical approach to the EOG, Electroencephalogr. Clin. Neurophysiol. 60 (1985) 455–463, https://doi.org/10.1016/0013-4694(85)91020-X.
[35] H. Kim, T. Hong, Enhancing emotion recognition using multimodal fusion of physiological, environmental, personal data, Expert Syst. Appl. (2024), https://doi.org/10.1016/j.eswa.2024.123723.
[36] J. Hong, J. Choi, J.H. Lee, S.H. Cho, T. Hong, S.U. Han, H.S. Park, D.E. Lee, Virtual reality-based analysis of the effect of construction noise exposure on masonry work productivity, Autom. ConStruct. 150 (2023) 104844, https://doi.org/10.1016/j.autcon.2023.104844.
[37] D.C. Niehorster, R.S. Hessels, J.S. Benjamins, GlassesViewer: open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker, Behav. Res. Methods 52 (2020) 1244–1253, https://doi.org/10.3758/s13428-019-01314-1.
[38] J. Li, H. Li, W. Umer, H. Wang, X. Xing, S. Zhao, J. Hou, Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology, Autom. ConStruct. 109 (2020) 103000, https://doi.org/10.1016/j.autcon.2019.103000.
[39] D. Jung, H. Kim, J. An, T. Hong, Thermoregulatory responses of young and elderly adults under temperature ramps, Build. Environ. (2023), https://doi.org/10.1016/j.buildenv.2023.110760.
[40] M. Papinutto, J. Nembrini, D. Lalanne, "Working in the dark?" investigation of physiological and psychological indices and prediction of back-lit screen users' reactions to light dimming, Build. Environ. (2020), https://doi.org/10.1016/j.buildenv.2020.107356.
[41] J. Jeoung, S. Jung, T. Hong, J.K. Choi, Blockchain-based IoT system for personalized indoor temperature control, Autom. ConStruct. 140 (2022) 104339, https://doi.org/10.1016/j.autcon.2022.104339.
[42] K.I. Fostervold, J. Nersveen, Proportions of direct and indirect indoor lighting - the effect on health, well-being and cognitive performance of office workers, Light. Res. Technol. 40 (2008) 175–197, https://doi.org/10.1177/1477153508090917.
[43] K.S. Liu, C.M. Chiang, Y. Sen Lin, Influences of visual fatigue on the productivity of subjects using visual display terminals in a light-emitting diode lighting environment, Architect. Sci. Rev. (2010), https://doi.org/10.3763/asre.2009.0083.
[44] P.S. Sachdev, D. Blacker, D.G. Blazer, M. Ganguli, D.V. Jeste, J.S. Paulsen, R.C. Petersen, Classifying neurocognitive disorders: the DSM-5 approach, Nat. Rev. Neurol. (2014), https://doi.org/10.1038/nrneurol.2014.181.
[45] J.P. Guilford, The Nature of Human Intelligence, 1967.
[46] J.D. Patterson, H.M. Merseal, D.R.A. Johnson, M.B. Sergio Baas, B.S. Barbot, B. Benedek, M. Borhani, K. Chen, Q. Christensen, J.F. Corazza, G.E. Forthmann, B. Karwowski, M. Kazemian, N. Kreisberg-Nitzav, A. Kenett, Y.N. Link, A. Lubart, T. Mercier, M. Miroshnik, K. Ovando-Tellez, M. Primi, R. Puente-Díaz, R. Said-Metwaly, S. Stevenson, C. Vartanian, M. Volle, E. van Hell, J.G. Beaty, R. E, Multilingual semantic distance: automatic verbal creativity assessment in many languages, Psychol. Aesthetics, Creat. Arts 17 (2023) 495–507, https://doi.org/10.1037/aca0000618.
[47] X. Xu, M. Xia, J. Zhao, W. Pang, Be real, open, and creative: how openness to experience and to change mediate the authenticity-creativity association, Think. Skills Creativ. (2021), https://doi.org/10.1016/j.tsc.2021.100857.
[48] F.E. Subekti, Wardono Zaenuri, Students' mathematical reasoning abilities on number sequence pattern material: viewed from a gender perspective, J. Phys. Conf. Ser. (2021), https://doi.org/10.1088/1742-6596/1918/4/042107.

[49] Z. Hamedani, E. Solgi, T. Hine, H. Skates, Revealing the relationships between luminous environment characteristics and physiological, ocular and performance measures: an experimental study, Build. Environ. (2020), https://doi.org/10.1016/j.buildenv.2020.106702.
[50] M. Bach, FrACT-landolt-vision, Optom. Vis. Sci. 73 (1996) 49–53.
[51] F. Linhart, J.L. Scartezzini, Evening office lighting - visual comfort vs. energy efficiency vs. performance? Build. Environ. 46 (2011) 981–989, https://doi.org/10.1016/j.buildenv.2010.10.002.
[52] X. Zheng, G. Xu, Y. Wu, Y. Wang, C. Du, Y. Wu, S. Zhang, C. Han, Comparison of the performance of six stimulus paradigms in visual acuity assessment based on steady-state visual evoked potentials, Doc. Ophthalmol. (2020), https://doi.org/10.1007/s10633-020-09768-x.
[53] J. Kim, H. Kim, T. Hong, Automated classification of indoor environmental quality control using stacked ensembles based on electroencephalograms, Comput. Civ. Infrastruct. Eng. 35 (2020) 448–464, https://doi.org/10.1111/mice.12515.
