

Article

Electroencephalogram Signals for Detecting Confused Students in Online Education Platforms with Probability-Based Features

Talal Daghriri ^{1,*}, Furqan Rustam ^{2,t}, Wajdi Aljedaani ³, Abdullateef H. Bashiri ⁴ and Imran Ashraf ^{5,*}¹ Department of Industrial Engineering, Jazan University, Jazan 82822, Saudi Arabia² School of Computer Science, University College Dublin, D04 V1W8 Dublin, Ireland³ Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA⁴ Department of Mechanical Engineering, Jazan University, Jazan 82822, Saudi Arabia⁵ Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea

* Correspondence: tdaghriri@knights.ucf.edu (T.D.); imranashraf@ynu.ac.kr (I.A.)

† These authors contributed equally to this work.

Abstract: Online education has emerged as an important educational medium during the COVID-19 pandemic. Despite the advantages of online education, it lacks face-to-face settings, which makes it very difficult to analyze the students' level of interaction, understanding, and confusion. This study makes use of electroencephalogram (EEG) data for student confusion detection for the massive open online course (MOOC) platform. Existing approaches for confusion detection predominantly focus on model optimization and feature engineering is not very well studied. This study proposes a novel engineering approach that uses probability-based features (PBF) for increasing the efficacy of machine learning models. The PBF approach utilizes the probabilistic output from the random forest (RF) and gradient-boosting machine (GBM) as a feature vector to train machine learning models. Extensive experiments are performed by using the original features and PBF approach through several machine learning models with EEG data. Experimental results suggest that by using the PBF approach on EEG data, a 100% accuracy can be obtained for detecting confused students. K-fold cross-validation and performance comparison with existing approaches further corroborates the results.

Keywords: confused student detection; MOOC platform; electroencephalogram; feature engineering



check for updates

Citation: Daghriri, T.; Rustam, F.; Aljedaani, W.; Bashiri, A.H.; Ashraf, I. Electroencephalogram Signals for Detecting Confused Students in Online Education Platforms with Probability-Based Features. *Electronics* **2022**, *11*, 2855. <https://doi.org/10.3390/electronics11182855>

Academic Editors: Cristian Marian Mihaescu, Vicente Julian and Savvas A. Chatzichristofis

Received: 9 August 2022

Accepted: 7 September 2022

Published: 9 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of online education platforms, unlike traditional classroom settings, is growing rapidly, and the number of students attending online courses through the massive open online course (MOOC) platform is increasing day by day. MOOC is a large-scale non-campus setup that is extensively used for online education [1]. Despite the advantages of 24/7 access, wide reach, and access around the world, this mode of education has several drawbacks. For example, students are not fully attentive in the online education program, and there is no face-to-face interaction with the instructor. Online education is very different from the traditional classroom or face-to-face education [2]. In a face-to-face classroom setting, a teacher can assess the level of student understanding by verbal questions, body language, facial expressions, etc., and help the students to increase their understanding. However, online learning lacks this feature, which affects the performance of the students. However, there are some guidelines for the stakeholders in e-learning for a better online course design. These days, MOOC offers interactive sessions where the teacher and the students can talk about a variety of topics, and student reviews are also collected for feedback [3]. There is no doubt that students can benefit greatly from online education, especially during the COVID-19 pandemic. The majority of the students' learning activities were shifted to the online mode of education, but online education still has deficiencies. Students may feel confused while watching the MOOC videos [4]. Predominantly, existing studies make use of video data, student physical attributes, and

textual data for student confusion detection. On the contrary, this study leverages a novel technology, electroencephalogram (EEG), for student confusion detection in the MOOC platform.

EEG indicates brain activity, and EEG analysis is an important area of research in the field of artificial intelligence [5]. It can help medical professionals in making intelligent diagnoses for conditions including epilepsy and Alzheimer's [6,7]. The EEG signal is used to determine how the voltage between brain neurons fluctuate. The frequency and amplitude of electrical activity generated by the brain are both estimated by EEG [8]. With any high-resolution imaging technology, millisecond-range temporal resolution is not possible, but EEG data can provide this resolution [9]. The relative intensity of activity within each frequency band has been linked to brain states like focused attentional processing, engagement, and frustration, which are important for and predictive of learning. Rhythmic fluctuations in the EEG signal occur within a set of specific frequency bands. The information captured by the EEG indicates the electrical activity of different parts of the brain. It contains the functional state information of the human brain and is represented by five types of waves. These waves are distinguished by frequency bands which are characterized by the range of frequency. The delta band is from 0 to 4 Hz, theta band falls between 3.5 to 7.5 Hz, the alpha band is between 7.5 and 13 Hz, the beta band ranges from 13 to 26 Hz, and the gamma band is between 26 and 70 Hz [10]. Each wave shows a different state of mind. For example, delta waves show the deep sleep state while theta waves indicate the meditation state where the body is asleep while the mind is awake. Alpha waves indicate the case of dreaming and relaxation, beta waves show the waking state of mind with large attention while the gamma waves are related to the state of the brain when it is in a decision-making mode. A raised activity of any of the waves can be used to diagnose the abnormal state of mind [11].

1.1. Motivation

Confusion detection is an important research area in EEG because confusion can be analyzed by using the EEG data [12]. For example, the waveform on EEG is different for fear and confusion. Although these waveforms are not obvious, they are important in the field of confusion detection and brain-computer interface (BCI) research [13]. Online learning emerged as an important application during the COVID-19 pandemic when regular education at institutes was closed. Despite being successful and attractive, online education has several drawbacks and gaps. This study focuses on one of the major gaps between online and in-class learning and aims at detecting students' confusion, which they experience during online classes [14]. Contrary to in-class instruction, where a teacher may determine if the student understands the content by verbal questioning or their body language (such as a furrowed brow, head-scratching, etc.), online learning makes it more difficult to get immediate feedback from the students [15]. Several additional factors can also cause confusion like the delivery tool, environment, content of the lecture and tutor method, etc. For an efficient online education system, it is necessary to have an automatic system that can detect confusion levels among students in the online education system.

1.2. Contributions

In several fields, including text classification, image processing, and sentiment analysis, machine learning algorithms have demonstrated their strength and superiority over conventional approaches. According to a number of recent studies, machine learning techniques outperform traditional methods for EEG data classification tasks. Therefore, this study leverage machine learning for the detection of confusion by using the EEG data and makes the following contributions.

- The novel use of EEG data is made for student confusion detection in the MOOC platform. For this purpose, EEG data collected from students when interacting with online courses are used for experiments.

- For increasing the performance of the machine learning models, an intuitive feature engineering approach, probability-based features (PBF), is designed. The PBF takes the class probabilities from the random forest (RF), and gradient boosting machine (GBM) and combines them to make the feature space.
- Extensive experiments are performed by using both the original feature set and the proposed PBF with many machine learning models including logistic regression (LR), RF, GBM, support vector classifier (SVC), and extra tree classifier (ETC). In addition, two state-of-the-art deep learning models are also employed, including convolutional neural network (CNN) and long short-term memory (LSTM). Performance analysis is carried out with existing methods and k-fold cross-validation is also performed.

The rest of the paper is divided into five sections. Section 2, contains the related work. Section 3 describes the proposed methodology, models, and data used for experiments. Section 4 contains results and discussion. The conclusion is summarized in Section 5.

2. Related Work

Many researchers used machine learning techniques to analyze the EEG data for various purposes including epilepsy detection, Alzheimer's detection, driver drowsiness detection, emotion detection, etc. There is a consensus that confusion in students during MOOC can be consistently detected through visual inspection of EEG waves from patterns. To access the cognitive processing and mental state by using EEG signals, several studies were conducted in this regard.

Keeping in view the importance of online education systems, especially in the context of COVID-19 [16], a large body of literature can be found regarding online education. For example, Kumar et al. [17] worked on the improvement of the quality and effectiveness of the online education system. They used 32 supervised learning algorithms with various parameter settings to detect whether the student is confused or not confused while watching MOOC videos. Results of the study show that bagging with RF gives the accuracy value of 61.89% for the pre-defined confusion level detection and achieved 66.6% for user-defined confusion level detection.

Zhaoheng Ni et al. [18] proposed a deep learning-based system to classify students' confusion in watching online course videos using EEG data. The study used the recurrent neural network (RNN), long short-term memory (LSTM), and bi-directional LSTM deep learning algorithms. The proposed bi-directional LSTM shows strong robustness that was evaluated by cross-validation. The proposed bi-directional LSTM achieved an accuracy of 73.3% in predicting whether a student is confused or not.

