Application of Machine Learning Models to Predict e-Learning Engagement Using EEG Data

Elias Dritsas[1] [a], Maria Trigka[1] [b] and Phivos Mylonas[2] [c]
[1] *Industrial Systems Institute, Athena Research and Innovation Center, Patras, Greece*
[2] *Department of Informatics and Computer Engineering, University of West Attica, Greece*
{dritsas, trigka}@isi.gr, mylonasf@uniwa.gr

Keywords: e-Learning, EEG, Machine Learning, Classification.

Abstract: The rapid evolution of e-learning platforms necessitates the development of innovative methods to enhance learner engagement. This study leverages machine learning (ML) techniques and models to predict e-learning engagement with the aid of Electroencephalography (EEG). Various ML models, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Machine (GBM), and Neural Networks (NN), were applied to a dataset comprising EEG signals collected during e-learning sessions. Among these models, NN demonstrated the highest accuracy (90%), with precision and F1-score of 88%, a recall of 89%, and an Area Under the Curve (AUC) of 0.92 for predicting engagement levels. The results underscore the potential of EEG-based analysis combined with advanced ML techniques to optimize e-learning environments by accurately monitoring and responding to learner engagement.

1 INTRODUCTION

The advent of e-learning has significantly transformed the educational landscape, offering unprecedented opportunities for flexible and accessible learning experiences. However, this paradigm shift has brought about new challenges, especially in maintaining learner concentration on a task. Engagement is a critical factor in educational success, influencing both the retention of information and the overall learning experience. Traditional methods of assessing engagement, such as self-reports and behavioural observations, are often subjective and prone to biases. Consequently, there is a growing interest in leveraging objective physiological measures to gain deeper insights into learner engagement (Herbig et al., 2020; Mejbri et al., 2022).

In this context, EEG, a neuroimaging technique that records human brain activity, has emerged as a promising tool. EEG can provide real-time knowledge of cognitive and emotional states by capturing brainwave patterns across different frequency bands. These patterns can indicate various cognitive and mental states (Trigka et al., 2023a; Maimaiti et al., 2022), including attention, relaxation, and cognitive load, which are all relevant to engagement in learning activities. By analyzing EEG data, researchers can obtain a more accurate and dynamic understanding of how learners interact with online materials (Chrysanthakopoulou et al., 2023; Trigka et al., 2023b).

In recent years, ML has gained significant traction as a powerful approach for analyzing complex physiological data, including EEG signals. ML algorithms can identify patterns and correlations in large-scale datasets, making them well-suited for predicting engagement levels based on EEG recordings. With the increasing prevalence of online education, the motivation for this research stems from the need to enhance the effectiveness of e-learning platforms by developing methods that monitor learners' status to ensure that they remain engaged and motivated (Trigka et al., 2024).

Hence, this study explores the application of various ML models to predict engagement using EEG data in an e-learning context. The contributions are threefold:

- A meticulous description of all involved components in an EEG-based framework that exploits the Emotiv Epoc-X device, for collecting, processing and extracting accurate EEG features for e-learning engagement modelling.

[a] https://orcid.org/0000-0001-5647-2929
[b] https://orcid.org/0000-0001-7793-0407
[c] https://orcid.org/0000-0002-6916-3129

Dritsas, E., Trigka, M. and Mylonas, P.
Application of Machine Learning Models to Predict e-Learning Engagement Using EEG Data.
DOI: 10.5220/0013016000003825
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024)*, pages 323-330
ISBN: 978-989-758-718-4; ISSN: 2184-3252
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

- Spectral features analysis is applied to reveal potential differences between engaged and non-engaged states. Understanding these differences will provide insights into the neural mechanisms underlying engagement and their connection to lecture comprehension.
- Evaluation of ML models' effectiveness in predicting engagement levels from EEG data, offering insights into their relative performance. These findings highlight the potential for such predictions to inform the design of adaptive e-learning systems making them responsive to learners' cognitive and mental state in real-time, ultimately enhancing personalized learning experiences and improving educational outcomes.

The rest of this paper is organized as follows. Section 2 presents related works for the subject under consideration. In Section 3, the adopted methodology is outlined, while Section 4 discusses the experimental results. Finally, in Section 5 the conclusions are outlined.

2 RELATED WORKS

Capturing attention in educational settings has seen significant advancements with the application of EEG-based brain-computer interface (BCI) systems (Trigka et al., 2022). Numerous studies have explored various computational methods and classification approaches to effectively monitor and enhance student engagement in both traditional and e-learning environments.

