

Article

Not peer-reviewed version

---

# Mental Workload Modeling by Using Machine Learning Techniques Based on EEG and Eye Tracking Data

---

[Seniz Harputlu Aksu](#), [Erman Çakıt](#)<sup>\*</sup>, Metin Dağdeviren

Posted Date: 8 February 2024

doi: 10.20944/preprints202402.0452.v1

Keywords: EEG; eye tracking; mental workload; machine learning; neuroergonomics; prediction; NASA-TLX; N-back



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Mental Workload Modeling by Using Machine Learning Techniques Based on EEG and Eye Tracking Data

Şeniz Harputlu Aksu <sup>1</sup>, Erman Çakıt <sup>2,\*</sup> and Metin Dağdeviren <sup>2</sup>

<sup>1</sup> The Scientific and Technological Research Council of Türkiye, Ankara 06530, Türkiye  
seniz.harputlu@tubitak.gov.tr

<sup>2</sup> Department of Industrial Engineering, Gazi University, Ankara 06570, Türkiye; ecakit@gazi.edu.tr (E.Ç.);  
metindag@gazi.edu.tr (M.D.)

\* Correspondence: ecakit@gazi.edu.tr

**Abstract:** The objective of this study was to analyze the mental workload using EEG and eye tracking data and classify it using machine learning algorithms. The machine learning model was developed based on the simultaneous recording of eye tracking and EEG measurements during the experimental process. The experiments involved 15 university students, consisting of 7 women and 8 men. Throughout the experiments, the researchers utilized the n-back memory task and the NASA-Task Load Index (TLX) subjective rating scale to assess various levels of mental workload. The findings revealed that as the task difficulty increased, there was an increase in the diameter of both the right and left pupils, the number of fixations, the number and duration of saccades, and the number and duration of blinks. Conversely, variables related to fixation duration decreased. The EEG results indicated that theta power in the prefrontal, frontal, and front central regions increased with task difficulty. Additionally, alpha power increased in the frontal regions but decreased in the temporal, parietal, and occipital regions as the task became more challenging. Furthermore, low beta power significantly decreased in almost all brain regions as the task difficulty increased. In terms of the four-class classification problem, the mental workload level can be predicted with an accuracy rate of 76.59% using 34 selected features. This study has made a significant contribution to the literature by presenting a four-class mental workload estimation model that utilizes different machine learning algorithms.

**Keywords:** EEG; eye tracking; mental workload; machine learning; neuroergonomics; prediction; NASA-TLX; N-back

## 1. Introduction

The elucidation and measurement of mental workload are essential endeavors in the field of cognitive sciences. In a time marked by a constant influx of information and a growing range of cognitive requirements, the significance of methodically assessing and understanding mental workload becomes paramount. According to Sweller (1998) in cognitive load theory, it is believed that an individual's working memory capacity is limited. Therefore, task designers should minimize external cognitive load sources to allow employees to effectively focus their cognitive resources on completing the primary task. If the task requires more mental resources than the available capacity, it will result in mental overload (Bommer and Fendley, 2018). Considering mental load is not just a niche pursuit, but a crucial aspect in optimizing cognitive performance, reducing mental fatigue, and protecting against the potential negative consequences of mental overload. In essence, the assessment of mental workload remains a significant component in striving for more comfortable, satisfying, productive, and safer work environments (Rubio et al., 2004).

Given the significance of the concept of mental workload, there is a necessity for precise and efficient tools to measure it. However, measuring mental workload directly and effortlessly is not feasible due to its multidimensional nature and susceptibility to various factors. Consequently, recent

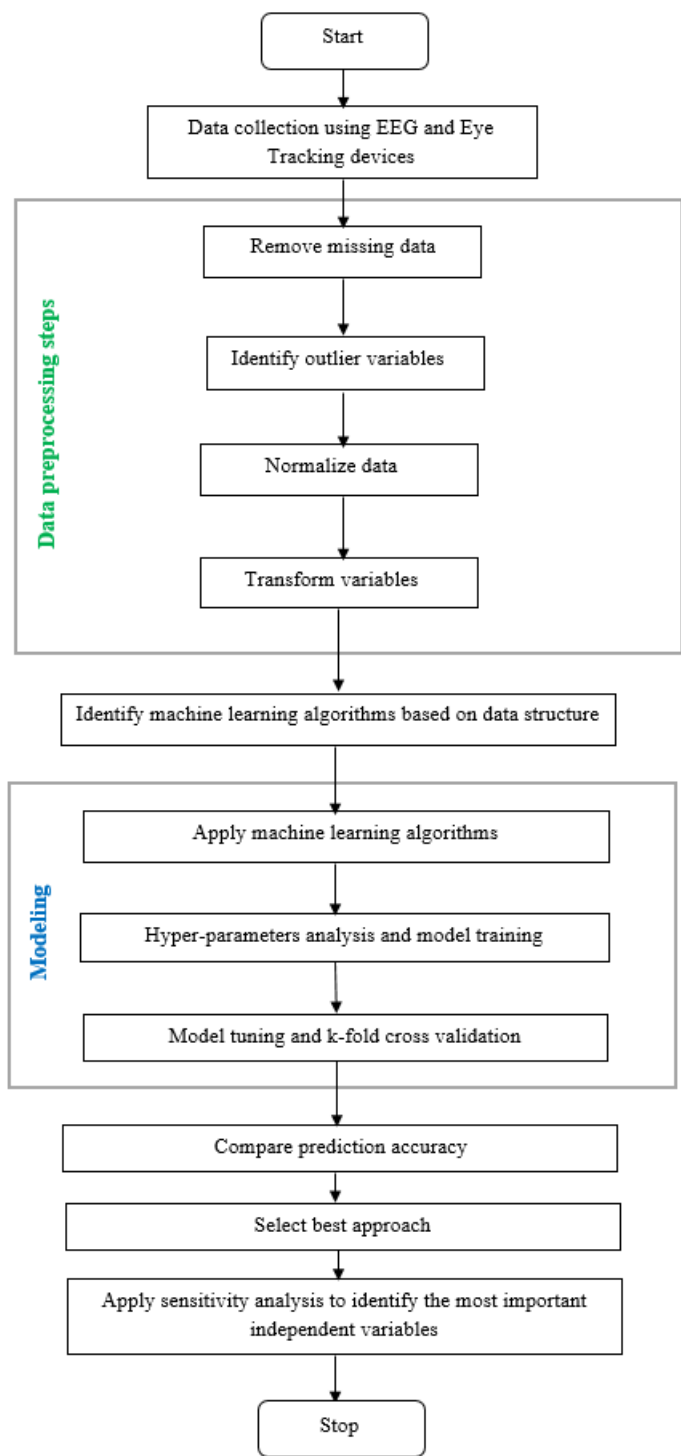
studies have focused on predicting mental workload, primarily approaching it as a classification problem with the objective of estimating the level of mental workload categorically. These studies differ in terms of the task type employed in the experiments, the techniques utilized to obtain input variables for the model, the number of predicted classes, and the algorithms employed to address the problem. Despite these efforts, developing a model that can accurately predict mental workload remains an unresolved issue in the existing literature.

One of the primary objectives of this study is to investigate the correlation between physiological variables and mental workload, and compare it with existing literature. Another crucial objective is to expand the scope of the problem, which is typically categorized into two or three classes in previous studies, by considering four classes. This will provide a more detailed understanding of workload levels and enable the development of a classification model with the highest possible accuracy. Furthermore, this study stands out due to its unique approach of simultaneously collecting eye tracking and EEG data from participants. Although synchronizing the data from both devices presents challenges during the experiment and analysis process, previous findings have shown that combining physiological methods, which are known to be influenced by fluctuations in mental workload, enhances the performance of the model. Hence, EEG and eye tracking methods were employed together in this study, which is a novel contribution to the literature as no previous studies have tested both n-back tasks while utilizing EEG and eye tracking methods simultaneously.