Haohan Wang et al. [19] proposed a system to improve MOOC feedback interaction by using EEG signals. They used two ways to label mental states. According to the experimental design, one way is a pre-defined confusion level, and the second way is a user-defined confusion level. For the pre-defined confusion level, student-specific classifiers achieved an accuracy of 67% and student-independent classifiers achieved an accuracy value of 57%. For the user-defined confusion level, student-specific and student-independent classifiers achieved the accuracy of 56% and 51%, respectively.

In another study, Haohan Wang et al. [20] proposed a deep learning system for improving the prediction accuracy for healthcare applications. To check the efficacy of their system they performed experiments across CT-scan, MRA, and EEG brainwave. CT-scan used for lung adenocarcinoma prediction. MRA is used for segmentation on the right ventricle (RV) of the heart, and EEG brainwave data is used for the prediction of student confusion in MOOC. They used SVM, KNN, CNN, DBN, RNN-LSTM, BiLSTM, and proposed CF-BiLSTM. The proposed CF-BiLSTM achieved the accuracy value of 75% for student confusion status prediction. Harsh Kumar et al. [21] used the EEG signals collected by using a Neuro Sky Mindwave headset for the estimation of mental confusion levels. They used machine learning algorithms such as KNN, NB, XGBoost, and RF. The RF achieved an accuracy value of 96.48% for this brain-computer interface (BCI). Edla et al. [22] proposed a system for human mental state analysis by using EEG data. They acquired the real-time

EEG data of 40 subjects (7 female, 33 male). Statistical measures such as standard deviation, mean, minimum, and maximum amplitudes are used to derive the features of the EEG data. Analysis of the results shows that the proposed random forest achieved an accuracy of 75%. Table 1, shows the summary of related work on the confused student prediction.

Table 1. Summary of the systematic analysis studies in related work.

Study	Year	Dataset	Classifiers	Achieved Accuracy
[17]	2019	Same dataset	32 algorithms	66.6% ETC
[13]	2020	Same dataset	KNN, CNN, NN, LSTM, BiLSTM	88% BiLSTM
[18]	2017	Same dataset	RNN,LSTM, Bi directional LSTM	Bi directional LSTM
[19]	2013	Same dataset	Student specific and student independent	67% student independent
[20]	2019	Same dataset	SVM, KNN, CNN, DBN, RNN-LSTM, BiLSTM, CF-BiLSTM	75% CF-BiLSTM
[21]	2019	Self-collected using neurosky mind wave hand wave	KNN, NB, XGBoost, RF	96.48% RF
[22]	2018	Self-collected using neurosky mind wave hand wave	RF, SVM	75%

Yeo et al. [23] used the EEG signals for the detection of drowsiness in car drivers. They extract the features from four EEG frequency bands. The SVM achieved an accuracy value of 99.3%. To extract the features from the EEG, Sun et al. [24] used an unsupervised learning technique. They observed that the efficiency of EEG classification suffers when supervised learning techniques are used. Deep belief networks (DBN) were used by Hajinoroozi et al. [25] to predict the driver’s cognitive states by extracting features and dimensionality reduction from EEG data. Results of their study reveal that DBN-C is a potential technique to extract features. Petrosian et al. [26] proposed a deep learning system that can detect the symptoms of Alzheimer’s disease from the long-term EEG signals. Gen et al. [13] proposed a maximum marginal approach for EEG signal processing for emotion detection. The proposed approach selects the least similar segments between EEG signals as features that can differentiate between EEG signals caused by different emotions. To find the features, the method defines a signal similarity described as the distance between two EEG signals. Wavelet transform make use of a wavelet to calculate the frequency domain of EEG. They used KNN, CNN, NN, LSTM, and BiLSTM models in their study and achieved the accuracy value of 88% by BiLSTM.

3. Materials and Methods

This study performed experiments for student confusion prediction by using EEG signals and a machine learning approach. We implement this approach on a Corei7 11th generation machine with Windows operating system. We used Sci-kit learn, TensorFlow, and Keras framework for the implementation of the proposed approach by using the Python language. The proposed approach’s architecture is shown in Figure 1.

The proposed approach consists of several steps. First, we acquired the dataset from Kaggle named “confused student EEG brainwave data”. This dataset consists of two target classes. After acquiring the dataset, we perform data engineering techniques to improve the accuracy of learning models. We used a probability-based feature engineering technique to generate new features from original features. We pass the original dataset to the machine learning models. These models generate probabilities against each sample of the dataset. Each model predicts two probabilities; one for a confused target and one for a non-confused target class. Data splitting is the next step after feature engineering and we split the dataset with a 0.8 to 0.2 ratio, where 80% data is used for training whereas 20% is used for testing of models. We used several machine learning and deep learning models for student confusion

prediction, and in the end, we evaluate the performance of learning models in terms of accuracy, precision, recall, and F1 score.

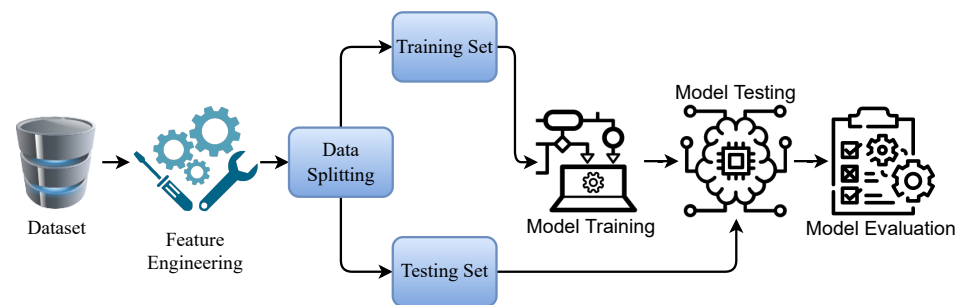


Figure 1. Flow of the adopted methodology for confusion detection.

3.1. Dataset

The dataset is acquired from Kaggle [19,27]. Ten students participated in the data collection while watching MOOC video clips. Ten videos were watched by each student which generate 100 data points consisting of 12,000+ rows. According to this count, each data point consists of approximately 120+ rows (100 data points \times 120+ rows = 12,000+ rows). The students wore a single-channel wireless MindSet that measured activity over the frontal lobe. Three electrodes, one placed on the forehead and two contacted with an ear, record the brain's spontaneous electrical activity over a period of time. This electrical activity generated a specific pattern to show if the student is confused, which is later labeled by the student himself to verify whether he was confused or not. Consequently, an EEG signal that is verified by a student who was confused during the lecture used for modeled training can be generated through an automatic system, which in turn can predict EEG data showing whether the student is confused or not. The electrodes collect the following signal streams by using NeuroSky's API:

- the raw EEG signals, sampled at 512 Hz;
- an indicator of signal quality reported at 1 Hz;
- MindSet's proprietary "attention" and "meditation" signals are said to measure the user's level of mental focus and calmness, reported at 1 Hz; and
- a power spectrum, reported at 8 Hz, clustered into the standard named frequency bands, i.e., delta (1–3 Hz), theta (4–7 Hz), alpha (8–11 Hz), beta (12–29 Hz), and gamma (30–100 Hz).

For data annotation, the participants confirmed their state of mind after each session of watching online videos. The subjects rate their confusion level on a scale of 1–7 from low to high. For binary classification, these labels are quantized into confused or not confused. Further details on the data collection and quantization can be found in [18].

The dataset has 17 columns and 12,811 samples, as shown in Table 2. The dataset features consist of two categories, one extracted by using an EEG signal when they were watching a MOOC video clip. The second type of features are demographic features that contain demographic information about students, such as gender, age, and language as shown in Table 3.

Table 2. Dataset details for training and testing.

Data Splitting		
Dataset	Confused	Not Confused
Training set	5255	4993
Testing Set	1251	1312
Total	6567	6244
Dataset Stats		
Total Features	17	
Total Samples	12,811	

Table 3. Dataset features description.

Feature Type	Feature Name	Description
EEG Features (10 students EEG recorded data)	Attention	A proprietary measure of mental focus
	Mediation	A proprietary measure of calmness
	Raw	EEG signal
	Delta	1–3 Hz of the power spectrum
	Theta	4–7 Hz of the power spectrum
	Alpha1	Lower 8–11 Hz of the power spectrum
	Alpha2	Higher 8–11 Hz of the power spectrum
	Beta1	Lower 12–29 Hz of the power spectrum
	Beta2	Higher 12–29 Hz of power spectrum)
	Gamma1	Lower 30–100 Hz of the power spectrum
	Gamma2	Higher 30–100 Hz of the power spectrum
	User-definelabel	Is the student confused or not confused (actual label)
Demographic Features (Each student’s demographic information)	Age	Age of students
	Ethnicity	Chines
		English
		Bengali
Gender	Student gender (Male, Female)	

The histogram distribution of these features is provided in Figure 2. This figure contains the features’ value ranges and the number of sample counts against the values. As in the attention feature, most of the samples have values ranging from 25 to 75. Similarly in the raw features, most of the samples have values ranging from –500 to 500. Because not all features are equally important, a feature correlation analysis is carried out to understand and analyze the correlation level. The sample of the dataset is present in Table 4.