Firstly, in (Nandi et al., 2021), a novel approach was presented for real-time emotion classification leveraging EEG data streams. The proposed system called the "Real-time Emotion Classification System" (RECS), employed LR trained online with the Stochastic Gradient Descent (SGD) algorithm. The research used the DEAP dataset for validation, demonstrating that RECS could classify emotional states more effectively in real-time compared to existing offline and online classifiers, including Hoeffding Tree (HT), Adaptive Random Forest (ARF), and others. The system was designed for practical applications, particularly in e-learning environments, where real-time emotional feedback can enhance learning. The authors in (Trigka et al., 2023a) introduced an ML methodology by comparing various classifiers trained and tested on EEG data, specifically focusing on band power, attention, and mediation features collected by the MindSet device. The goal was to effectively differentiate between "Confused" and "Not-Confused" individuals. Notably, the J48 model emerged as the most effective, achieving optimal performance with accuracy, precision, and recall rates of 99.9%, and an AUC of 1.

Moreover, (Al-Nafjan and Aldayel, 2022) proposed a BCI system to enhance the quality of distance education by using EEG signals to detect students' attention during online classes. The study extracted power spectral density (PSD) features from a public dataset and calculated various attention indexes using the fast Fourier transform (FFT). K-nearest neighbours (KNN), SVM, and RF models were employed to assess their performance in recognizing students' attentive states. The results showed that the RF classifier achieved the highest accuracy of 96%, indicating its effectiveness in distinguishing attention states in online learning environments.

In (Pathak and Kashyap, 2023), a novel solution that employed real-time EEG data collected from individuals wearing EEG headsets during online courses was presented. This method focuses on a convolutional neural network (CNN) model, which efficiently classifies these EEG signals with an accuracy rate of 70%. The performance highlighted the speed of processing and accuracy of the developed models, offering a promising solution to current e-learning validation challenges. Research work (Pathak and Kashyap, 2022) introduced deep learning (DL) model to address the limitations of existing ones in ML, which rely on manual feature extraction and training with limited data. Real-time e-learning data was gathered from students wearing EEG headsets during online classes. This approach overcame the challenges associated with traditional ML models and historical data. The proposed CNN model classified students into different grade levels, aiding in the creation of an automated system to monitor student learning progress and provide recommendations to enhance e-learning course materials.

Also, (Daghriri et al., 2022) presented a novel approach utilizing Probability-Based Features (PBF) derived from RF and GBM models to enhance the performance of ML classifiers for detecting confusion in students during online learning sessions. The study evaluated various classifiers, including RF, GBM, LR, SVC, and Extra Trees Classifier (ETC), achieving an accuracy, precision, recall and f1-score of 100%, with the proposed PBF approach. Additionally, the approach was validated using a separate EEG dataset, demonstrating superior performance compared to existing methodologies. The best-performing model numerically was the proposed PBF using RF and GBM features, achieving consistent top scores across all evaluation metrics.

Finally, (Aggarwal et al., 2021) evaluated learners' attention levels in MOOC (Massive Open Online Courses) environments and compared them with traditional classroom settings using brain signals. The proposed approach involved capturing EEG frequency bands from various subjects during short lectures in both e-learning and classroom environments. An SVM model was employed to classify students' mental states as either attentive or non-attentive.

3 METHODOLOGY

In this section, we analyze the dataset's characteristics in which our ML models were evaluated. Also, we describe the adopted methodology, and finally, we capture the ensemble models we experimented with, as well as the metrics for their evaluation.

3.1 Experiment and Dataset Collection

The dataset used in this study comprised EEG recordings collected from participants engaged in an e-learning activity. More specifically, 8 students, with varying levels of education (High school, Middle school, Undergraduate) were invited to watch 11 online video lectures (e.g., Quantum Physics, Statistics, String Theory, Photosynthesis, Linear Algebra, Biology, Numbers and Operations, Computational Geometry, Mythology). During these lectures, the student's EEG brain waves were recorded using a multichannel EEG system, the Emotiv Epoc X 14-channel headset with a sampling rate of 128/256 Hz (Dadebayev et al., 2022).