In this study, we employed EEG and eye tracking as physiological techniques, which were further complemented by the subjective evaluation scale NASA-TLX (Hart and Staveland, 1988). Additionally, the raw EEG data was processed using the MATLAB software-based EEGLAB tool, which not only facilitates the visualization and interpretation of statistical analysis results but also serves as an exemplary application of this tool in the literature.

In summary, this study has three main objectives:

1. Analyzing the relationship between EEG and eye-related variables with task difficulty level and subjective workload evaluation, and comparing them with similar studies in the literature.
2. Predicting task difficulty level, which is evaluated as the mental workload, using EEG and eye tracking data.
3. Incorporating machine learning algorithms and utilizing the EEGLAB tool to enhance the analysis and interpretation of the results, thereby contributing to the existing literature. This research was carried out in the order shown in Figure 1.



**Figure 1.** Study Steps.

The remaining part of this study is structured in the following manner. The literature review is provided in the subsequent section titled “Background”. The section titled “Materials and methods” outlines the study design and introduces the implementation of machine learning techniques. The comparison of the results obtained from the various machine learning methods is discussed in the section titled “Results and discussion”. Lastly, the section titled “Conclusions” evaluates the findings and provides recommendations for future research endeavors.

## 2. Background

Mental workload is a “construct that reflects the relation between the environmental demands imposed on the human operator and the operator’s capabilities to meet those demands” (Parasuraman and Hancock, 2001). For mental workload assessment, the eye tracking technique is favored due to its ability to minimize task interruption compared to other physiological methods. Additionally, it can gather multiple types of information about the eye using just one sensor. On the other hand, the EEG method is selected for its effectiveness in capturing real-time temporal information. One potential advantage of utilizing eye movements in the assessment of mental workload is the ability to capture rapid fluctuations in workload that occur within a short timeframe (Ahlstrom and Friedman-Berg, 2006). Eye tracking also offers benefits such as quick device setup, precise and sufficient data collection, and the ability to accommodate slight head movements. Specifically, studies have demonstrated a correlation between pupil response, blink movement, and visual as well as auditory tasks (Benedetto et al., 2011; Gao et al., 2013; Reiner and Gelfeld, 2014; Wanyan et al., 2018; Radhakrishnan et al., 2023).

In contrast to other neuroimaging techniques, EEG exhibits a relatively high temporal resolution. It offers a real-time and unbiased assessment, and its ease of application is attributed to its portable device. Moreover, EEG is considered a safe method for long-term and repeated use due to its non-invasive nature. Notably, it does not necessitate the presence of a skilled operator or involve any medical procedures (Grimes et al., 2008). EEG indices have demonstrated a remarkable sensitivity to fluctuations in brain activity. Several studies applying machine learning algorithms using eye tracking and EEG data (Table 1). In the research conducted by Liu et al. (2017), individuals were instructed to perform an n-back memory task in order to assess the workload levels of 21 participants across 3 distinct scenarios. Simultaneously, the participants' EEG, fNIRS data, and additional physiological measurements including heart rate, respiratory rate, and heart rate variability were recorded. The findings of the study unveiled that the combination of EEG and fNIRS techniques yielded more precise outcomes compared to their individual usage, with an accuracy rate of 65%.

Jusas and Samuvel (2019) conducted a study involving 9 individuals, where they employed 4 distinct motor imagery tasks. These tasks required participants to engage in mental processes where they practiced or simulated specific actions. The study solely utilized EEG physiological data, and the highest classification performance achieved was 64% through the implementation of Linear Discriminant Analysis. In their research on 7 individuals, Yin and Zhang (2017) performed a prediction model for mental workload using binary classification. The study concluded with a comparison of the model's estimator accuracy performance, which achieved a remarkable 85.7%, against traditional classifiers such as support vector machines and nearest neighbors. The findings demonstrated the superiority of this method over static classifiers. Pei et al. (2021) conducted a study involving 7 individuals, where they achieved an accuracy performance of 84.3% using the RF algorithm. They established a model by utilizing various EEG attributes obtained from a 64-channel EEG device during a simulation flight mission. Furthermore, the classification performance of the 3-level model, which solely utilized band power variables, reached 75.9%. In a study conducted by Wang et al. (2012), the hierarchical bayes model's cross-subject classifier was trained and tested using EEG data from a cohort of 8 participants. The findings of the research indicate that the performance of the cross-subject classifier is similar to that of individual-specific classifiers.

In the research carried out by Kaczorowska et al. (2021), a number symbol matching test with three levels of difficulty was administered to a group of 29 individuals. The accuracy of the 20-variable model, constructed using eye tracking data, was determined to be 95% using logistic regression and Random Forest algorithms. In the research conducted by Sassaroli et al. (2008), the fNIRS technique was utilized on three individuals to differentiate various levels of mental workload by analyzing the hemodynamic changes in the prefrontal cortex. To determine the mental workload levels, the K Nearest Neighbor algorithm was employed, considering the amplitude of oxyhemoglobin and deoxyhemoglobin ( $k=3$ ). The study revealed classification success rates ranging from 44% to 72%. Grimes et al. (2008) conducted an early study that classified mental workload through n-back tasks, achieving a commendable level of accuracy. By employing the Naive Bayes



algorithm, they determined that the 2-difficulty level model had an accuracy rate of 99%, while the 4-difficulty level model achieved an accuracy rate of 88%. These measurements were obtained using a 32-channel EEG device and involved a sample of 8 individuals.

Wu et al. (2019) conducted a study on prediction using the eye tracking method. The study involved 32 non-experts and 7 experts who were tasked with performing operational tasks at 2 varying difficulty levels. The researchers used pupil diameter, blink rate, focusing speed, and saccade speed as inputs in their model, while the NASA-TLX score served as the output. The results showed that the accuracy of the model, obtained through the use of ANN, was 97% when tested with the data. A research conducted on the cognitive burden of automobile operators demonstrated that a feed-forward neural network classifier achieved a prediction accuracy of over 90% in determining levels of alertness, drowsiness, and sleep states (Subasi, 2005). In another study, a group of 28 individuals engaged in a computer-based flight mission. During this study, the brain signals of the participants were recorded using a 32-channel EEG device, while eye data was collected using an eye-tracking device operating at a frequency of 256 Hz (Li et al., 2023). The study employed various algorithms such as linear regression, support vector machines, random forests, and artificial neural networks. Among these algorithms, the model that incorporated both eye tracking and EEG variables yielded the highest accuracy rate of 54.2% in the 3-category classification task. SVM has been utilized in various studies within the literature to obtain mental workload classification results using eye tracking or EEG data (Borys et al., 2017; Kaczorowska et al., 2020; Qu et al., 2020; Lim et al., 2018; Şaşmaz et al., 2023; Zhou et al., 2022; Plechawska-Wojcik et al., 2019; Le et al., 2018).

Upon reviewing the pertinent literature, it became evident that both EEG and eye tracking techniques have significant implications in the analysis of mental workload. Nevertheless, it was observed that there exists incongruity in the findings regarding the correlation between the variables in both methods and mental workload. Notably, certain studies have demonstrated an augmentation of eye-related variables with increased mental workload, while others have shown a decline. The current study distinguishes itself from other studies by incorporating machine learning algorithms that have been developed in the past decade (Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM)) but have not been previously utilized in studies aiming to estimate mental workload using EEG and eye tracking techniques.