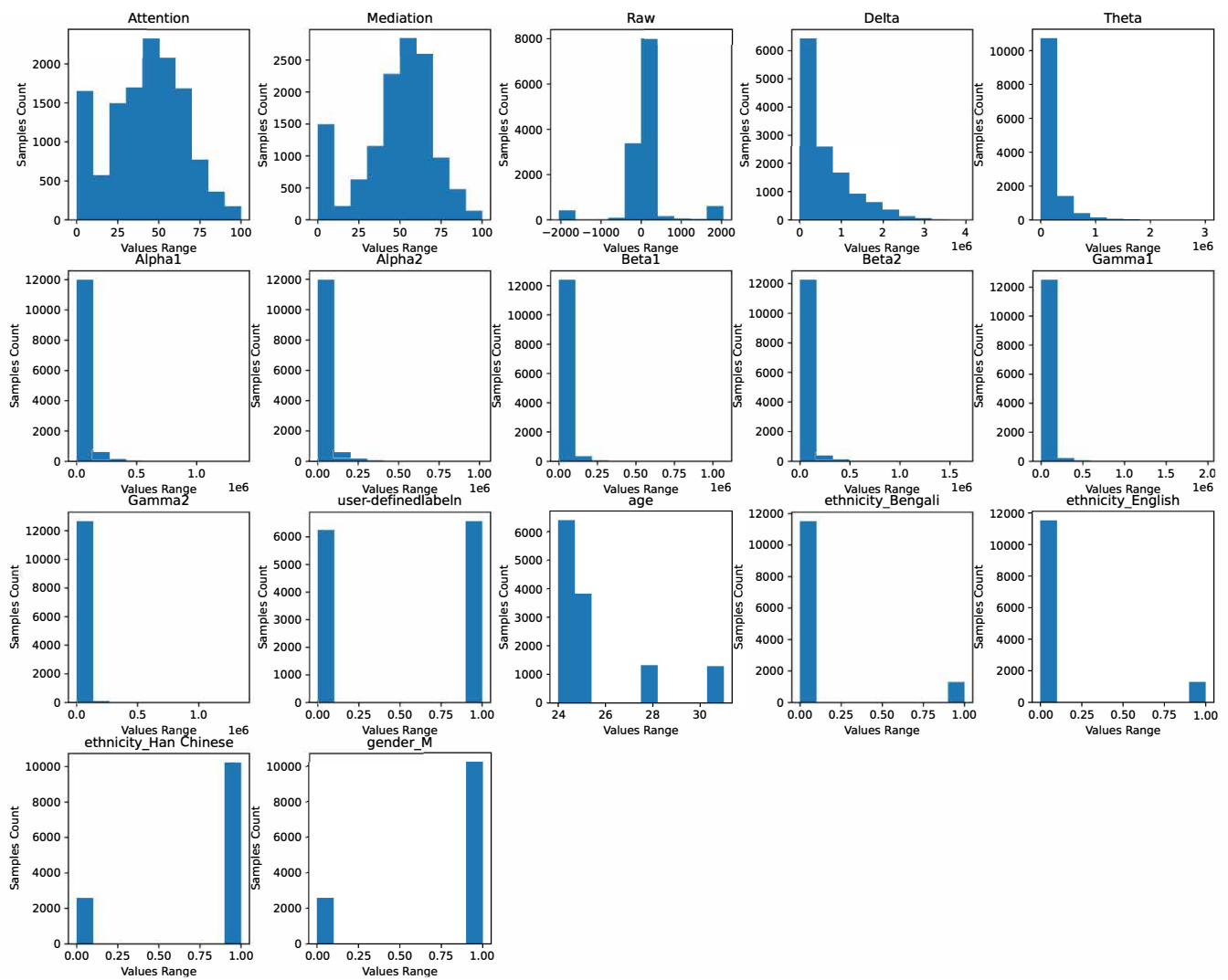


Figure 2. Histogram of EEG dataset features.

Table 4. Sample of dataset.

No.	Attention	Mediation	Raw	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2
1	56	43	278	301,963	90,612	33,735	23,991	27,946	45,097
2	40	35	−50	73,787	28,083	1439	2240	2746	3687
3	57	53	−73	2,265,079	48,307	82,437	140,472	15,464	227,432
4	64	64	−42	83,208	11,927	6755	811	2141	4271
5	63	66	279	901,346	44,037	18,886	27,924	8475	8999
No.	Gamma1	Gamma2	User-Definedlabel	Age	ethnicity_Bengali	ethnicity_English	ethnicity_Han Chinese	gender_M	
1	33,228	8293	0	25	0	0	1	1	
2	5293	2740	0	25	0	0	1	1	
3	30,097	6403	1	24	0	0	1	0	
4	6877	274	1	24	0	0	1	0	
5	16,990	2883	1	24	0	0	1	0	

3.2. Probability-Based Features

The used dataset consists of 17 features. The original dataset is not linearly separable to a great extent, which leads to poor performance from machine learning models. So in this study, we worked on feature engineering and proposed a novel approach called the probability-based features (PBS) technique. Figure 3 shows the architecture of the PBF approach.

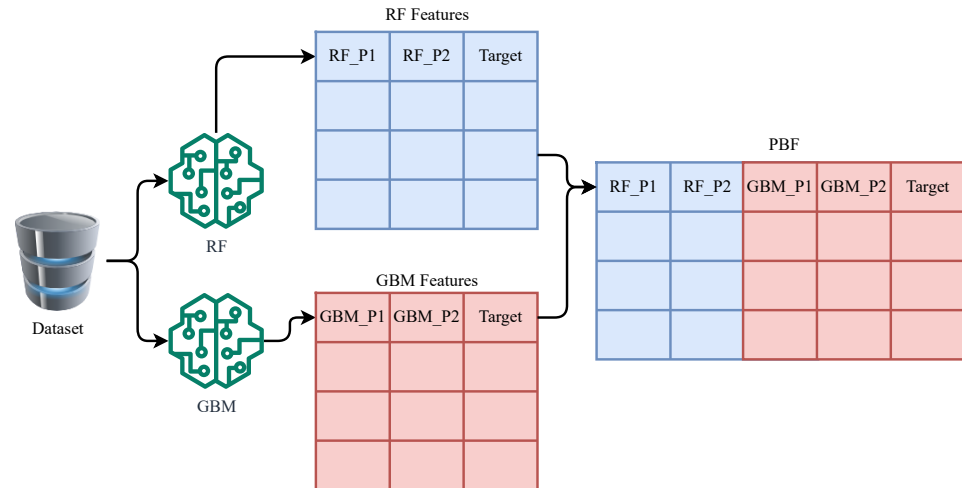


Figure 3. Schematic diagram of the proposed PBF approach.

Algorithm 1 shows the working of the proposed PBF approach. The approach utilizes two machine learning models including RF and GBM. Both models are tree-based ensemble models. These models are selected for feature engineering because the original feature is small in size and is not linearly separable. Predominantly, linear models do not show good performance on small-sized feature sets. For this reason, we design the probability-based feature set by using RF and GBM.

The whole dataset is passed to RF and GBM separately. The dataset has two target classes: “confused” and “not confused”. Consequently, the output from RF and GBM is in two classes. RF and GBM both provide the output in the form of class probability each for “confused” and “not confused”. These probabilities are then joined to make a feature set comprising two probabilities from RF and two from GBM. PBS has a total of four feature sets, which is more linearly separable and distinguishes both target classes with a higher margin.

Algorithm 1 Algorithm for PBF

Input: EEG Features

Output: Confused or Not-Confused

- 1: $RF \leftarrow RF \text{ Model}$
 - 2: $GBM \leftarrow GBM \text{ Model}$
 - 3: **for** i in Corpus **do**
 - 4: $Pob_{RF} \leftarrow RF(i)$
 - 5: $Pob_{GBM} \leftarrow GBM(i)$
 - 6: **end for**
 - 7: $PBF \leftarrow Concatenate(Pob_{RF}, Pob_{GBM})$
-

3.3. Machine Learning Classifiers

The use of machine learning algorithms for the EEG brainwave data has produced good results. Consequently, many algorithms and their variants can be found in the literature. For the current study, LR, RF, GBM, linear SVC, and ETC are used for confusion

detection in students while watching MOOC videos. For the implementation of these algorithms, the Scikit-Learn library is used. By fine-tuning several parameters, the performance of these algorithms has been optimized. A brief description of these algorithms is provided in this section, and Table 5 shows the hyperparameter setting of each model.