The dataset contained preprocessed data from the channels AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, as shown in Figure 1. The letters in the electrode names indicate the lobe locations: F (frontal), P (parietal), T (temporal), O (occipital), and C (central). Odd numbers correspond to the right hemisphere and even numbers to the left hemisphere. Also, for each of the 14 channels, the PSD was estimated in five different frequency bands (i.e., θ (4-8Hz), α (8-12Hz), low β (12 - 20Hz), high β (20-30 Hz), γ (30 - 45 Hz)) providing a quantitative measure of the brain's electrical activity. The target class captures whether a student understood the lecture or not. In total, the dataset consists of 85 features, 54370 samples in class "engaged" and 14461 samples in class "non-engaged".

3.2 Dataset Preprocessing

Effective preprocessing of EEG data is essential for accurate and reliable analysis. Raw EEG signals often contain noise and artifacts that can obscure the neural activity related to engagement. To address this, a multi-step preprocessing framework is considered to clean and prepare the data for ML analysis.

To extract PSD features, the Emotiv Epoc-X used specific digital filters that preprocessed and properly prepared the raw EEG data. Firstly, band-pass filtering was applied to retain frequencies within the range of 0.2 to 45 Hz, which are most relevant for cognitive and emotional state analysis. This step effectively removes high-frequency noise and low-frequency drifts that are not informative for the study. Also, this device includes built-in digital notch filters at 50 Hz and 60 Hz to eliminate power line interference, which could otherwise contaminate the EEG signal and affect the accuracy of the PSD calculation. The Sinc filter is used to smooth the signal and remove high-frequency noise and aliasing artefacts. The built-in digital 5th-order Sinc filter helps to refine the EEG data by providing a sharp cutoff for unwanted high-frequency components, ensuring that only the frequencies of interest are retained for PSD analysis.

Next, artefact removal was performed using Independent Component Analysis (ICA) to eliminate physiological artefacts like eye blinks, muscle movements, and heartbeats, preserving the true neural signals relevant to engagement. Following artefact removal, the EEG signals were normalized to reduce inter-subject variability. Z-score normalization is applied to each EEG channel, transforming the data to have a zero mean and a standard deviation of one. This standardization ensures that the features extracted from the EEG data are on a comparable scale, facilitating better performance of the machine learning models.

After normalization, band-pass filters were applied to extract features from specific frequency bands of the EEG signals to derive meaningful information. These filters are essential for accurate PSD calculation by keeping frequencies in the desired band. The PSD was computed using the FFT and/or other related methods. These frequency bands are known to correlate with various cognitive states, such as attention, relaxation, and cognitive load, which are critical for assessing engagement.

3.3 Spectral Features Analysis

To gain a deeper understanding of the dataset, an exploratory data analysis was conducted. Figure 2 summarizes, across all participants, the statistical measures of PSD, namely, mean, minimum, maximum and standard deviation values across different frequency bands per engagement class, allowing for easy

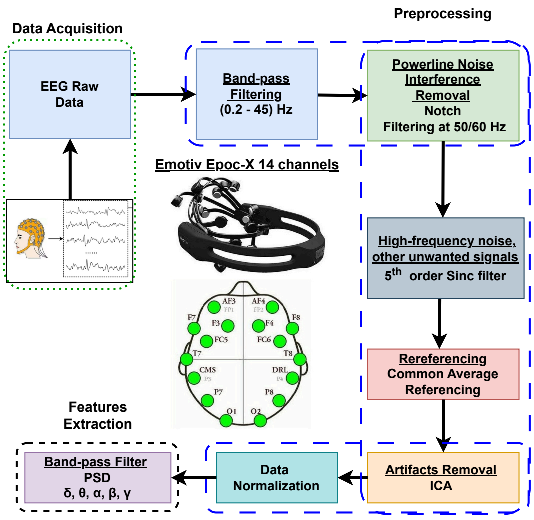

Figure 1: EEG-based processing pipeline in multi-channel Emotiv Epoc-X device.

comparison. In the following, such an analysis is presented.

In the non-engaged group, the θ band exhibited significantly higher mean (12111.71) and maximum (17170.77) PSD values compared to the engaged group, which had a mean of 1758.19 and a maximum of 7214.34. This suggested that higher θ activity in the non-engaged group might indicate increased cognitive effort without effective engagement or comprehension. Additionally, the standard deviation in the non-engaged group (2440.01) was higher than that in the engaged group (1709.85), reflecting greater variability in cognitive processing.