### 3. Materials and Methods

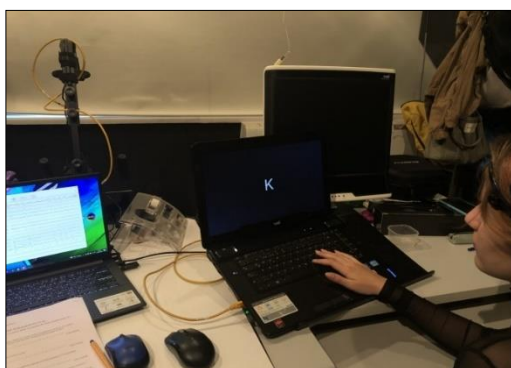
#### 3.1. N-Back Task

Given that cognitive load theory was developed based on the limitations of working memory capacity, it is logical to select task types that specifically target working memory for experimental studies on mental workload manipulation. N-back tasks are commonly utilized in research literature due to their reliability in assessing working memory capacity and their validity, which has been demonstrated in numerous studies (Grimes et al., 2008; Herff et al., 2014; Ke et al., 2015; Liu et al., 2017; Tjolleng et al., 2017; Aksu and Çakıt, 2023; Aksu et al., 2023; Harputlu Aksu and Çakıt, 2022)). These memory tests were also employed in this particular study due to their established validity in previous research. N-back tasks inherently involve both the storage and manipulation of information, surpassing traditional working memory tasks. The level of "n" in the n-back task is directly related to the "working memory load," representing the mental demand required for the storage and/or manipulation processes involved (Öztürk, 2018). The n-back paradigm is well-established, with strong correlations between difficulty level and cortical activation associated with working memory. Monod and Kapitaniak (1999) also suggested that task difficulty, an internal cognitive load factor, directly influences cognitive load. Previous research has shown that task difficulty, determined by the number of items to be remembered, hampers performance in memory tasks and affects psychophysiological measurements, particularly components regulated by the autonomic nervous system (Galy et al., 2012). As task difficulty is a crucial determinant of mental workload, it was selected as the target variable in the classification model developed for this study.

### 3.2. Participants and Experimental Procedure

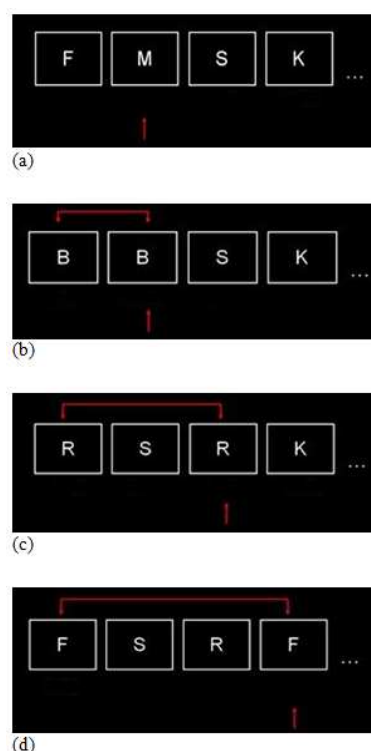
A total of 15 healthy undergraduate students (8 males, 7 females) between the ages of 19-25 were selected for the experiments. The average age of the participants was 21.6.

The experimental protocol for this study were approved by the Institutional Ethics Committee (2020-483 and 08.09.2020) of Gazi University, Ankara, Türkiye. All participants provided informed consent in accordance with human research ethics guidelines. All experiments took place in the Human Computer Interaction Laboratory, under consistent physical conditions. The laboratory is well-lit with fluorescent lighting and is sound insulated. Participants were verbally briefed on the study's purpose and the n-back tests they would be performing, as well as the required actions. Prior to starting the task, written instructions were displayed on the screen, explaining how to perform the task and its difficulty level. To ensure participants understood the tests, a trial test was conducted twice, once off-record and once after recording began. However, the trial test data included in the recording was excluded from the analysis. Experiments were only initiated after confirming the quality of the EEG signals and eye tracking calibration. Figure 2 depicts a photograph taken during the implementation of the experiment.



**Figure 2.** A shot taken from the experiment.

A standard visual n-back task set was utilized in the Inquisit Lab 6 software, offering four different difficulty levels (0-back, 1-back, 2-back, 3-back). The task becomes more challenging as the value of “n” increases. Each difficulty level was repeated three times, resulting in a total of 12 trials per participant. The order of the trials was randomized to prevent any bias caused by knowing the difficulty level in advance. Participants were instructed to press the “M” key when they saw the letter “M” in the 0-back task, and the “L” key if the letter was not “M”. In the 1-back, 2-back, and 3-back difficulty levels, they were required to press the “M” key if the letter on the screen matched the letter from 1, 2, or 3 positions back, respectively (target/match condition). If the letter did not match (non-target/non-match condition), they were instructed to press the “L” key. The task rules are provided in Figure 3. Each letter was displayed for 500 milliseconds, with a 2000 millisecond interval until the next letter appeared. Therefore, participants had a total response time of 2500 milliseconds.



**Figure 3.** N-back task levels: (a) 0-back (b) 1-back (c) 2-back (d) 3-back.

After each block, participants were asked to subjectively assess their mental workload using the NASA-TLX scale. Additionally, they were asked to prioritize the six sub-dimensions of the NASA-TLX through pairwise comparison tables. Weighted total NASA-TLX scores were calculated based on the weights derived from the 15 pairwise comparisons. The duration of the experiment with one participant was approximately 20 minutes.

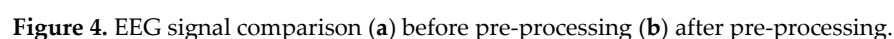
### 3.3. Data Acquisition and Pre-Processing

#### 3.3.1. EEG Data

The EEG data was captured using the EMOTIV EPOC X device, which has a resolution of 14 bits and a sampling rate of 128 Hertz (Hz). This device is equipped with 14 channels, and the sensor placements adhere to the international 10-20 system. The EMOTIV Pro software conducts real-time frequency analysis on each channel and also allows for analysis of recorded data. It visually presents the power in different frequency bands for each selected channel. The power changes of each frequency can be displayed in decibels (dB). The power densities obtained for each frequency band of the EEG signals, which are divided into sub-bands using FFT (Fast Fourier Transform), can be observed in the "Band Power" graph located at the bottom of the screen.

The frequency ranges for different bands in the software are as follows: theta band (4-8 Hz), alpha band (8-12 Hz), gamma band (25-45 Hz), low beta (12-16 Hz), and high beta (16-25 Hz). The variables analyzed in the study were aligned with these band intervals. The "FFT/Band Power" data can be transferred to an external environment, allowing access to power data for five frequency bands per channel. Unlike raw EEG data, which has a frequency of 128 Hz, the band power data is transferred with a frequency of 8 Hz. For EEG recordings with a quality measurement above 75%, the average values of band power variables provided by the software were obtained based on stimuli. The variable "AF3.Theta" represents the average theta power from the AF3 channel, while "AF3.HighBeta" represents the average high beta power from the same channel. A total of 70 EEG variables, representing power in five frequency bands (theta, alpha, low beta, high beta, gamma), were used for 14 channels. Additionally, based on previous studies indicating a positive relationship between the theta/alpha power ratio and difficulty level, 14 more variables corresponding to this ratio





### 3.3.2. Eye Tracking Data

The Tobii X2-60 device was used to collect eye tracking data. Tobii Studio software recorded three types of eye events: “fixation”, “saccade”, and “unclassified”. The software can detect fixation and saccade movements, but blink data cannot be directly obtained. As a result, a separate analysis was developed for records classified as “unclassified”. The software allows for direct measurement of the sizes of the right and left pupils. Additionally, the analysis includes variables that were calculated by taking the mean and maximum values of both eyes. Prior to adding the fixation variables to the analysis, records with fixation durations below 80 ms and above 1200 ms were excluded. This exclusion was based on the studies conducted by Borys et al. (2017a), Borys et al. (2017b), Wu et al., (2019), and Yan et al. (2019).

Blinking refers to the action of either fully or partially closing the eye, as stated by Ahlstrom and Friedman-Berg (2006). In a study focused on detecting blinking movements, the center of the pupil was used as a reference point. If the eye was open, the pupil center could be calculated, and a value of 1 was assigned to the relevant variable. Conversely, if the eye was closed, the center point could not be determined, resulting in a value of 0 for the variable, as explained by Naveed et al. (2012). Following this methodology, the Tobii Studio software provided variables for the right pupil (“PupilRight”) and left pupil (“PupilLeft”). If calculations could be performed using these variables, it was assumed that the eye was open. On the other hand, if there was no data available for either variable, it indicated that the eye was closed. To represent the open and closed states of the eye, a variable named “BlinkValid” was introduced, with values of 1 and 0 assigned, respectively.