Table 5. Machine learning models' hyperparameters setting.

Model	Hyperparameters	Hyperparameters Tuning
ETC	n_estimators = 300, max_depth = 15	max_depth = {2 to 300}, n_estimators = {50 to 500}
RF	n_estimators = 300, max_depth = 15	max_depth = {2 to 300}, n_estimators = {2 to 300}
GBC	n_estimators = 300, max_depth = 15, learning_rate = 0.8	max_depth = {2 to 300}, n_estimators = {50 to 500}, learning_rate = {0.0 to 1.0}
SVC	Kernel = linear, C = 3.0	Kernel = {poly, linear, sigmoid} C = {1.0 to 5.0}
LR	solver = liblinear, C = 3.0, multi_class = ovr	Kernel = {liblinear, sag, saga} C = {1.0 to 5.0}, multi_class = ovr

LR is a machine learning algorithm used for classification problems [28]. It is a statistical method that is based on the logistic function and works on the concept of probability. The values of the S-shaped curve and variable v of the logistic function range from $-\infty$ to $+\infty$ for the actual number. In LR, the "liblinear" hyperparameter is used to boost the performance because it has a small corpus [29]. For the binary classification problems, the "multi-class" parameter is set to "multidimensional".

RF is an ensemble learning model used for constructing predictions with high precision by composing the results of sub-trees [30]. For the training of several decision trees, RF used bagging by using samples of bootstraps [31]. In RF, to achieve the best accuracy value, RF has been applied with many weak learners, and with n -estimator in RF shows the number of trees taking part in the prediction process to reduce the overfitting phenomenon [32]. The max-depth parameter is used in RF and "random state" is used for the randomness of the sample at the time of training.

GBM is a boosting model which works by a model formed by an ensemble of weak prediction models, commonly called decision trees. In boosting, weak learners are transformed into strong learners [33]. Every tree which is formed is a modified version of one before it and uses gradient as a loss function. Loss determines how well a model coefficient fits the underlying data. For the optimization of the model loss, function is used.

Linear SVC is a good choice for practical applications. SVC generates a hyperplane or line that divides the data into classes [34]. With the help of the kernel function, low dimensional input space is transformed into higher dimensional space. This means that non-separable issues are transformed into separable ones. Linear SVC mainly handles non-linear differential problems [35]. SVC separates the data based on labels and performs complex data transformations.

ETC is an ensemble learning model that considers the results of multiple uncorrelated decision trees for the final decision. Each decision tree in the forest used for further classification is formed by using training samples [36]. On the random sample of features, multiple uncorrelated decision trees are constructed. During the construction of trees, feature selection is done to split the data by using the Gini index for each feature [37].

3.4. Deep Learning Models for Experiments

In this study, we also deployed deep learning models in comparison with machine learning models. We used two state-of-the-art deep learning models, LSTM and CNN. These models are used with their best hyperparameters setting according to the dataset and their architecture is shown in Table 6. We used these models because they are mostly used in literature for this type of dataset. We compared it with our proposed approach for machine learning models. Both models are used with a vocabulary size of 50 and 20 output

dimensions. LSTM is used with 100 units; similarly, CNN is also used with 100 filters and 2×2 kernel size. We used binary_cross-entropy loss function because of the binary class problem and also used the Adam optimizer. We used 100 epochs and a batch size of 8 for model fitting.

Table 6. Architecture of deep learning models.

Model	Hyperparameters
LSTM	Embedding(50, 20, input_length = ..) Dropout(0.5) LSTM(100) Dense({2}, activation = 'softmax')
CNN	Embedding(50,20, input_length = ..) Conv1D(100, 2, activation = 'relu') MaxPooling1D(pool_size = 2) Activation('relu') Dropout(rate = 0.5) Flatten() Dense({2}, activation = 'softmax')
{loss = binary_crossentropy}', optimizer = 'adam', epochs = 100, batch_size = 8	

3.5. Performance Evaluation Parameters

To check the performance of machine learning algorithms, various evaluation matrices are used in this research. We used the confusion matrix. Every observation in the testing set is predicted in exactly one box. A confusion matrix is a tabular representation of a classification model's performance, and it consists of four parameters—true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). We used accuracy, precision, recall, and F1 score for the evaluation of learning models.

The measure of prediction correctness is called accuracy [38]. It is measured as

$$Accuracy = \frac{\text{total number of correct predictions}}{\text{total number of predictions}}. \quad (1)$$

It can also be represented as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (2)$$

Precision represents the ratio of true positives to all events predicted as true [39]. Its value lies between 0 and 1 and is calculated as

$$Precision = \frac{TP}{TP + FP}. \quad (3)$$

The recall represents the total number of positive classifications out of true class [40]. It is calculated as

$$Recall = \frac{TP}{TP + FN}. \quad (4)$$

F1 score represents a tradeoff between precision and recall, or it is a harmonic mean between precision and recall [41]. It is calculated as

$$F_1 \text{ Score} = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (5)$$

4. Results and Discussion

In this section, we discussed the results obtained from the three sets of experiments. We evaluate and compare the performance of all used machine learning models applied to the features obtained by using RF, GBM, LR, and SVM. The machine learning models employed in this study can be classified into two categories: tree-based classifiers, and regression-based classifiers. RF, GBC, and ETC are tree-based classifiers, whereas LR and linear SVC are regression-based models.

4.1. Experimental Setup

The analysis of the results for the confusion detection in students while watching MOOC videos is presented in this section. The SciKit-learn library and Natural Language Process Tool Kit (NLTK) are used to implement the machine learning models. Python's SciKit module is used to deploy machine learning algorithms. For conducting experiments, the Jupyter notebook is used. The data is split in the ratio of 0.7 to 0.3 for training and testing, respectively, because this ratio is adopted by many studies to avoid overfitting [42]. All the experiments are carried out on a 2 GB Dell PowerEdge T430 graphical processing unit on a 2X Intel Xeon 8 cores 2.4GHz machine equipped with a 32 GB DDR core random access memory (RAM).

4.2. Performance of Machine Learning Models Using Proposed PBF

For this set of experiments, we carried out experiments by using the proposed approach where the extracted features using RF and GBM learning algorithms are combined to make PBF. A comparative analysis of machine learning models has been conducted, and the results are given in Table 7. Results indicate that all models show 100% accuracy, precision, and recall when trained and tested by using the proposed approach PBF which shows the superiority of the approach over traditional feature engineering approaches.

Table 7. Machine learning models results by using RF and GBM extracted features.

Classifiers	Accuracy	Precision	Recall	F1 Score
RF	1.00	1.00	1.00	1.00
GBC	1.00	1.00	1.00	1.00
LR	1.00	1.00	1.00	1.00
SVC	1.00	1.00	1.00	1.00
ETC	1.00	1.00	1.00	1.00

Table 8 shows the 10-fold cross-validation results of our proposed approach and we can analyze that models are also significant with a 10-fold cross-validation approach when they used PBF. All models achieved a 1.00 mean accuracy score with ± 0.00 standard deviation which shows the significance of our proposed approach.

Table 8. 10-fold cross-validation results of RF and GBM features using machine learning algorithms.

Classifiers	Accuracy	Standard Deviation
RF	1.00	0.00
GBC	1.00	0.00
LR	1.00	0.00
SVC	1.00	0.00
ETC	1.00	0.00

4.3. Performance of Models Using LR and SVM Extracted Features

To show the efficacy of the proposed approach, we performed additional experiments. The purpose of the second set of experiments is to analyze the performance of the features extracted by using LR and SVM and compare the performance of machine learning models with that of PBF. The process of combining the features is the same as followed in PBF. In the second set of experiments, we used the features extracted by using LR and SVM learning algorithms. A comparative analysis of machine learning models has been conducted on these sets of features, and results are given in Table 9. Table 9 shows that the highest accuracy score of 60% is achieved by the LR and linear SVC with the same 60% precision, recall, and F1 score. The second-best results are obtained by the RF classifier which obtains a 59% accuracy, precision, recall, and F1 score, respectively. From the results, it can be observed that the regression-based models perform better than tree-based models.

Table 9. Machine learning models results using LR and SVM extracted features.

Classifiers	Accuracy	Precision	Recall	F1 Score
RF	0.59	0.59	0.59	0.59
GBM	0.55	0.55	0.55	0.55
LR	0.60	0.60	0.60	0.60
SVC	0.60	0.60	0.60	0.60
ETC	0.56	0.56	0.56	0.56

Figure 4 shows the confusion matrix for all models using LR and SVM extracted features. The “confused” class is represented by 1 while the “not confused” class is represented by 0 in the confusion matrix. From the total of 2563, LR obtains the highest number of 1543 correct predictions, followed by 1541 correct predictions by the SVC and RF with 1508 correct predictions. GBC has the lowest number of correct predictions, i.e., 1419 when using the features extracted from LR and SVM. Table 10 shows per-class precision, recall, and F1 scores of all the models.