The α band activity in the non-engaged group showed a higher mean PSD (12022.24) and maximum PSD (17418.27) than the engaged group, which had a mean of 915.58 and a maximum of 1748.40. The increased α activity might indicate a state of relaxation or inattentiveness, which is counterproductive to effective learning. The variability in α band activity was also greater in the non-engaged group, as indicated by the standard deviation (2342.06 versus 501.08).

For the β low band, the non-engaged group displayed significantly higher mean (10466.68) and maximum (15438.50) PSD values than the engaged group (mean: 521.01, maximum: 924.49). This suggested that while the non-engaged individuals might be concentrated in cognitive processes, these processes are not effectively directed towards understanding the lesson. The higher standard deviation in the non-engaged group (2051.02 versus 241.23) indicated more unstable cognitive activity. In the β high band, the non-engaged group's mean (6925.70) and maximum (8995.99) PSD values were markedly higher than those of the engaged group (mean: 359.02, maximum: 657.95). This further supported the notion that the non-engaged group was experiencing cognitive activity that was not aligned with effective learning. The standard deviation was also higher in the non-engaged group (1141.96 versus 152.38), reflecting less consistent cognitive concentration.

Finally, γ band activity was associated with information processing and integration. The non-engaged group showed higher mean (4295.32) and maximum (5796.70) PSD values compared to the engaged group (mean: 330.62, maximum: 519.33). This suggested that while the non-engaged individuals may have been processing information, they were not effectively integrating it in a manner conducive to understanding the lesson. The higher standard deviation in the non-engaged group (759.34 versus 108.39) demonstrated greater fluctuations in cognitive processing. Overall, these results highlighted significant differences in the PSD values across various frequency bands between the engaged and non-engaged groups, pointing to differences in cognitive activity and engagement levels.

Additionally, Figure 3 depicts the correlation (or linear dependency) between PSD features including engagement class (engaged, non-engaged) using a heatmap. The heatmap visually represents the Pearson correlation coefficients (PCCs), with colour intensity indicating the strength of the correlation. Positive correlations are shown in shades of red, while negative correlations are in shades of blue. Also, this visualization helps identify which frequency bands are most closely associated with the engagement class.

It was observed that power-based features are highly linear-dependent on one another, but, their importance in improving the predictive performance of the ML models is low according to the PCCs in the blue area of the heatmap. Hence, further and extensive analysis should be conducted to understand the features' importance and apply proper selection techniques to indicate the most important ones that raise the model's performance while reducing complexity.

3.4 Machine Learning Models

The selection of appropriate ML models is critical to the successful prediction of e-learning engagement using EEG data. In this work, several ML models were investigated, each with distinct strengths and capabilities, to determine the most effective approach for this task. The models evaluated include LR, SVM, RF, GBM, and NNs. In the following, a detailed description of each model and the rationale behind their selection are provided.

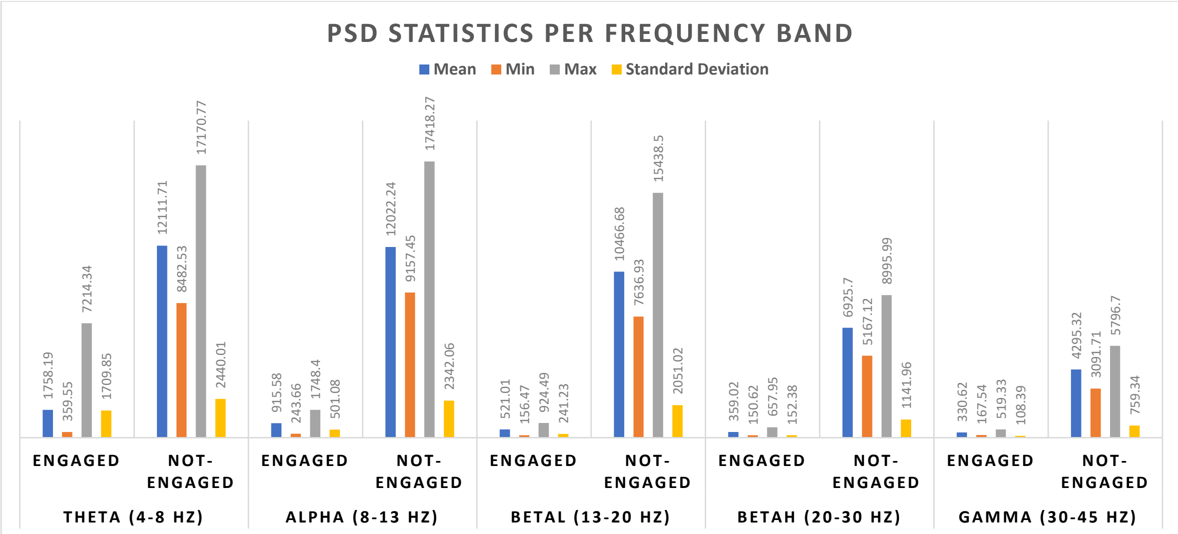

Figure 2: Statistics of PSD per frequency band and engagement state.