A blink typically lasts between 100-300 ms, with blinks lasting over 500 ms considered as drowsiness (Johns, 2003). In this study, the device used had a frequency of 60 Hz, meaning that the duration of a recording line was 1/60 second. To classify situations as blinking for durations of 100 ms and above, the “BlinkValid” variable, indicating the number of records where the eye remained closed, must be at least six. Additionally, recordings identified as fixation by the Tobii Studio software should be excluded from the blink analysis, as blinks occur between two fixation movements. Using Excel Visual Basic for Applications (VBA), records meeting the specified conditions were identified and assigned a “BlinkIndex” value, representing the number of blinks. The durations of the blink recordings were calculated by multiplying the duration of a recording (1/60 second) by the number of recordings in each blink event, and these values were stored in the “BlinkDuration” variable. The number of records identified as “blink” were also counted for each stimulus, and the total durations were determined on a stimulus basis.

In total, 27 eye tracking variables were used, including the total, average, maximum, minimum, and standard deviation of fixation numbers, fixation durations, saccade numbers, saccade amplitudes, pupil sizes, and blink movements. These variables were used to create an eye-related dataset for each stimulus, consisting of the calculated values.

### 3.4. Datasets for Analysis

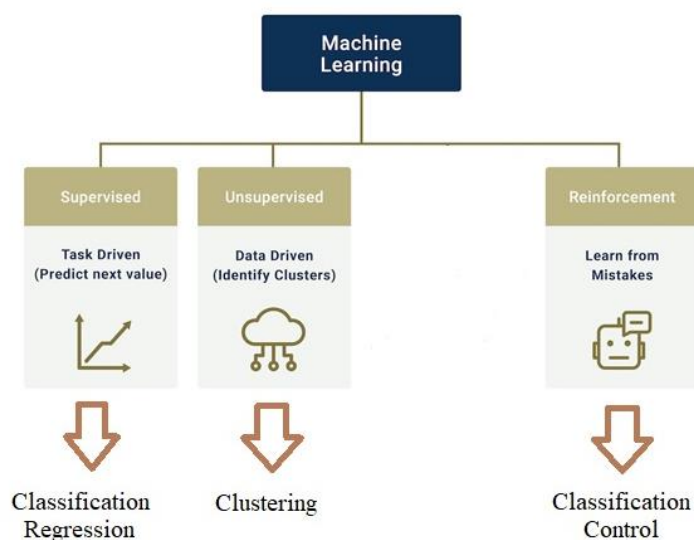
Two distinct sets of data were examined. The initial two objectives of the study focused on analyzing the first dataset, which comprised 2700 stimulus-based samples (15 subjects x 12 sessions x 15 stimuli). This dataset aimed to investigate changes in EEG and eye-related variables during tests of varying difficulty levels and to develop machine learning algorithms for predicting mental workload levels. The reason for considering the dataset on a stimulus basis is that the mental processes of visual perception, recall, matching, and reaction occur within a 2500-millisecond timeframe after each stimulus is presented on the screen. Statistical analysis was conducted using SPSS 21.0, while Python 3.8 was utilized to establish classification models.

The second dataset aimed to explore changes in band power across different brain regions under various task conditions and visualize the statistical findings. It consisted of 180 set extension files of raw EEG data obtained through EEG device software. These files were converted into set extension files using MATLAB for each participant's session and underwent data pre-processing with EEGLAB. The dimensions of the files for the 0-back task were 4800 x 14, for the 1-back task were 5120 x 14, for the 2-back task were 5440 x 14, and for the 3-back task were 5760 x 14. The number of observations

varied based on the number of stimuli in each task. As the EEG device had 14 channels, the columns in the dataset corresponded to the electrodes.

### 3.5. Brief Overview of Machine Learning

The term “Machine Learning” (ML) was initially coined in 1959 by Arthur Samuel, a pioneering figure in the fields of artificial intelligence and computer games. Samuel's definition of ML was “the field of study that gives computers the ability to learn without being explicitly programmed” (Mahesh, 2020). However, there is no universally agreed-upon definition for ML, as different authors have their own interpretations. Some authors emphasize the significance of learning from experience and the ability to apply knowledge to new situations. Others focus on the statistical nature of ML models and the importance of data for training and validation. Certain definitions highlight the distinction between supervised, unsupervised, and reinforcement learning, while others describe ML as a means of optimization or function approximation. ML involves programming computers to enhance performance by utilizing example data or past experiences. A model is established with parameters, and the learning process involves optimizing those parameters through the execution of a computer program using training data or past experiences (Alpaydin, 2020). ML is a field that focuses on the development of algorithms and techniques that allow computer programs to learn from data and enhance their performance through experience (Mitchell, 2007). By implementing suitable ML algorithms, the effectiveness of data analysis and processing can be improved, addressing practical challenges arising from the increasing volume of data across various domains (Lou et al., 2021). Figure 5 illustrates the three main types of ML: supervised learning, unsupervised learning, and reinforcement learning. Numerous comprehensive texts delve into the intricacies of ML (Alpaydin, 2020; Mohri et al., 2018; Marsland, 2015). In this section, we present the eight ML techniques employed in our study, namely “k-nearest neighbors, random forests, artificial neural network, support vector machine, gradient boosting machines (GBM), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM).”



**Figure 5.** Types of Machine Learning (Adapted from (Swamynathan, 2019)).

#### 3.5.1. K-Nearest Neighbors (KNN)

The k-nearest neighbor (k-NN) algorithm is widely used in pattern recognition or classification tasks. Its approach involves analyzing the k-nearest training samples in the problem space to ascertain the class of a new, unlabeled data point. The k-NN algorithm operates under the assumption that points with similar features are likely to belong to the same class. KNN is categorized as an instance-based learning method as it depends on the specific instances or examples in the training dataset to classify new data points (Raikwal and Saxena, 2012).

### 3.5.2. Random Forests

The random forest (RF) technique, introduced by Breiman in 2001, offers a wide range of applications including regression, classification, and variable selection, as highlighted by Genuer et al. (2010). By employing ensemble learning, RF combines multiple trees into a single algorithm, as explained by Cun et al. (2021). The prediction for a new observation is then determined by aggregating the predicted values from each individual tree in the forest. The key parameters for RF algorithms are the number of trees, the minimum number of observations in the terminal node, and the number of suitable features for splitting. In the existing literature, comprehensive mathematical explanations for RFs can be found (Breiman, 2001).

### 3.5.3. Artificial Neural Networks (ANNs)

Artificial Neural Network (ANN) is a mathematical technique that utilizes the principles of the human brain's nervous system to analyze data and make predictions (Rucco et al., 2019; Çakıt and Dağdeviren, 2023; Çakıt et al., 2015). These networks are capable of efficiently processing large and intricate datasets (Noori, 2021; Çakıt et al., 2014). The complexity of a neural network is greatly influenced by the number of hidden layers it possesses. Typically, a neural network architecture consists of an input layer to receive initial data, one or more hidden layers to process the data, and an output layer to generate the final prediction or output (Haykin, 2007). To ensure accurate predictions, it is crucial to select the appropriate design for the neural network (Gnana Sheela and Deepa, 2014). For further explanations on ANN, there are additional resources available in the literature (Zurada, 1992; Haykin, 2007; Fausett, 2006).

### 3.5.4. Support Vector Machine (SVM)

The objective of pattern classification is to develop a model that surpasses the performance of existing models on the training data. Traditional training methods in machine learning aim to find models that accurately classify each input-output pair within their respective class. However, if the model is overly tailored to the training data, it may start memorizing the data instead of learning to generalize, resulting in a decline in its ability to correctly classify future data (Carvantes et al., 2020). The Support Vector Machine (SVM) is a supervised machine learning technique that has been widely utilized in the machine learning community since the late 1990s (Brereton and Lloyd, 2010). SVM is a powerful tool that can be applied to both classification and regression problems. It is particularly well-suited for binary classification problems. In SVM, the decision boundary is determined by selecting a hyperplane that effectively separates the classes. The hyperplane is chosen in a manner that maximizes the distance or margin between the support vectors, which are the data points closest to the decision boundary.