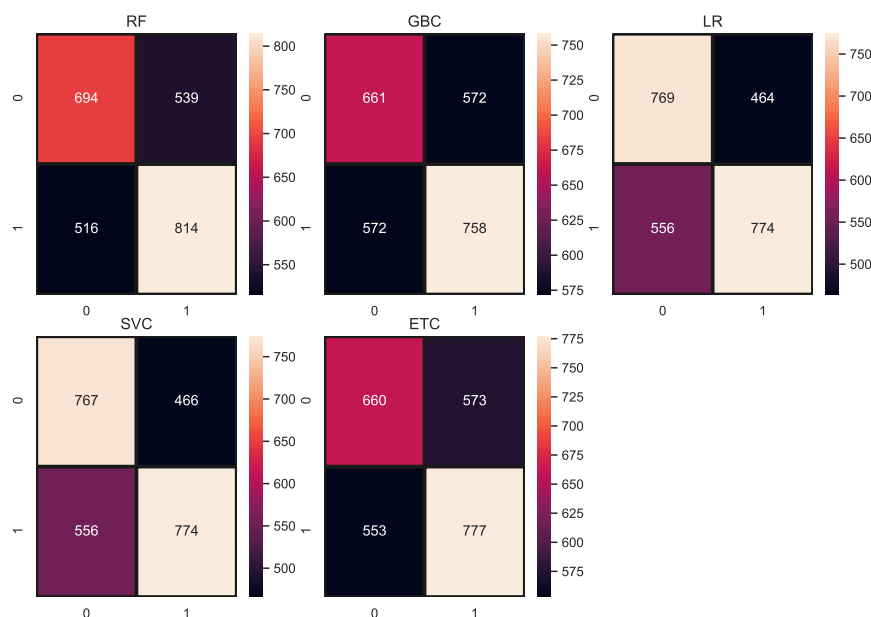


Figure 4. Confusion matrices of all models using LR and SVM extracted features.

Table 10 illustrates the per-class accuracy of the models, which demonstrates that all models are equally effective for both the target classes. The models are equally effective

for each target class, as demonstrated by these confusion matrices. After rounding off, the accuracy scores of RF, LR, and SVC are almost identical, although the number of accurate predictions varies slightly. These findings demonstrate that no class is overfitted by a model.

Table 10. Per-class accuracy of LR and SVM models.

Models	Class	Precision	Recall	F1 Score
RF	0	0.57	0.56	0.57
	1	0.60	0.61	0.61
GBC	0	0.54	0.54	0.54
	1	0.57	0.57	0.57
LR	0	0.58	0.62	0.60
	1	0.63	0.58	0.60
SVC	0	0.58	0.62	0.60
	1	0.62	0.58	0.60
ETC	0	0.54	0.54	0.54
	1	0.62	0.58	0.60

For validating the performance of machine learning models, 10-fold cross-validation is performed and the results of all machine learning algorithms are shown in Table 11. Results show that a 0.59% accuracy can be obtained by using both LR and SVC when features formed by using LR and SVM are utilized for training and testing the machine learning models.

Table 11. 10-fold cross-validation results on LR and SVM features using machine learning algorithms.

Classifiers	Accuracy	Standard Deviation
RF	0.56	0.04
GBC	0.53	0.03
LR	0.59	0.04
SVC	0.59	0.04
ETC	0.56	0.04

4.4. Performance of Machine Learning Models Using Original Features

In the final set of experiments, all the features from the dataset are used with the machine learning models. A comparative analysis of machine learning models has been conducted on this set of features, and results are given in Table 12.

Table 12 shows that the highest accuracy score of 77% is achieved by the RF with 77% precision, recall, and F1 score. The second best results are given by the ETC classifier by obtaining 75% accuracy, precision, recall, and F1 score. It can be observed that the tree-based models perform better than the regression-based models because enough data is available to construct the decision tree. This is the reason the tree-based models perform well as compared to the regression-based models.

Table 12. Results of machine learning models using original features.

Classifiers	Accuracy	Precision	Recall	F1 Score
RF	0.77	0.77	0.77	0.77
GBC	0.71	0.71	0.71	0.71
LR	0.61	0.61	0.61	0.61
SVC	0.52	0.56	0.51	0.38
ETC	0.75	0.75	0.75	0.75

Figure 5 presents the confusion matrices for all of the models. These confusion matrices show how the models are exposed to varying predictions for each target class. For a larger sample size, RF and ETC’s correct predictions are 1971 and 1934, respectively, out of a total of 2563. Table 13 shows F1 scores for both the RF and ETC models, which indicate that they make similar predictions across all classes. Out of 2563 total predictions, the GBC and LR have correct predictions of 1823 and 1557, respectively. The SVC model has the least accurate predictions of 1323.

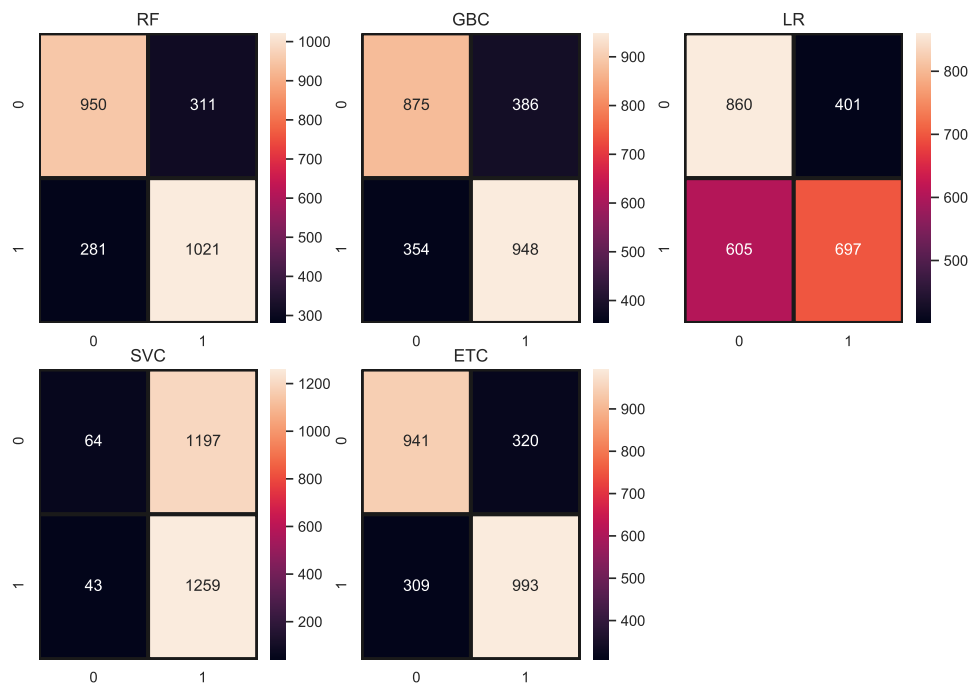


Figure 5. Confusion matrices of all models using original features.

Table 13 shows class-wise performance evaluation of all the models. It can be seen that the highest F1 score of 0.78 is obtained from the RF model for class 0. Class-wise performance from the models is almost similar. It is clear that SVC fails to predict class 0 and has the lowest F1 score of all the models.

We have zero false rates with the proposed approach because in this study, the machine learning model’s performance depends on the base machine learning model probabilities. We used two kinds of models for probability-based feature generation—one tree-based model (RF, GBM) and a second linear model (LR, SVM). The combination of RF+GBM generated a more correlated feature set as compared to the LR+SVM approach, and the reason is that LR and SVM require a large feature set to get a good fit and generate more accurate probabilities, whereas RF and GBM can also perform well on small feature sets. When RF and GBM generate a probability feature set, it is more correlated to the target class, which means that in the new feature, one target class value becomes totally different

compared to other class values. The new features set has clear patterns for “confused” or “not confused” target classes, which leads to a 100% accuracy score.

Table 13. Per-class accuracy using the original features.

Models	Class	Precision	Recall	F1 Score
RF	0	0.77	0.75	0.76
	1	0.77	0.78	0.78
GBC	0	0.71	0.69	0.70
	1	0.71	0.73	0.72
LR	0	0.59	0.68	0.63
	1	0.63	0.54	0.58
SVC	0	0.60	0.05	0.09
	1	0.51	0.97	0.67
ETC	0	0.75	0.75	0.75
	1	0.76	0.76	0.76

The 10-fold cross-validation results using machine learning algorithms with original features are shown in Table 14. It can be seen that the best performance can be obtained by using the ETC model which shows a 0.69% accuracy with a 0.03 standard deviation. RF also has the same accuracy, but its standard deviation is 0.04.