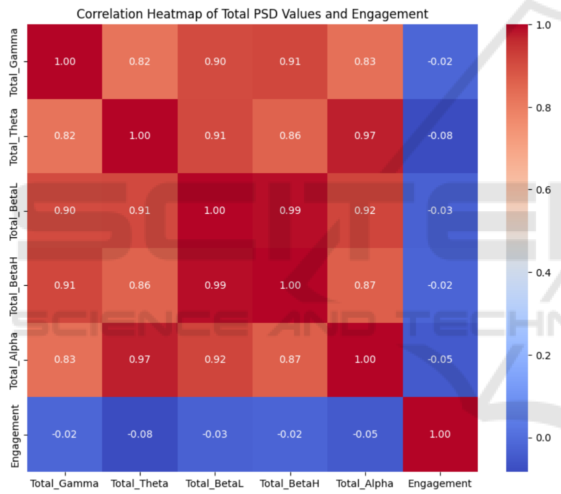

Figure 3: Correlation between PSD features and engagement class.

LR (Lu and Wang, 2024) model is based on the logistic function (a special case of sigmoid function), which maps any real-valued number to a value between 0 and 1. This function is particularly useful for binary classification tasks. The LR equation can be expressed as follows: $P(y=1 \mid X) = \sigma(z) = \frac{1}{1+e^{-z}}$, where $P(y=1 \mid X)$ is the probability that the output y is 1 (engaged) given the input features X. Also, z is defined as $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$, where $\sigma(z)$ is the logistic function, $(\beta_0, \beta_1, \beta_2, \ldots, \beta_n)$ are the coefficients of the model and $(X_1, X_2, \ldots, X_n)$ are the input features. Putting it all together, the LR model can be written as $P(y=1 \mid X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$. This equation calculates the probability that the input X belongs to class 1 "engaged". The predicted class label can be determined by applying a threshold (typically 0.5) to this probability.

SVM (Pisner and Schnyer, 2020) with a radial basis function (RBF) kernel, mainly aims to find the optimal hyperplane that separates the classes with the maximum margin. The mathematical formulation involves solving a quadratic optimization problem. The decision function for SVM is given by: $f(X) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(X_i, X) + b\right)$, where α_i are the Lagrange multipliers, y_i are the class labels, $K(X_i, X)$ is the kernel function and b is the bias term.

Our focus here is on the RBF kernel whose function K is defined as: $K(X_i, X) = \exp\left(-\gamma \|X_i - X\|^2\right)$, where γ is a parameter that determines the spread of the kernel. Summarizing these together, the decision function with the RBF kernel is $f(X) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i \exp\left(-\gamma \|X_i - X\|^2\right) + b\right)$.

RF (Genuer et al., 2020) is an ensemble learning method that combines multiple decision trees to improve the robustness and generalizability of the model. The overall prediction of the RF model is obtained by aggregating the predictions of individual trees, often by taking the mode (majority vote) in classification tasks. Here's the mathematical formulation for RFs:

1. **Individual Decision Tree Prediction -** Let $h_m(X)$ be the prediction of the m-th decision tree in the forest for input X.

2. **Random Forest Prediction -** The final prediction $H(X)$ of the RF is obtained by taking the majority vote of all M trees' predictions: $H(X) = \text{mode}\{h_1(X), h_2(X), \ldots, h_M(X)\}$.