### 3.5.5. Gradient Boosting Machines (GBM)

Friedman (2001) introduced gradient boosting machines (GBMs) as an alternative method for implementing supervised machine learning techniques. The "gbm" model consists of three primary tuning parameters: "ntree" for the maximum number of trees, "tree depth" for the maximum number of interactions between independent values, and "learning rate" (Kuhn and Johnson, 2013).

### 3.5.6. Extreme Gradient Boosting (XGBoost)

The XGBoost technique also follows the fundamental principles of the gradient boosting machine algorithm (Chen and Guestrin., 2016). Achieving the optimal model performance frequently relies on finding the most suitable combination of the different parameters required by XGBoost. As follows is how the XGBoost algorithm works: consider a dataset with  $m$  features and an  $n$  number of instances  $DS = \{(x_i, y_i): i = 1 \dots n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}\}$ . By lowering the loss and regularization goal, we should ascertain the ideal mix of functions.

$$\mathcal{L}(\phi) = \sum_i l(y_i, \phi(x_i)) + \sum_k \Omega(f_k) \quad (3)$$



where  $l$  represents the loss function,  $f_k$  represents the ( $k$ -th tree), to solve the above equation, while  $\Omega$  is a measure of the model's complexity, this prevents over-fitting of the model (Çakıt and Dağdeviren., 2022).

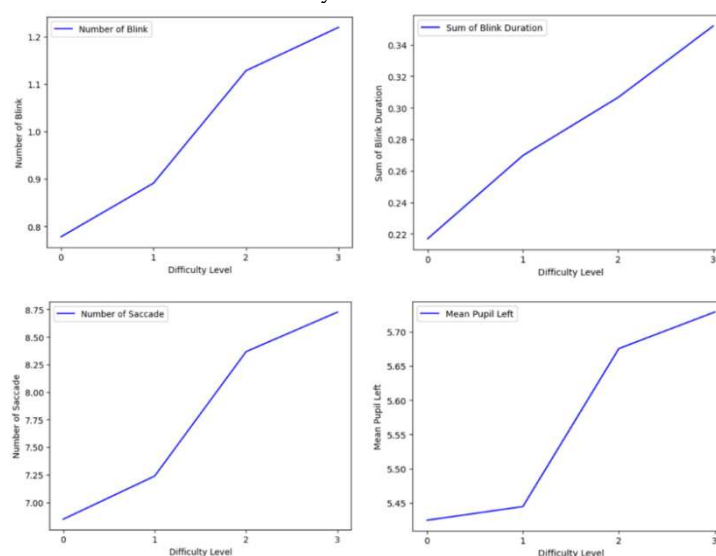
### 3.5.7. Light Gradient Boosting Machine (LightGBM)

LightGBM, a gradient-boosting decision tree method, is an open-source implementation that employs a leaf-wise approach to identify optimal splits for maximizing profits (Ke et al., 2017). To enhance prediction accuracy, various model parameters such as the number of leaves, learning rate, maximum depth, and boosting type need to be fine-tuned (Ke et al., 2017).

## 4. Results and Discussion

### 4.1. Statistical Results

The investigation aimed to determine whether there was a correlation between eye tracking variables and NASA-TLX scores and task difficulty levels. The analysis revealed that the number and total duration of blinks, the number and total duration of saccades, and the mean of the left and right pupil diameter had the highest correlation with task difficulty level ( $\rho > 0.1$ ,  $p < 0.01$ ). Similarly, the number and total duration of blinks, the mean, maximum value, and standard deviation of saccade amplitude, and the number and total duration of saccades were found to be the variables most correlated with the weighted NASA-TLX score ( $\rho > 0.1$ ,  $p < 0.01$ ). All variables, except for fixation duration, exhibited a positive correlation with mental workload. Furthermore, the Kruskal-Wallis (K-W) test demonstrated significant differences in all eye tracking variables based on task difficulty levels. Figure 6 presents graphs illustrating the changes in the mean values of blink, saccade, and pupil-related variables in relation to the difficulty level.



**Figure 6.** Change in eye tracking variables according to difficulty level.

The Spearman correlation test yielded results indicating that the EEG variables associated with theta power from the AF3, AF4, F7, F8, and FC5 channels exhibited the strongest correlation with the level of task difficulty ( $\rho > 0.3$ ,  $p < 0.01$ ). These findings demonstrate that as the task difficulty increases, theta power in the prefrontal, frontal, and frontal central regions also increase. Additionally, the correlation coefficients for variables related to theta/alpha power ratios were slightly higher in these same brain regions. Furthermore, it was observed that the level of task difficulty had a negative correlation with alpha power measured in the O1, O2, T7, and P7 channels, while it had a positive correlation with alpha power measured in the F7 and F8 channels. Specifically, as task difficulty increased, alpha power in the frontal regions increased, whereas alpha power in the temporal, parietal, and occipital regions decreased.

The decrease in low beta power is a noticeable trend across various brain regions as the task difficulty increases. A significant correlation was found between the weighted NASA-TLX total score and 75 EEG variables, with a confidence level of 99%. The EEG variables that exhibited the strongest correlation with the weighted NASA-TLX total score were theta power from AF4 and F8, as well as theta/alpha power from AF3, AF4, and F8 (correlation coefficient > 0.3,  $p < 0.01$ ). Essentially, as the perceived mental workload intensifies, there is an increase in prefrontal and frontal theta power. Conversely, as the perceived workload increases, there is a decrease in low beta power in the prefrontal, frontal, parietal, and occipital brain regions ( $\rho > 0.2$ ,  $p < 0.01$ ). The K-W test was employed to assess whether there were significant differences in EEG variables across different difficulty levels. The results revealed that there were indeed differences in difficulty levels at a confidence level of 99% for 49 EEG variables. Upon examining the rank values of these variables, it was observed that 34 out of the 49 variables exhibited a smooth increase or decrease in accordance with the change in task difficulty. This suggests that these variables may play a role in decomposing mental workload into four levels. In more challenging task conditions (2-back and 3-back), prefrontal, frontal, and central regions displayed higher theta power, while temporal, parietal, and occipital regions exhibited lower alpha power. Additionally, beta power was lower in almost all brain regions.

4.2. EEGLAB Study Results

The band power data provided by the EEG device software was supported through the processing and visual analysis of the raw EEG data. By utilizing the EEGLAB STUDY tool, changes in EEG signals under different task conditions were examined graphically. Figure 7 presents the frequency ranges that exhibit significant differences in band power based on task conditions, on a channel basis. The power spectra graphs illustrate the frequency range of 4-45 Hz. Notably, there is a substantial power change, particularly in the low-frequency theta band, in channels AF3, AF4, F7, F8, FC5, and FC6, which correspond to the prefrontal, frontal, and central regions. Additionally, a significant power change is observed in channels O1, O2, P7, P8, T7, and T8, representing the temporal, parietal, and occipital regions, particularly in the alpha band.

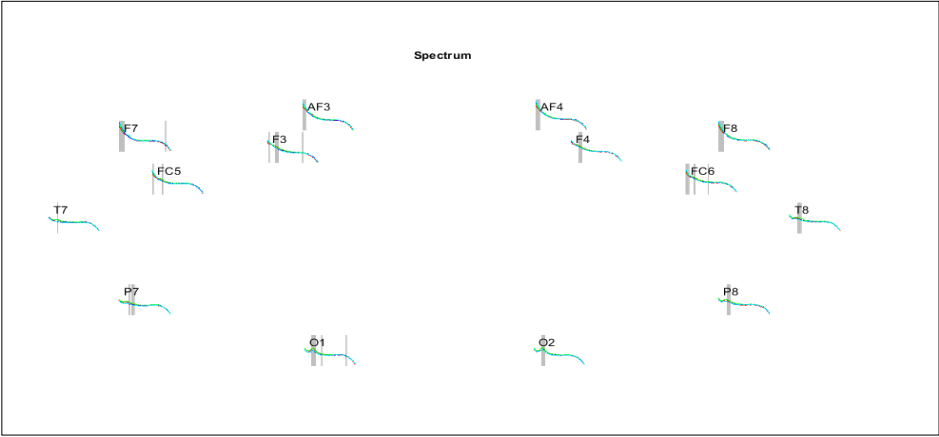


Figure 7. Significant frequencies in band power depending on task difficulty.