[54] J.D. Hunter, J.G. Milton, H. Lüdtke, B. Wilhelm, H. Wilhelm, Spontaneous fluctuations in pupil size are not triggered by lens accommodation, Vis. Res. 40 (2000) 567–573, https://doi.org/10.1016/S0042-6989(99)00200-X.
[55] H. Lüdtke, B. Wilhelm, M. Adler, F. Schaeffel, H. Wilhelm, Mathematical procedures in data recording and processing of pupillary fatigue waves, Vis. Res. (1998), https://doi.org/10.1016/S0042-6989(98)00081-9.
[56] I. Primpas, G. Tsirtsis, M. Karydis, G.D. Kokkoris, Principal component analysis: development of a multivariate index for assessing eutrophication according to the European water framework directive, Ecol. Indicat. (2010), https://doi.org/10.1016/j.ecolind.2009.04.007.
[57] S. Giri, A.K. Singh, Assessment of metal pollution in groundwater using a novel multivariate metal pollution index in the mining areas of the Singhbhum copper belt, Environ. Earth Sci. (2019), https://doi.org/10.1007/s12665-019-8200-9.
[58] H.S. Huh, C.Y. Park, Asia-Pacific regional integration index: construction, interpretation, and comparison, J. Asian Econ. (2018), https://doi.org/10.1016/j.asieco.2017.12.001.
[59] M. Deng, X. Wang, C.C. Menassa, Measurement and prediction of work engagement under different indoor lighting conditions using physiological sensing, Build. Environ. 203 (2021) 108098, https://doi.org/10.1016/j.buildenv.2021.108098.
[60] M.I. Jordan, T.M. Mitchell, Machine Learning: Trends, Perspectives, and Prospects, 2015, p. 349.
[61] S. Makridakis, E. Spiliotis, V. Assimakopoulos, A.A. Semenoglou, G. Mulder, K. Nikolopoulos, Statistical, machine learning and deep learning forecasting methods: comparisons and ways forward, J. Oper. Res. Soc. (2023), https://doi.org/10.1080/01605682.2022.2118629.
[62] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, J. Pharm. Biomed. Anal. 22 (2000) 717–727, https://doi.org/10.1016/S0731-7085(99)00272-1.
[63] T. Guillod, P. Papamanolis, J.W. Kolar, Artificial neural network (ann) based fast and accurate inductor modeling and design, IEEE Open J. Power Electron. (2020), https://doi.org/10.1109/OJPEL.2020.3012777.
[64] S. Suthanharan, Support Vector Machine, 2016.
[65] D.M. Abdullah, A.M. Abdulazeez, Machine learning applications based on SVM classification: a review, Qubahan Acad. J. (2021), https://doi.org/10.48161/qaj.v1n2a50.
[66] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorob. 7 (2013), https://doi.org/10.3389/fnbot.2013.00021.
[67] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. (2021), https://doi.org/10.1007/s10462-020-09896-5.
[68] G. Biau, E. Scornet, A random forest guided tour, Test 25 (2016) 197–227, https://doi.org/10.1007/s11749-016-0481-7.
[69] R. Genuer, J.-M. Poggi, Random forests. https://doi.org/10.1007/978-3-030-56485-8_1, 2020.
[70] A. Hoła, S. Czarnecki, Random forest algorithm and support vector machine for nondestructive assessment of mass moisture content of brick walls in historic buildings, Autom. ConStruct. 149 (2023), https://doi.org/10.1016/j.autcon.2023.104793.
[71] V.R. Joseph, Optimal ratio for data splitting, Stat. Anal. Data Min. 15 (2022) 531–538, https://doi.org/10.1002/sam.11583.
[72] K. Haruehansapong, M. Kliangkhlao, K. Yeranee, B. Sahoh, Personal thermal comfort prediction using multi-physiological sensors: the design and development of deep neural network models based on individual preferences, Build. Environ. (2023), https://doi.org/10.1016/j.buildenv.2023.110940.
[73] C. Shan, J. Hu, J. Wu, A. Zhang, G. Ding, L.X. Xu, Towards non-intrusive and high accuracy prediction of personal thermal comfort using a few sensitive physiological parameters, Energy Build. (2020), https://doi.org/10.1016/j.enbuild.2019.109594.
[74] R. Iranzad, X. Liu, A review of random forest-based feature selection methods for data science education and applications, Int. J. Data Sci. Anal. (2024), https://doi.org/10.1007/s41060-024-00509-w.
[75] A. Campagne, T. Pebayle, A. Muzet, Correlation between driving errors and vigilance level: influence of the driver's age, Physiol. Behav. (2004), https://doi.org/10.1016/j.physbeh.2003.10.004.
[76] L.E. Ismail, W. Karwowski, Applications of EEG indices for the quantification of human cognitive performance: a systematic review and bibliometric analysis, PLoS One (2020), https://doi.org/10.1371/journal.pone.0242857.
[77] B.W. Hsu, M.J.J. Wang, Evaluating the effectiveness of using electroencephalogram power indices to measure visual fatigue, Percept, Mot. Skills (2013), https://doi.org/10.2466/29.15.24.PMS.116.1.235-252.
[78] T. Chaudhuri, D. Zhai, Y.C. Soh, H. Li, L. Xie, Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology, Energy Build. 166 (2018) 391–406, https://doi.org/10.1016/j.enbuild.2018.02.035.