GBM (Ayyadevara and Ayyadevara, 2018) is an ensemble learning technique that builds models sequentially, with each new model correcting errors made by the previous ones. The goal is to optimize the overall prediction by minimizing the loss function. Here's the mathematical formulation for GBMs:

1. **Model Initialization** $F_0(X) = \arg\min_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$, where L is the loss function, and y_i are the actual target values.
2. **Additive Model -** The model is built in a stage-wise manner. At each stage m, a new model $h_m(X)$ is added to minimize the loss: $F_m(X) = F_{m-1}(X) + \eta h_m(X)$, where η is the learning rate, and $h_m(X)$ is the new model fitted to the residuals of the previous model.
3. **Residual Calculation -** For each stage m, compute the residuals r_{im}: $r_{im} = -\frac{\partial L(y_i, F(X_i))}{\partial F(X_i)}\bigg|_{F(X_i)=F_{m-1}(X_i)}$
4. **Fit New Model** $h_m(X)$ to the residuals: $h_m(X) = \arg\min_h \sum_{i=1}^{n} (r_{im} - h(X_i))^2$
5. **Update the Model** with the new fitted model: $F_m(X) = F_{m-1}(X) + \eta h_m(X)$.

NNs (Gurney, 2018) are DL models that use multiple layers of neurons to capture intricate patterns in data. In a feedforward neural network, the data flows from the input layer through multiple hidden layers to the output layer. Each neuron computes a weighted sum of its inputs, applies an activation function, and passes the result to the next layer. The training process involves backpropagation to adjust the weights. The mathematical formulation for a feedforward neural network is as follows:

1. **Weighted Sum and Activation for a Single Neuron -** For each neuron in layer l, the output $a_i^{(l)}$ is computed as: $z_i^{(l)} = \sum_{j=1}^{n^{(l-1)}} w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)}, a_i^{(l)} = \sigma(z_i^{(l)})$, where $z_i^{(l)}$ is the weighted sum of inputs to the i-th neuron in layer l, $w_{ij}^{(l)}$ are the weights from neuron j in layer $l-1$ to neuron i in layer l, $b_i^{(l)}$ is the bias term for the i-th neuron in layer l, σ is the activation function (e.g., ReLU, sigmoid, tanh), and $a_j^{(l-1)}$ is the activation of the j-th neuron in the previous layer.
2. **Output Layer**, the process is similar: $z_k^{(L)} = \sum_{j=1}^{n^{(L-1)}} w_{jk}^{(L)} a_j^{(L-1)} + b_k^{(L)}, \hat{y}_k = \sigma(z_k^{(L)})$, where L is the final layer, and $\hat{y}_k$ is the predicted output.
3. **Loss Function** L measures the difference between the predicted outputs $\hat{y}$ and the true targets y. For example, using Mean Squared Error (MSE): $L = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, where N is the number of training examples.
4. **Backpropagation:** During this step, gradients of the loss with respect to the weights and biases are computed and used to update the parameters. For weights $w_{ij}^{(l)}$: $w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \frac{\partial L}{\partial w_{ij}^{(l)}}$, where η is the learning rate.

3.5 Model Training and Optimization

The ML modelS' evaluation was carried out using WEKA (WEK,), a free software suite offering a range of tools for data preprocessing, classification, regression, clustering, and visualization. The experiments were executed on an Apple MacBook Pro with a 13.3" Retina Display, equipped with an M2 chip, 16GB of RAM, and a 256GB SSD. Each model was trained on the preprocessed EEG dataset using a stratified 10-fold cross-validation to ensure robust performance evaluation. Hyperparameter tuning was performed using grid search to identify the optimal parameter settings for each model as shown in Table 1.

Table 1: Optimal Hyperparameter Tuning for Machine Learning Models.

Model	**Hyperparameter**	**Optimal Value**
Logistic Regression	Regularization (C)	1.0
SVM	Kernel Type	RBF
	Kernel Coefficient (γ)	0.01
	Regularization (C)	10
RF	Number of Trees	100
	Maximum Depth	None (unlimited)
	Minimum Samples Split	2
GBM	Number of Estimators	200
	Learning Rate	0.1
	Maximum Depth	3
NN	Number of Layers	3
	Neurons per Layer	[64, 128, 64]
	Activation Function	ReLU
	Learning Rate	0.001
	Batch Size	32
	Epochs	150

3.6 Evaluation Metrics

Several metrics were used to evaluate the performance of the ML models, accuracy, precision, recall, F1-score, and AUC (Naidu et al., 2023). These metrics provide insights into various aspects of models' performance, ensuring a robust assessment of their predictive capabilities. It should be noted that the ultimate value in each metric was derived by averaging the outcomes of both classes from all folds. The definition of these metrics is based on the confusion matrix consisting of the elements true-positive (Tp), true-negative (Tn), false-positive (Fp) and false-negative (Fn). Below is a brief description of each metric:

- **Accuracy** is the proportion of correctly predicted instances out of the total instances. It is a straightforward metric indicating the overall correctness of the model. Accuracy $= \frac{\text{Tp+Tn}}{\text{Total Instances}}$.