4.3. Classification Results

Different algorithms such as kNN, SVM, ANN, RF, GBM, LightGBM, XGBoost were utilized for modeling purposes. Table 1 presents the classification performance measures of 111-variable, 4-class classification models performed by using various algorithms. The values provided represent the average of five distinct performance outcomes achieved through 5-fold cross validation. It is evident that the Light-GBM algorithm yielded the highest accuracy performance (71.96%), while xGBoost, SVM, and GBM also demonstrated commendable results.



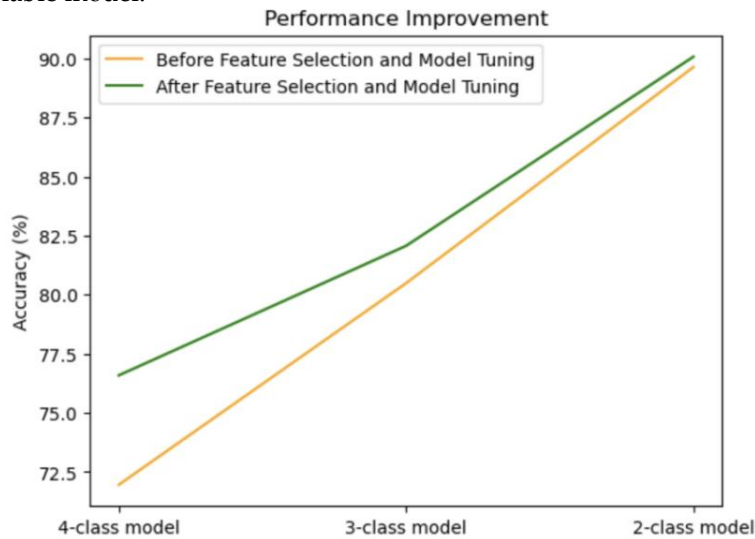
**Table 1.** Performance results of four-class models.

Model	Accuracy (%)	Kappa	Matthews Correlation Coefficient (MCC)	AUC
KNN	54.85	0.40	0.40	0.80
SVM	62.15	0.49	0.50	0.85
ANN	59.00	0.45	0.45	0.82
RF	61.44	0.49	0.49	0.85
GBM	61.93	0.49	0.49	0.85
xGBoost	70.22	0.60	0.60	0.91
Light-GBM	71.96	0.63	0.63	0.92

The performance of models using only EEG or only eye variables was investigated using the Light-GBM algorithm. The findings revealed that the model incorporating both EEG and eye variables yielded superior results, as anticipated, compared to the models that utilized them separately. Notably, the model solely based on eye variables exhibited better performance (accuracy: 65.67%) than the model solely based on EEG variables (accuracy: 56.15%).

Upon examining the classification performances of the five models obtained through cross-validation, it was observed that the model mostly made incorrect predictions for observations belonging to the 1-coded class. Additionally, the misclassification predictions were predominantly observed between classes 0-1 and 2-3. To investigate the impact of reducing the complexity of the problem by reducing the number of classes, the problem was also analyzed as a three-class and two-class scenario. Initially, the observations related to the class coded 1, which had the lowest sensitivity value, were removed from the dataset. The aim was to classify the mental workload as low-medium-high by considering only the observations for the 0-back, 2-back, and 3-back tasks. The accuracy of the three-class model, which included all variables, was determined to be 80.49%. In the two-class model, where only observations for the 0-back and 3-back tasks were considered, an accuracy performance of 89.63% was achieved. Notably, both the three-class and two-class models exhibited strong classifier characteristics, as indicated by Kappa and MCC values exceeding 0.70.

We performed dimension reduction procedure to improve both the classification performance and the comprehensibility of the model. The classification performance exhibited a noticeable shift after implementing dimension reduction and adjusting hyper-parameters, as depicted in Figure 8. It was noted that the accuracy performance experienced a more significant change when a larger number of variables were eliminated, particularly in the four-class model. Through five-fold cross validation, the average accuracy performance of the Light-GBM algorithm achieved a commendable 76.59% in the 34-variable model.



**Figure 8.** Performance change after dimension reduction and hyper-parameter tuning.

The LightGBM algorithm demonstrated superior prediction accuracy compared to other machine learning methods in the preceding section. The impact of the input parameters on the determination of the output parameter was assessed through a sensitivity analysis utilizing the LightGBM technique. The Light-GBM algorithm determined the importance ranking of the 34 features, which is presented in Figure 9. The key variables that significantly contributed to the performance of the four-class model, as determined by the LightGBM algorithm, included right and left pupil diameter, variables associated with blink, saccade and saccadic amplitude, prefrontal and frontal theta and theta/alpha, occipital alpha, and temporal, parietal, and occipital gamma power.

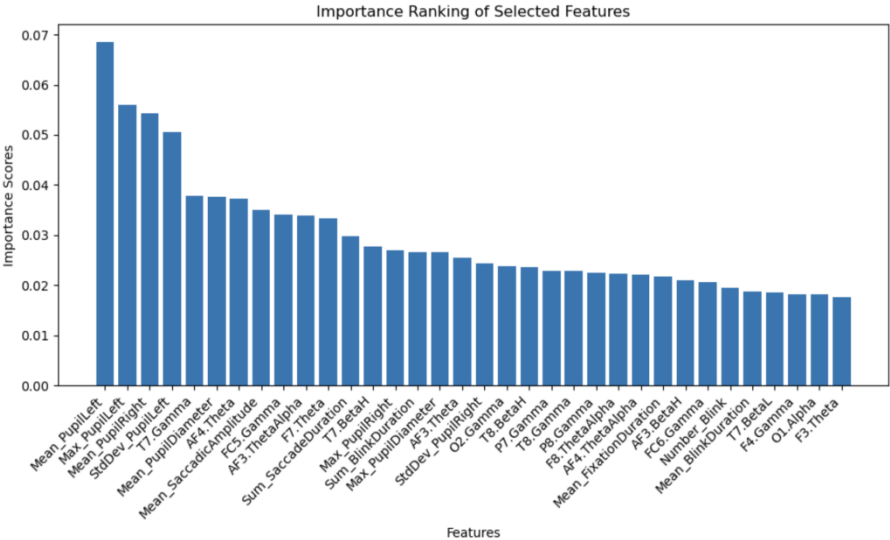


Figure 9. Importance ranking of the 34 features selected by the LightGBM algorithm.

High accuracy rates were obtained in the analysis of mental workload classification using recently developed algorithms such as LightGBM, XGBoost, and GBM. These algorithms are not commonly used in studies on this subject in the literature. The study demonstrated the extent to which performance improvement can be achieved through dimension reduction and hyper-parameter tuning. Furthermore, it was observed that reducing the number of classes leads to significantly better classification performance. As expected, the three-class and two-class models performed better. Feature selection, hyper-parameter tuning, and early stopping methods were applied only to the focal model, while the other algorithms were initially used for comparison purposes. In the four-class classification model, which utilized both EEG and eye variables as input, an accuracy performance of 76.59% was achieved after dimension reduction, hyper-parameter adjustment, and early stopping. The analysis revealed that the models made the most errors in classifying observations related to the 1-back task. Therefore, three-class models excluding the data class obtained during the 1-back task were studied. These models achieved an accuracy rate of approximately 80% after early stopping. The two-class classification model, which aimed to predict low and high mental workload, achieved an accuracy rate of approximately 90%.