Table 14. K-fold cross-validation results on the original dataset using machine learning algorithms.

Classifiers	Accuracy	Standard Deviation
RF	0.69	0.04
GBC	0.67	0.03
LR	0.59	0.04
SVC	0.51	0.04
ETC	0.69	0.03

4.5. Comparison of Original and Probability-Based Features

Experimental results show that using the proposed PBF, the performance of the machine learning models is greatly enhanced. For verifying this performance, a visual representation of feature distribution is given in Figure 6. Figure 6a shows that with the original feature, the data is not linearly separable. However, when the proposed method of feature engineering is used, the feature space becomes more linearly separable, as shown in Figure 6b. This is the reason the performance of the machine learning models is elevated.

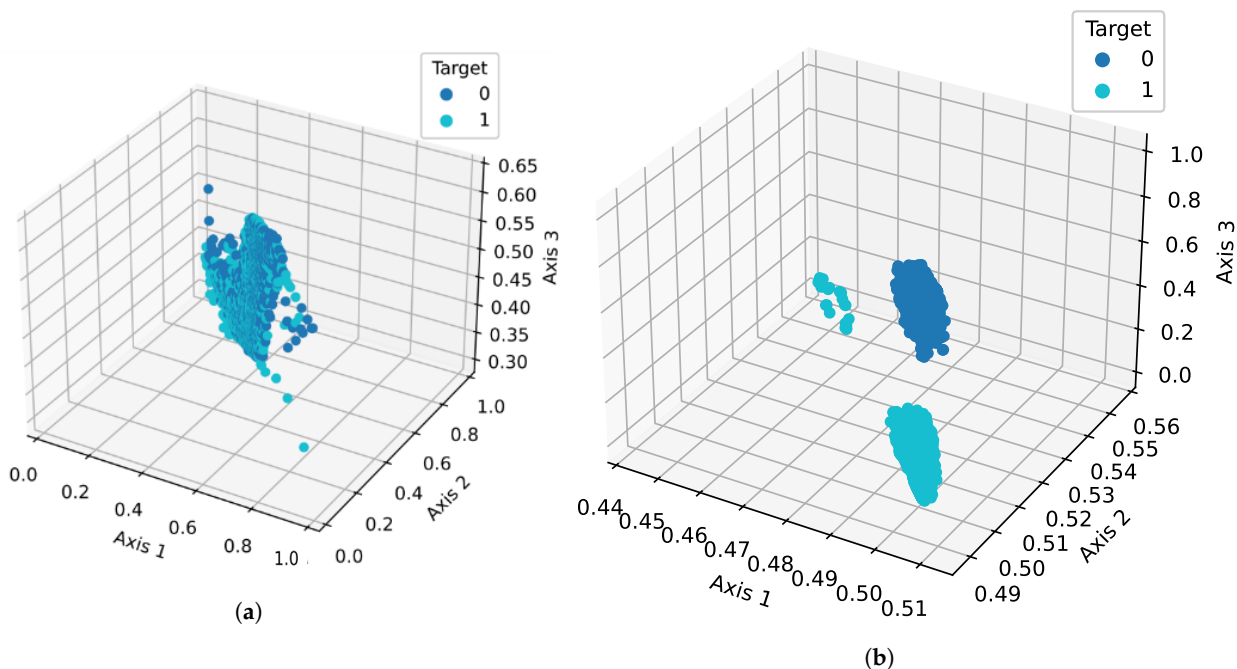


Figure 6. Feature space, (a) Class distribution using the original features, and (b) Class distribution using the proposed PBF.

Figure 7 shows the comparison between the machine learning model’s performance using RF+GBM, LR+SVM, and original features. We can see that models outperform tree-based model features (RF+GBM) whereas original features are poor as compared to RF+GBM features but it is better than LR+SVM features.

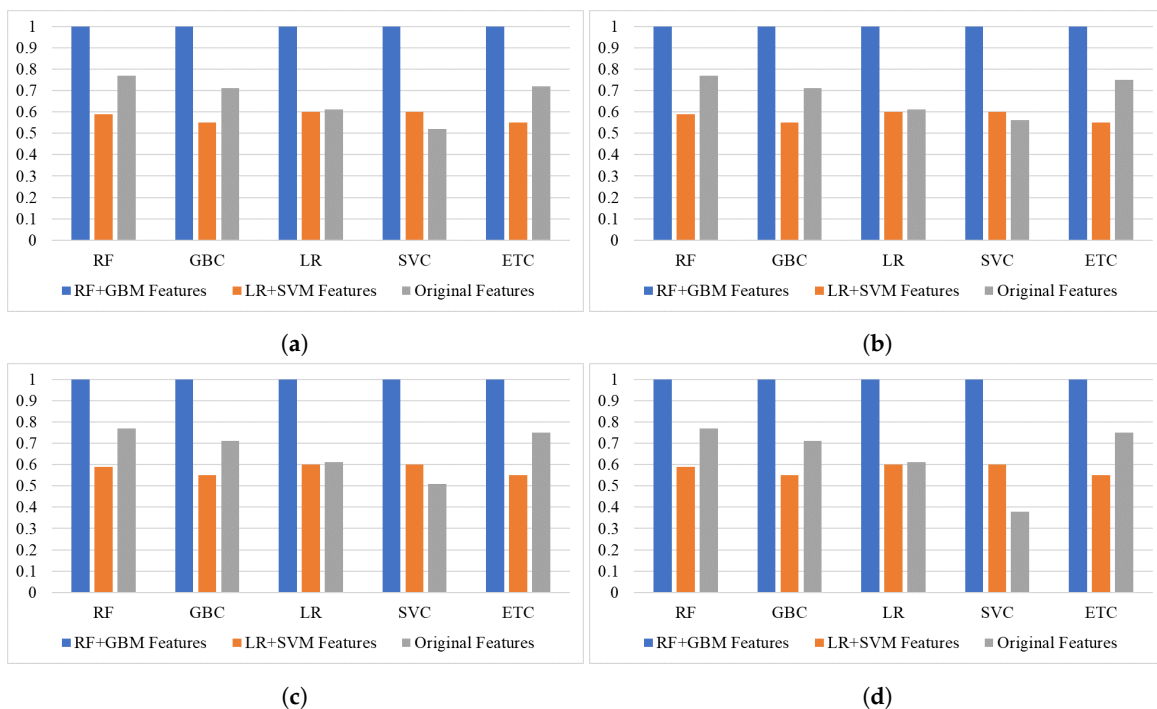


Figure 7. Performance comparison of models (a) Accuracy, (b) Precision, (c) Recall, and (d) F1 score.

We also calculated the computational cost for each model with different features engineering techniques. We find that the proposed approach is significant in terms of accuracy as well as efficiency. The computational cost for the proposed approach is significantly bet-

ter than other approaches. We achieved the highest accuracy with the lowest computational time as shown in Table 15.

Table 15. Computational cost (time in seconds) of machine learning models.

Classifiers	PBF	LR+SVM	Original
RF	0.43	2.87	3.41
GBC	0.14	2.73	2.65
LR	0.04	0.02	0.11
SVC	0.02	0.03	0.64
ETC	0.42	1.42	1.66

4.6. Results of Deep Learning Models

We deployed LSTM and CNN models on the original dataset. Figure 8 shows the results of the deep learning model per epoch. It shows that the LSTM obtains the highest accuracy at the 54th epoch; after that, there are ups and downs in the accuracy. Similarly, CNN obtains the best accuracy at the 81st epoch, and after that, there is no change.

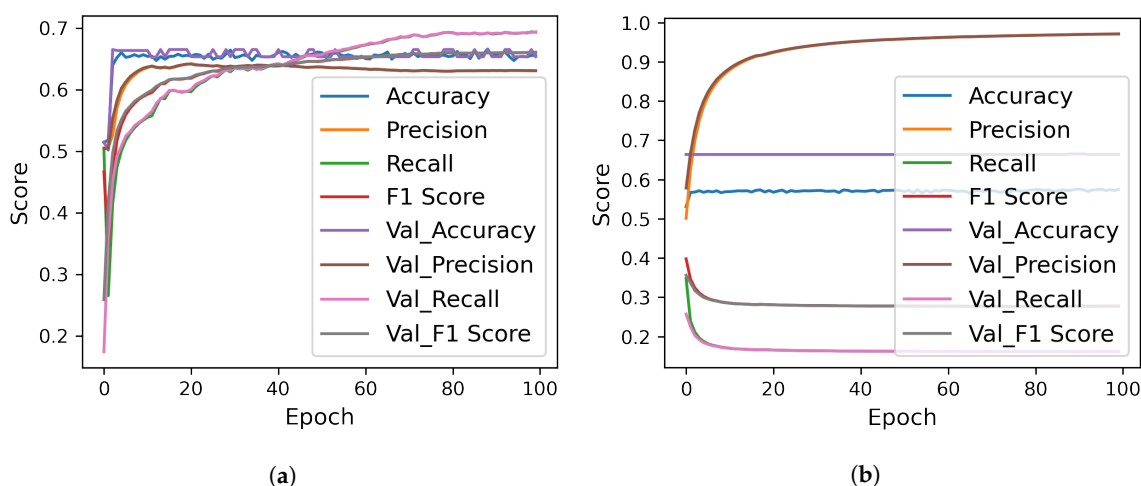


Figure 8. Deep learning models score per epochs. (a) LSTM; (b) CNN.