- **Precision** is the ratio of correctly predicted positive observations to the total predicted positives. It reflects the accuracy of the positive predictions made by the model. Precision $= \frac{\text{Tp}}{\text{Tp+Fp}}$.
- **Recall** is the ratio of correctly predicted positive observations to all the observations in the actual class. It measures the model's ability to capture all relevant instances. Recall $= \frac{\text{Tp}}{\text{Tp+Fn}}$.
- **F1-score** is the harmonic mean of Precision and Recall. It provides a single metric that balances the trade-off between Precision and Recall, especially useful when the class distribution is imbalanced: F1-Score $= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}}$.
- **AUC** measures the ability of the model to distinguish between classes. It represents the degree of separability achieved by the model. An AUC of 1 indicates a perfect model, while an AUC of 0.5 suggests no discriminative power.

These metrics provided a comprehensive view of the model performance, enabling the identification of the most effective model for predicting e-learning engagement based on EEG data.

4 RESULTS AND DISCUSSION

The performance results are summarized in Table 2. The NN model outperformed all other models, achieving the highest scores across all evaluation metrics. The GBM also showed strong performance, indicating its effectiveness in handling complex, nonlinear relationships in the EEG data.

Table 2: Performance of Machine Learning Models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
LR	0.78	0.75	0.76	0.75	0.8
SVM	0.82	0.80	0.81	0.80	0.84
RF	0.85	0.83	0.84	0.83	0.87
GBM	0.87	0.85	0.86	0.85	0.89
NN	0.90	0.88	0.89	0.88	0.92

The NN model achieved an accuracy of 90%, precision of 88%, recall of 89%, F1-score of 88%, and an AUC of 0.92. These results demonstrated the model's superior ability to accurately predict learner engagement. The high AUC value indicated excellent discrimination between engaged and not-engaged states. The GBM also performed well, with an accuracy of 87%, precision of 85%, recall of 86%, F1-score of 85%, and an AUC of 0.89. The ensemble nature of this model allows it to capture complex patterns and interactions in the data, contributing to its robust performance. RF, while slightly less accurate than GBM, still showed strong performance with an accuracy of 85%, precision of 83%, recall of 84%, F1-score of 83%, and an AUC of 0.87. Its ability to handle high-dimensional data and reduce overfitting by averaging multiple trees makes it a reliable choice for EEG data analysis.

The RBF-based SVM model achieved an accuracy of 82%, precision of 80%, recall of 81%, F1-score of 80%, and an AUC of 0.84. Its performance demonstrated the effectiveness of kernel methods in capturing non-linear relationships in the EEG data. LR, despite being the simplest model, performed reasonably well with an accuracy of 78%, precision of 75%, recall of 76%, F1-score of 75%, and an AUC of 0.8. This indicated that even linear models can provide valuable insights when applied to EEG data.

The results of this study are expected to have significant implications for the design and implementation of e-learning systems. By integrating EEG-based engagement prediction models, e-learning platforms can adapt in real-time to the learners' cognitive and emotional states. This personalization can enhance learner engagement, improve learning outcomes, and reduce dropout rates.

5 CONCLUSIONS

This study demonstrated the efficacy of various ML models in predicting e-learning engagement using EEG data, with NN emerging as the most effective model. The experimental results underscored the superiority of NN, which achieved the highest metrics across all evaluation parameters; an accuracy of 90%, a precision and F1-score of 88%, a recall equal to 89%, and an AUC of 0.92. These results indicated that NN provides a robust model for accurately predicting learner engagement.

The findings reveal that the most significant EEG features contributing to engagement predictions were the power spectral densities in the alpha and beta frequency bands. These bands are well-documented in literature for their associations with relaxation, attention, and cognitive processing, respectively. The implications of this research are substantial for the design and implementation of adaptive e-learning systems. By incorporating EEG-based engagement prediction models, e-learning platforms can dynamically adapt to the cognitive and emotional states of learners, thereby enhancing engagement, improving learning outcomes, and potentially reducing dropout rates.

In future research, we aim to expand the analysis to datasets that include a broader range of physiological signals, enhancing the robustness and generalizability of the engagement prediction models. Additionally, exploring the real-time implementation of these models within e-learning platforms will be a crucial step towards creating more personalized and responsive learning environments.