4.4. Comparison with Previous Studies

A comparison summary of the nineteen applications of machine learning algorithms in the field of mental workload assessment and the current study is presented in Table 2, including author names, publication years, number of participants, number of classes, measurement tools, methods, and performance accuracy. The performance results of the current study surpass those of several other studies (Jusas and Samuvel, 2019; Li et al., 2023; Lim et al., 2018; Qu et al., 2020; Yin and Zhang, 2017), as shown in Table 2. Conversely, it has been observed that certain studies (Grimes et al., 2008; Kaczorowska et al., 2021) achieved better results than this study. Upon examining EEG studies with superior outcomes, it is evident that the device used in those studies had a higher number of channels, which differs from the current study. When eye tracking studies that yield better results are examined, higher participant numbers are noteworthy. Although the number of participants used in

this study is sufficient for the reliability of the analyses, increasing the number of participants is deemed crucial for obtaining even more dependable data to be incorporated in the analysis.

**Table 2.** Performance comparison with previous studies.

Authors, year	Number of participants	Number of classes	Measurement Tool	Method	Accuracy
Subasi (2005)	30	3	EEG	ANN	87-98%
Grimes et al. (2008)	8	2-4	EEG	NB	88- 99%
Sassaroli et al. (2008)	3	3	fNIRS	KNN (k=3)	44–72%
Wang et al. (2012)	8	3	EEG	ANN, NB	30-84%
Borys et al. (2017a)	13	2-3	EEG +Eye tracking	DT, LDA, LR, SVM, KNN	73-90%
Liu et al. (2017)	21	3	EEG+fNIRS	LDA, NB	65%
Yin and Zhang (2017)	7	2	EEG	DL	85,7%
Le et al. (2018)	5	3	fNIRS	DT, LDA, LR, SVM, KNN	81,3-95,4%
Lim et al. (2018)	48	3	EEG	SVM	69%
Jusas and Samuvel (2019)	9	4	EEG	LDA	64%
Plechawsk a-Wojcik et al. (2019)	11	3	EEG	SVM, DT, KNN, RF	70,4-91,5%
Wu et al. (2019)	39	2	Eye tracking	ANN	97%
Kaczorow ska et al. (2020)	26	2	Eye tracking	SVM, KNN, RF	97%
Qu et al. (2020)	10	3	EEG	SVM	79,8%
Kaczorow ska et al. (2021)	29	3	Eye tracking	LR, RF	97%
Pei et al. (2021)	7	3	EEG	RF	75.9-84.3%
Zhou et al. (2022)	45	7	EEG	KNN, SVM,LDA, ANN	56%
Li et al. (2023)	28	2-3	EEG +Eye tracking	SVM, RF, ANN	54,2-82,7%
Şaşmaz et al. (2023)	45	3	EEG	SVM, RF, LDA, ANN	83,4%
Current study	15	2-3-4	EEG +Eye tracking	KNN,SVM,ANN, RF,GBM,	76-90%

---

XGBoost,  
LightGBM

---

\* EEG: Electroencephalography; fNIRS: functional near-infrared spectroscopy; KNN: k-nearest neighbors; NB: naive bayes DT: decision trees; SVM: support vector machines; RF: random forest; ANN: artificial neural network; DL: deep learning; LDA: linear discriminant analysis; LR: logistic regression .

## 5. Conclusions

This study introduced the combined use of EEG and eye tracking methods in an experimental procedure alongside the NASA-TLX scale and performance-based methods. The data was collected during n-back tasks. By simultaneously recording EEG and eye data, the study successfully developed mental workload prediction models using machine learning algorithms. The model performances were found to be satisfactory, particularly in the four-class model, which achieved a high level of accuracy compared to existing literature. Furthermore, the study introduced a unique approach to deriving the blink variable. Additionally, the research provides an application example to the literature by showcasing the use of the EEGLAB STUDY program and interpreting its results. In future research, the models can be evaluated by testing them with a diverse group of participants or by using data collected under varying task conditions. Although the number of participants used in this study is sufficient for the reliability of the analyses, increasing the number of participants is considered important in order to obtain more reliable data that can be included in the analysis.

**Author Contributions:** Conceptualization, Ş.H.A., E.Ç. and M.D.; methodology, Ş.H.A.; validation, Ş.H.A. and E.Ç.; formal analysis, Ş.H.A.; investigation, Ş.H.A.; resources, Ş.H.A.; data curation, Ş.H.A.; writing—original draft preparation, Ş.H.A.; writing—review and editing, E.Ç. and M.D.; visualization, Ş.H.A. and E.Ç.; supervision, E.Ç. and M.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by Grant No. 6483 from the Scientific Research Projects Unit at Gazi University, Ankara, Türkiye.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Gazi University (2020-483 and 08.09.2020) for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This study was financially supported by Grant No. 6483 from the Scientific Research Projects Unit at Gazi University. Additionally, the Middle East Technical University Human-Computer Interaction Research and Application Laboratory generously provided the eye tracking device. Our sincere gratitude goes to these institutions and all the volunteers who actively participated in this research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Ahlstrom, U. ve Friedman-Berg, F. (2006). Using eye movement activity as a correlate of cognitive workload, *International Journal of Industrial Ergonomics*, 36, 623-636.
- Aksu, Ş. H., & Çakıt, E. (2023). Göz izleme verilerine bağlı olarak zihinsel iş yükünü sınıflandırmada makine öğrenmesi algoritmalarının kullanılması. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 38(2), 1027-1040.
- Aksu, Ş. H., Çakıt, E., & Dagdeviren, M. (2023). Investigating the Relationship Between EEG Features and N-Back Task Difficulty Levels With NASA-TLX Scores Among Undergraduate Students. *Intelligent Human Systems Integration (IHSI 2023): Integrating People and Intelligent Systems*, 69(69).
- Alpaydin, E. (2020). In: *Introduction to Machine Learning*, fourth ed. The MIT Press, London, England.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A. ve Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14, 199-208.
- Bommer, S. ve Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, 63, 7-17.

- Borys, M., Plechawska-Wojcik, M., Wawrzyk, M. ve Wesolowska, K. (2017a). Classifying Cognitive Workload Using Eye Activity and EEG Features in Arithmetic Tasks.
- Borys, M., Tokovarov, M., Wawrzyk, M., Wesolowska, K., Plechawska-Wojcik, M., Dmytruk, R. ve Kaczorowska, M. (2017b). An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks. 287-292.
- Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst*, 135(2), 230-267.
- Breiman, L. (2001) Random forests. *Machine Learning* 45, 5–32
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Cun, W., Mo, R., Chu, J., Yu, S., Zhang, H., Fan, H., Chen, Y., Wang, M., Wang, H., and Chen, C. (2021) Sitting posture detection and recognition of aircraft passengers using machine learning. *AI EDAM* 35, 284–294.
- Çakıt, E., Karwowski, W., Bozkurt, H., Ahram, T., Thompson, W., Mikusinski, P., & Lee, G. (2014). Investigating the relationship between adverse events and infrastructure development in an active war theater using soft computing techniques. *Applied Soft Computing*, 25, 204-214.
- Çakıt, E., & Dağdeviren, M. (2022). Predicting the percentage of student placement: A comparative study of machine learning algorithms. *Education and Information Technologies*, 27(1), 997-1022.
- Çakıt, E., & Dağdeviren, M. (2023). Comparative analysis of machine learning algorithms for predicting standard time in a manufacturing environment. *AI EDAM*, 37, e2.
- Çakıt, E., Durgun, B., & Cetik, O. (2015). A neural network approach for assessing the relationship between grip strength and hand anthropometry.
- Delorme, A. ve Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-12.
- Fausett, L. V. (2006). *Fundamentals of neural networks: architectures, algorithms and applications*. Pearson Education India.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Galy, E., Cariou, M. ve Melan, C. (2012). What is the relationship between mental workload factors and cognitive load types?. *International Journal of Psychophysiology*, 83(3), 269-275.
- Gao, Q., Wang, Y., Song, F., Li, Z. ve Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56, 1070 - 1085.
- Genuer, R., Poggi, J.M., and Tuleau-Malot, C. (2010) Variable selection using random forests. *Pattern Recognition Letters* 31, 2225–2236.
- Gnana Sheela, K., & Deepa, S. N. (2014). Performance analysis of modeling framework for prediction in wind farms employing artificial neural networks. *Soft Computing*, 18, 607-615.
- Grimes D., Tan D. S., Hudson S. E., Shenoy P. ve Rao R. P. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, Floransa, İtalya, 835-844.
- Guan, K., Chai, X., Zhang, Z., Li, Q. ve Niu, H. (2021). Evaluation of mental workload in working memory tasks with different information types based on EEG. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Meksika, 5682-5685.
- Harputlu Aksu, Ş., & Çakıt, E. (2022). Classifying mental workload using EEG data: A machine learning approach. *International Conference on Applied Human Factors and Ergonomics (AHFE)*.
- Hart, S. G. ve Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183.
- Haykin, S. (2007). *Neural networks: a comprehensive foundation*. Prentice-Hall, Inc..
- Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F. ve Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, 7, 935.
- Johns, M. (2003). The amplitude velocity ratio of blinks: A new method for monitoring drowsiness. *Sleep*, 26.
- Jusas, V. ve Samuvel, S.G. (2019). Classification of motor imagery using a combination of user-specific band and subject-specific band for brain-computer interface. *Applied Sciences*, 9, 4990.
- Kaczorowska, M., Wawrzyk, M. ve Plechawska-Wojcik, M. (2020). Binary Classification of Cognitive Workload Levels with Oculography Features. In: Saeed, K., Dvorský, J. (eds) *Computer Information Systems and Industrial Management*, 12133.
- Kaczorowska M., Plechawska-Wojcik, M. ve Tokovarov, M. (2021). Interpretable machine learning models for three-way classification of cognitive workload levels for eye-tracking features. *Brain Sciences*, 11(2), 210.