Experimental results for deep learning models are shown in Table 16. According to the results models, performance is poor as compared to the machine learning models, even on the original dataset. The performance of deep learning models is better on a large feature set as compared to using a small dataset. The used dataset has a small feature set which is the reason the performance of deep learning models is low as compared to machine learning models. LSTM achieved a 0.67 accuracy score and CNN achieved only a 0.66 accuracy score.

Table 16. Results of deep learning models.

Model	Accuracy	Class	Precision	Recall	F1 Score
LSTM	0.67	0	1.00	0.31	0.47
		1	0.61	1.00	0.75
		Avg.	0.80	0.65	0.61
CNN	0.66	0	0.99	0.31	0.47
		1	0.61	1.00	0.75
		Avg.	0.80	0.65	0.61

4.7. Statistical T-Test

We have done a statistical T-test to compare models' performance by using RF+GBM features with other used features. A T-test will find whether compared results are statistically different and significant or not. To evaluate the T-test, there is a null hypothesis. If the value of the T-score value given by the T-test is greater than the critical value (cv), the null hypothesis will be rejected and the alternative hypothesis will be accepted. Our null hypothesis and alternative hypothesis are:

- Null hypothesis; the proposed approach (RF+GBM) is not statistically significant as compared to the other approach; and
- Alternative hypothesis; the proposed approach (RF+GBM) is statistically significant as compared to the other approach.

Table 17 shows the results of the T-test with two cases. When we compared model results by using RF+GBM features and model performance by using original features. The T-test rejected the null hypothesis and accepted the alternative hypothesis. Similarly the RF+GBM vs. the LR+SVM approach shows that in our model performance with our proposed features, RF+GBM is statistically significant.

Table 17. Statistical T-test results.

Case	T-Score	df	cv	Null Hypothesis
RF+GBM vs. Original features	6.98	4	7.06×10^{-17}	Rejected
RF+GBM vs. LR+SVM	33.03	4	7.06×10^{-17}	Rejected

4.8. Comparison with Existing Studies

In this section, we compare the proposed approach with existing studies on the same dataset. These studies all worked on different machine learning or deep learning approaches to achieve significant results but no one focused on feature engineering. The study [18] proposed bi-directional long short-term memory (Bi-LSTM) for confused student prediction. Similarly, the study [13] also deployed Bi-LSTM for confused students' predictions and also adopted a feature selection technique to enhance the performance. The study [17] also worked on student confusion prediction by using the machine learning model ETC. In comparison with all approaches, we worked on feature engineering and deployed machine learning models to achieve significant results. A comparison between previous studies and our approach is shown in Table 18, which indicates that the proposed approach outperforms existing models.

Table 18. Comparison with existing studies.

Study	Year	Models	Reported Accuracy
[18]	2017	Bi-LSTM	73.3 %
[20]	2019	CF-BiLSTM	75%
[13]	2020	BiLSTM	88%
[17]	2022	ETC	66.6%
This study	2022	RF, GBC, LR, SVC, ETC	100%

4.9. Performance of Proposed Approach with "Feeling Emotions" Dataset

For validating the performance of the proposed approach, we deployed the proposed approach on another dataset which is the EEG Brainwave Dataset: Feeling Emotions. This dataset consists of three sentiments—positive, neutral, and negative. The dataset is collected by using the Muse EEG headband. We performed experiments by using the proposed feature set approach with machine learning models. Experiments are performed

with a two-fold purpose. First, the performance of the proposed approach is validated by using the EEG data for emotions. Secondly, the performance of the proposed approach is analyzed for emotion classification by using the EEG dataset. Table 19 shows the comparison between the proposed approach and other studies that utilized the same dataset. The results demonstrate that by using the proposed approach, an accuracy of 100% is achieved. Results also prove that the proposed approach outperforms state-of-the-art approaches that use the same EEG data for emotion classification.

Table 19. Performance of the proposed approach on the Feeling Emotions dataset.

Ref.	Year	Model	Accuracy
[43]	2018	SVM, RF	87%
[44]	2019	RF	97.89%
[45]	2021	RF, XGBoost	96.88%, 96.41%
[46]	2022	RNN	Training: 97.50% Testing: 96.50%
		SVM, LR	Training: 100.00% Testing: 97.25%
[47]	2022	RNN	97%
This study	2022	RF, GBC, LR, SVC, ETC	100%

5. Conclusions

Online education became an attractive mode of education recently, especially during the COVID-19 outbreak. Unlike the traditional modes of education, it does not require being in the classrooms physically, and classes are conducted online. However, lacking face-to-face interaction with the instructor, student's level of understanding, or confusion regarding particular topics cannot be judged, which raises serious concerns. This study leverages the electroencephalogram data to detect confused students by using a machine learning approach in the context of the MOOC platform. An intuitive feature engineering approach is proposed, which utilizes the class probabilities output from RF and GBM to make the feature vector. Experiments are performed by using machine learning and deep learning models with the original features, as well as the proposed PBF approach. It is found that machine learning models tend to show better results than deep learning models. Results indicate that by using the proposed feature engineering approach a 100% accuracy for confused student detection can be obtained. Results are further corroborated by using k-fold cross-validation and the Feeling Emotions dataset. Furthermore, performance comparison with the state-of-the-art approaches shows that the proposed approach outperforms existing studies. In the future, we intend to work on feature augmentation. Because deep learning models could not perform well on the small feature set, we plan to increase the number of features. In addition, we want to work with real-time confusion detection in the future.