REFERENCES

Weka. https://www.weka.io/. (accessed on 17 July 2024).

Aggarwal, S., Lamba, M., Verma, K., Khuttan, S., and Gautam, H. (2021). A preliminary investigation for assessing attention levels for massive online open courses learning environment using eeg signals: An experimental study. *Human Behavior and emerging technologies*, 3(5):933–941.

Al-Nafjan, A. and Aldayel, M. (2022). Predict students' attention in online learning using eeg data. *Sustainability*, 14(11):6553.

Ayyadevara, V. K. and Ayyadevara, V. K. (2018). Gradient boosting machine. *Pro machine learning algorithms: A hands-on approach to implementing algorithms in python and R*, pages 117–134.

Chrysanthakopoulou, A., Dritsas, E., Trigka, M., and Mylonas, P. (2023). An eeg-based application for real-time mental state recognition in adaptive e-learning environment. In *2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP) 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP 2023)*, pages 1–6. IEEE.

Dadebayev, D., Goh, W. W., and Tan, E. X. (2022). Eeg-based emotion recognition: Review of commercial eeg devices and machine learning techniques. *Journal of King Saud University-Computer and Information Sciences*, 34(7):4385–4401.

Daghriri, T., Rustam, F., Aljedaani, W., Bashiri, A. H., and Ashraf, I. (2022). Electroencephalogram signals for detecting confused students in online education platforms with probability-based features. *Electronics*, 11(18):2855.

Genuer, R., Poggi, J.-M., Genuer, R., and Poggi, J.-M. (2020). *Random forests*. Springer.

Gurney, K. (2018). *An introduction to neural networks*. CRC press.

Herbig, N., Düwel, T., Helali, M., Eckhart, L., Schuck, P., Choudhury, S., and Krüger, A. (2020). Investigating multi-modal measures for cognitive load detection in e-learning. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, pages 88–97.

Lu, W. and Wang, Y. (2024). Logistic regression. In *Textbook of Medical Statistics: For Medical Students*, pages 181–189. Springer.

Maimaiti, B., Meng, H., Lv, Y., Qiu, J., Zhu, Z., Xie, Y., Li, Y., Zhao, W., Liu, J., Li, M., et al. (2022). An overview of eeg-based machine learning methods in seizure prediction and opportunities for neurologists in this field. *Neuroscience*, 481:197–218.

Mejbri, N., Essalmi, F., Jemni, M., and Alyoubi, B. A. (2022). Trends in the use of affective computing in e-learning environments. *Education and Information Technologies*, pages 1–23.

Naidu, G., Zuva, T., and Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. In *Computer Science On-line Conference*, pages 15–25. Springer.

Nandi, A., Xhafa, F., Subirats, L., and Fort, S. (2021). Real-time emotion classification using eeg data stream in e-learning contexts. *Sensors*, 21(5):1589.

Pathak, D. and Kashyap, R. (2022). Electroencephalogram-based deep learning framework for the proposed solution of e-learning challenges and limitations. *International Journal of Intelligent Information and Database Systems*, 15(3):295–310.

Pathak, D. and Kashyap, R. (2023). Evaluating e-learning engagement through eeg signal analysis with convolutional neural networks. In *International Conference on Computer & Communication Technologies*, pages 225–239. Springer.

Pisner, D. A. and Schnyer, D. M. (2020). Support vector machine. In *Machine learning*, pages 101–121. Elsevier.

Trigka, M., Dritsas, E., and Fidas, C. (2022). A survey on signal processing methods for eeg-based brain computer interface systems. In *Proceedings of the 26th Pan-Hellenic Conference on Informatics*, pages 213–218.

Trigka, M., Dritsas, E., and Mylonas, P. (2023a). Mental confusion prediction in e-learning contexts with eeg and machine learning. In *Novel & Intelligent Digital Systems Conferences*, pages 195–200. Springer.

Trigka, M., Dritsas, E., and Mylonas, P. (2024). New perspectives in e-learning: Eeg-based modelling of human cognition individual differences. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 290–299. Springer.

Trigka, M., Papadoulis, G., Dritsas, E., and Fidas, C. (2023b). Influences of cognitive styles on eeg-based activity: An empirical study on visual content comprehension. In *IFIP Conference on Human-Computer Interaction*, pages 496–500. Springer.