- Ke, Y., Qi, H., Zhang, L., Chen, S., Jiao, X., Zhou, P., Zhao, X., Wan, B. ve Ming, D. (2015). Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression, *International Journal of Psychophysiology*, 98 (2), 157-166.
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q and Liu TY (2017). LightGBM: a highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pp. 3146–3154.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- Le, A.S., Aoki, H., Murase, F. ve Ishida, K. (2018). A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy. *Frontiers in Human Neuroscience*, 12.
- Li, Q., Ng, K.K.H, Yu, S.C.M., Yiu, C.Y. ve Lyu, M. (2023). Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks. *Knowledge-Based Systems*, 260.
- Lim, W., Sourina, O. ve Wang, L. (2018). STEW: simultaneous task EEG workload dataset. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 1-1.
- Liu, Y., Ayaz, H. ve Shewokis, P. A. (2017). Multisubject “Learning” for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures. *Frontiers in Human Neuroscience*, 11(389).
- Lou, R., Lv, Z., Dang, S., Su, T., & Li, X. (2021). Application of machine learning in ocean data. *Multimedia Systems*, 1-10.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*, 9, 381-386.
- Marsland, S. 2015. *Machine learning: an algorithmic perspective*. CRC press.
- Mitchell, T. M. (2007). *Machine learning* (Vol. 1). New York: McGraw-hill.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. 2018. *Foundations of machine learning*. MIT press.
- Monod, H. ve Kapitaniak, B. (1999). *Ergonomie*. Paris: Masson Publishing.
- Naveed, S., Sikander, B. ve Khiyal, M. (2012). Eye Tracking System with Blink Detection. *Journal of Computing*, 4, 50-60.
- Noori, B. (2021). Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, 35(8), 567-588.
- Öztürk, A. (2018). Transfer and maintenance effects of n-back working memory training in interpreting students: A behavioural and optical brain imaging study, Ph.D. Dissertation, Orta Doğu Teknik Üniversitesi Enformatik Enstitüsü, Ankara.
- Parasuraman, R., Hancock, P.A. (2001). Adaptive control of mental workload. In: Hancock, P.A., Desmond, P.A. (Eds.), *Stress, Workload, and Fatigue*. Lawrence Erlbaum, Mahwah, NJ, pp. 305–333.
- Pei, Z., Wang, H., Bezerianos, A. ve Li, J. (2021). EEG-based multiclass workload identification using feature fusion and selection. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-8.
- Plechawska-Wojcik, M., Tokovarov, M., Kaczorowska, M. ve Zapala, D. (2019). A three-class classification of cognitive workload based on EEG spectral data. *Applied Sciences*, 9.
- Qu, H., Shan, Y., Liu, Y., Pang, L., Fan, Z., Zhang, J. ve Wanyan, X. (2020). Mental workload classification method based on EEG independent component features. *Applied Sciences*, 10, 3036.
- Radhakrishnan, V., Louw, T., Gonçalves, R., Torrao, G., Lenné, M. ve Merat, N. (2023). Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation Research Part F Traffic Psychology and Behaviour*, 94, 254-267.
- Raikwal, J. S., & Saxena, K. (2012). Performance evaluation of SVM and k-nearest neighbor algorithm over medical data set. *International Journal of Computer Applications*, 50(14).
- Rashid, A. ve Qureshi, I. (2015). Eliminating electroencephalogram artefacts using independent component analysis. *International Journal of Applied Mathematics Electronics and Computers*, 3 (1) , 48-52 .
- Reiner, M. ve Gelfeld, T.M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, 93(1), 38-44.
- Rubio, S., Diaz, E., Martin, J. ve Puente, J. M. (2004). Evaluation of Subjective Mental Workload : A Comparison of SWAT , NASA-TLX , and Workload Profile Methods. *Applied Psychology: An International Review*, 53(1), 61-86.
- Rucco, M., Giannini, F., Lupinetti, K., & Monti, M. (2019). A methodology for part classification with supervised machine learning. *AI EDAM*, 33(1), 100-113.
- Sareen, E., Singh, L., Varkey, B., Achary, K. ve Gupta, A. (2020). EEG dataset of individuals with intellectual and developmental disorder and healthy controls under rest and music stimuli. *Data in Brief*, 30, 105488.
- Sassaroli, A., Zheng, F., Hirshfield, L., Girouard, A., Solovey, E., Jacob, R. ve Fantini, S. (2008). Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, 31, 227-237.



- Subasi A. (2005). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications*, 28(4), 701-711.
- Swamynathan, M. (2019). Mastering machine learning with python in six steps: A practical implementation guide to predictive data analytics using python. Apress.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Şaşmaz, S., Saraoğlu, H., Kabay, S., Akdağ, G., Keskinçelik, C. ve Tosun, M. (2023). EEG-based mental workload estimation of multiple sclerosis patients. *Signal, Image and Video Processing*, 1-9.
- Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, H. ve Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, 59, 326-332.
- Wang, Z., Hope, R., Wang, Z., Ji, Q. ve Gray, W. (2012). Cross-subject workload classification with a hierarchical bayes model, *NeuroImage*, 59(1), 64-69.
- Wanyan, X., Zhuang, D., Lin, Y., Xiao, X. ve Song, J.W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *International Journal of Industrial Ergonomics*, 64, 1-7.
- Wu, Y., Liu, Z., Jia, M., Congchi, T. ve Yan, S. (2019). Using artificial neural networks for predicting mental workload in nuclear power plants based on eye tracking. *Nuclear Technology*, 206, 1-13.
- Yan, S., Wei, Y. ve Tran, C.C. (2019). Evaluation and prediction mental workload in user interface of maritime operations using eye response. *International Journal of Industrial Ergonomics*, 71, 117-127.
- Yin, Z. ve Zhang, J. (2017). Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomedical Signal Processing and Control*, 33, 30-47.
- Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X. ve Zhang, D. (2022). Cross-task cognitive workload recognition based on EEG and domain adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, 50-60.
- Zurada, J. (1992). Introduction to artificial neural systems (Vol. 8). St. Paul: West.

**Disclaimer/**