Author Contributions: Conceptualization, T.D. and F.R.; data curation, T.D.; formal analysis, F.R.; investigation, W.A.; methodology, T.D. and W.A.; project administration, A.H.B.; software, A.H.B.; supervision, I.A.; validation, I.A.; visualization, A.H.B.; writing—original draft, F.R., T.D. and W.A.; writing—review & editing, I.A. and F.R. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Data Availability Statement: The dataset and code for machine learning models used in this study are available via the following link: <https://github.com/furqanrustam/EEG-Brainwave> (accessed on 15 August 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Allen, I.E.; Seaman, J. *Going the Distance: Online Education in the United States*; ERIC: Washington, DC, USA, 2011.
2. Thompson, G. How can correspondence-based distance education be improved? A survey of attitudes of students who are not well disposed toward correspondence study. *J. Distance Educ.* **1990**, *5*, 53–65.
3. Sharma, R.C. Innovative applications of online pedagogy and course design. *Int. J. Inf. Commun. Technol. Educ.* **2019**, *15*, 451.
4. Sublett, C. What do we know about online coursetaking, persistence, transfer, and degree completion among community college students? *Community Coll. J. Res. Pract.* **2019**, *43*, 813–828. [[CrossRef](#)]
5. Suhaimi, N.S.; Mountstephens, J.; Teo, J. EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* **2020**, *2020*, 8875426. [[CrossRef](#)]
6. Li, Y.; Liu, Y.; Cui, W.G.; Guo, Y.Z.; Huang, H.; Hu, Z.Y. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 782–794. [[CrossRef](#)]
7. Khan, K.A.; Shanir, P.; Khan, Y.U.; Farooq, O. A hybrid Local Binary Pattern and wavelets based approach for EEG classification for diagnosing epilepsy. *Expert Syst. Appl.* **2020**, *140*, 112895. [[CrossRef](#)]
8. Marosi, E.; Bazán, O.; Yanez, G.; Bernal, J.; Fernandez, T.; Rodriguez, M.; Silva, J.; Reyes, A. Narrow-band spectral measurements of EEG during emotional tasks. *Int. J. Neurosci.* **2002**, *112*, 871–891. [[CrossRef](#)]
9. Ding, Y.; Chen, X.; Zhong, S.; Liu, L. Emotion Analysis of College Students Using a Fuzzy Support Vector Machine. *Math. Probl. Eng.* **2020**, *2020*, 8931486. [[CrossRef](#)]
10. Baars, B.J.; Gage, N.M. *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*; Academic Press: Cambridge, MA, USA, 2010.
11. Alotaiby, T.; Abd El-Samie, F.E.; Alshebeili, S.A.; Ahmad, I. A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process.* **2015**, *2015*, 1–21. [[CrossRef](#)]
12. Kumar, H.; Sethia, M.; Thakur, H.; Agrawal, I. Swarnalatha, P. Electroencephalogram with Machine Learning for Estimation of Mental Confusion Level. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 761–765.
13. Li, G.; Jung, J.J. Maximum marginal approach on eeg signal preprocessing for emotion detection. *Appl. Sci.* **2020**, *10*, 7677. [[CrossRef](#)]
14. Sarwat, S.; Ullah, N.; Sadiq, S.; Saleem, R.; Umer, M.; Eshmawi, A.; Mohamed, A.; Ashraf, I. Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM. *Sensors* **2022**, *22*, 4834. [[CrossRef](#)]
15. Li, Z.; Qiu, L.; Li, R.; He, Z.; Xiao, J.; Liang, Y.; Wang, F.; Pan, J. Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection. *Sensors* **2020**, *20*, 3028. [[CrossRef](#)] [[PubMed](#)]
16. Aljedaani, W.; Aljedaani, M.; AlOmar, E.A.; Mkaouer, M.W.; Ludi, S.; Khalaf, Y.B. I cannot see you—The perspectives of deaf students to online learning during covid-19 pandemic: Saudi arabia case study. *Educ. Sci.* **2021**, *11*, 712. [[CrossRef](#)]
17. Anala, V.A.S.M.; Bhumireddy, G. Comparison of Machine Learning Algorithms on Detecting the Confusion of Students While Watching MOOCs. 2022. Available online: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1641701&dswid=-4947> (accessed on 15 August 2022).
18. Ni, Z.; Yuksel, A.C.; Ni, X.; Mandel, M.I.; Xie, L. Confused or not confused? Disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Boston, MA, USA, 20–23 August 2017; pp. 241–246.
19. Wang, H.; Li, Y.; Hu, X.; Yang, Y.; Meng, Z.; Chang, K.M. Using EEG to Improve Massive Open Online Courses Feedback Interaction. In Proceedings of the AIED Workshops, Memphis, TN, USA, 9–13 July 2013.
20. Wang, H.; Wu, Z.; Xing, E.P. Removing confounding factors associated weights in deep neural networks improves the prediction accuracy for healthcare applications. In *BIOCOMPUTING 2019: Proceedings of the Pacific Symposium*; World Scientific: Singapore, 2018; pp. 54–65. Available online: <https://pubmed.ncbi.nlm.nih.gov/30864310/> (accessed on 15 August 2022).
21. Li, N.; Kelleher, J.D.; Ross, R. Detecting Interlocutor Confusion in Situated Human-Avatar Dialogue: A Pilot Study. *arXiv* **2022**, arXiv:2206.02436.
22. Edla, D.R.; Mangalorekar, K.; Dhavalikar, G.; Dodia, S. Classification of EEG data for human mental state analysis using Random Forest Classifier. *Procedia Comput. Sci.* **2018**, *132*, 1523–1532. [[CrossRef](#)]
23. Yeo, M.V.; Li, X.; Shen, K.; Wilder-Smith, E.P. Can SVM be used for automatic EEG detection of drowsiness during car driving? *Saf. Sci.* **2009**, *47*, 115–124. [[CrossRef](#)]
24. Sun, L.; Jin, B.; Yang, H.; Tong, J.; Liu, C.; Xiong, H. Unsupervised EEG feature extraction based on echo state network. *Inf. Sci.* **2019**, *475*, 1–17. [[CrossRef](#)]
25. Hajinoroozi, M.; Jung, T.P.; Lin, C.T.; Huang, Y. Feature extraction with deep belief networks for driver's cognitive states prediction from EEG data. In Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), Chengdu, China, 12–15 July 2015; pp. 812–815.
26. Petrosian, A.; Prokhorov, D.; Lajara-Nanson, W.; Schiffer, R. Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG. *Clin. Neurophysiol.* **2001**, *112*, 1378–1387. [[CrossRef](#)]

27. Confused Student EEG Brainwave Data. Available online: <https://www.kaggle.com/datasets/wanghaohan/confused-eeeg> (accessed on 3 September 2022).
28. Aljedaani, W.; Mkaouer, M.W.; Ludi, S.; Javed, Y. Automatic Classification of Accessibility User Reviews in Android Apps. In Proceedings of the 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 1–3 March 2022; pp. 133–138.
29. Sebastiani, F. Machine learning in automated text categorization. *ACM Comput. Surv. (CSUR)* **2002**, *34*, 1–47. [[CrossRef](#)]
30. AlOmar, E.A.; Aljedaani, W.; Tamjeed, M.; Mkaouer, M.W.; El-Glaly, Y.N. Finding the needle in a haystack: On the automatic identification of accessibility user reviews. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–15.
31. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]
32. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
33. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *19*, 1189–1232. [[CrossRef](#)]
34. Safdari, N.; Alrubaye, H.; Aljedaani, W.; Baez, B.B.; DiStasi, A.; Mkaouer, M.W. Learning to rank faulty source files for dependent bug reports. In *Big Data: Learning, Analytics, and Applications*; SPIE: Bellingham, WA, USA, 2019; Volume 10989, pp. 60–78. [[CrossRef](#)]
35. Xindong, W.; Kumar, J.V.; Quinlan, R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.; Angus, N.; Liu, B.; Philip, S.; et al. Top 10 algorithms in data mining. Knowledge and Information Systems. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
36. Sharaff, A.; Gupta, H. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 189–197.
37. Ossai, C.I.; Wickramasinghe, N. GLCM and statistical features extraction technique with Extra-Tree Classifier in Macular Oedema risk diagnosis. *Biomed. Signal Process. Control* **2022**, *73*, 103471. [[CrossRef](#)]
38. Abid, M.A.; Ullah, S.; Siddique, M.A.; Mushtaq, M.F.; Aljedaani, W.; Rustam, F. Spam SMS filtering based on text features and supervised machine learning techniques. *Multimed. Tools Appl.* **2022**, 1–19. [[CrossRef](#)]
39. Amaar, A.; Aljedaani, W.; Rustam, F.; Ullah, S.; Rupapara, V.; Ludi, S. Detection of fake job postings by utilizing machine learning and natural language processing approaches. *Neural Process. Lett.* **2022**, *54*, 2219–2247. [[CrossRef](#)]
40. Rupapara, V.; Rustam, F.; Aljedaani, W.; Shahzad, H.F.; Lee, E.; Ashraf, I. Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Sci. Rep.* **2022**, *12*, 1–15. [[CrossRef](#)]
41. Fang, F.; Wu, J.; Li, Y.; Ye, X.; Aljedaani, W.; Mkaouer, M.W. On the classification of bug reports to improve bug localization. *Soft Comput.* **2021**, *25*, 7307–7323. [[CrossRef](#)]
42. Ashraf, I.; Umer, M.; Majeed, R.; Mehmood, A.; Aslam, W.; Yasir, M.N.; Choi, G.S. Home automation using general purpose household electric appliances with Raspberry Pi and commercial smartphone. *PLoS ONE* **2020**, *15*, e0238480.
43. Bird, J.J.; Manso, L.J.; Ribeiro, E.P.; Ekárt, A.; Faria, D.R. A study on mental state classification using eeg-based brain-machine interface. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 795–800.
44. Bird, J.J.; Ekart, A.; Buckingham, C.D.; Faria, D.R. Mental emotional sentiment classification with an eeg-based brain-machine interface. In Proceedings of the International Conference on Digital Image and Signal Processing (DISP'19), Oxford, UK, 29–30 April 2019. Available online: https://www.researchgate.net/publication/329403546_Mental_Emotional_Sentiment_Classification_with_an_EEG-based_Brain-machine_Interface (accessed on 15 August 2022).
45. Klibi, S.; Mestiri, M.; Farah, I.R. Emotional behavior analysis based on EEG signal processing using Machine Learning: A case study. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Virtual, 4–5 July 2021; pp. 1–7.
46. Sarkar, A.; Singh, A.; Chakraborty, R. A deep learning-based comparative study to track mental depression from EEG data. *Neurosci. Inf.* **2022**, *2*, 100039. [[CrossRef](#)]
47. Chowdary, M.K.; Anitha, J.; Hemanth, D.J. Emotion Recognition from EEG Signals Using Recurrent Neural Networks. *Electronics* **2022**, *11*, 2387. [[CrossRef](#)]