ODL-BCI: Optimal deep learning model for brain-computer interface to classify students concentration via hyper-parameter tuning

Md. Ochiuddin Miah[1], and Umme Habiba[1], and Md Faisal Kabir[2]
Department of Computer Science, North Dakota State University, Fargo, USA[1]
Department of Computer Science, Pennsylvania State University - Harrisburg, USA[2]
Email: mdochiuddin.miah@ndsu.edu, umme.habiba@ndsu.edu, mpk5904@psu.edu

***Abstract*—Brain-computer interface (BCI) research has gained attention in education contexts, offering the potential to monitor and enhance student's cognitive states and this study focuses on developing an optimal deep learning model, ODL-BCI, for real-time classification of students' concentration levels. The model incorporates hyper-parameter tuning techniques and utilizes the publicly available "Confused student EEG brainwave data" dataset. We propose a deep learning model with hyperparameters optimized through Bayesian optimization. The model's architecture is constructed with an input layer, several hidden layers, and an output layer. The number of nodes and activation functions in the hidden layers is determined using Bayesian optimization. The learning rate is also optimized for each layer. The proposed model is evaluated and compared with several standard machine learning classifiers, including Decision Tree, AdaBoost, Bagging, MLP, Naïve Bayes, Random Forest, SVM, and XG Boost, on an EEG confusion dataset. Experimental results demonstrate that the optimized deep learning model outperforms all other classifiers, achieving an accuracy of 74 percent. The model's effectiveness in accurately classifying students' concentration levels highlights its potential as a valuable tool in educational settings. This research contributes to the advancement of BCI technology, providing insights into the optimization of deep learning models for EEG-based cognitive assessment. Future work involves exploring the model's generalizability on larger datasets and extending its applicability to other BCI applications.**

Keywords*—Brain Computer Interface; Electroencephalogram; Hyper-parameter Tuning; Machine Learning; Deep Learning; Neural Network; Bayesian Optimization;

I. Introduction

The emergence of brain-computer interface (BCI) technologies, coupled with advancements in machine learning algorithms, has opened up new frontiers in understanding and decoding human cognitive states [1], [2]. In education, one critical aspect lies in monitoring and assessing students' concentration levels, which can provide valuable insights into their learning behaviors and enable educators to tailor personalized teaching methods for improved learning outcomes. However, discerning and classifying cognitive states like concentration levels from electroencephalogram (EEG) signals remain a challenging task due to the complexity and high-dimensionality of these signals [3].

Addressing this difficulty can be accomplished through the utilization of deep learning methods. Deep learning, a subset of machine learning that employs deep neural networks, demonstrates exceptional proficiency in managing intricate and high-dimensional data like EEG signals. Its effectiveness is linked to its capacity to autonomously learn layered feature representations, eliminating the necessity for manual feature extraction. This attribute makes it a highly valuable instrument for scrutinizing EEG data [4], [5].

Nevertheless, the performance of deep learning models depends heavily on the appropriate selection of hyperparameters, including the number of hidden layers, the nodes per layer, the type of activation function, and the learning rate. The traditional methods to tune these hyperparameters, such as grid search or random search, have significant drawbacks. They can be computationally expensive and inefficient in exploring the hyperparameter space, and they do not consider the interactions between hyperparameters, which may impact the model's performance[6].

This paper presents an ODL-BCI approach, which optimizes a deep learning model for brain-computer interface, specifically designed to classify students' concentration levels using EEG signals. The novelty of ODL-BCI lies in the incorporation of Bayesian optimization for hyperparameter tuning. Bayesian optimization constructs a probabilistic model of the objective function that maps hyperparameters to a probability of a score on the objective function. Unlike traditional methods, it uses information from past evaluations for future selection, thus offering an efficient exploration of the hyperparameter space[7]. By incorporating the complex interaction of hyperparameters and balancing exploration and exploitation in the hyperparameter space, Bayesian optimization improves the performance of the deep learning model[6].

Our research findings provide new perspectives on the application of deep learning and Bayesian optimization in the context of BCI and EEG signal analysis, paving the way for more efficient and effective cognitive state classification model.

The rest of this paper is organized as Section II Literature Review, a summary of similar existing works. Then, we move to Section III to explain the Experimental Setup and how we got EEG data. In Section IV, we detail our Proposed Method. We share the Results of our experiments in Section V and then wrap up with Section VI, where we Conclude and discuss Future Work, suggesting where this research could go next.

II. Related Work

Several studies have explored the application of deep learning and machine learning models in brain-computer interfaces

(BCIs) for various cognitive tasks. Miah et al. [8] propose CluSem, a clustering-based ensemble method for predicting motor imagery tasks from multi-channel EEG data. Their approach achieves high accuracy in classifying motor imagery tasks, showcasing its potential for improving the performance of brain-computer interfaces. The study contributes to the development of more accurate and reliable brain signal analysis techniques. Kabir et al. [9] conduct a comprehensive performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. Their findings shed light on the impact of different dimensionality reduction techniques on the predictive accuracy of cancer prediction models. This research provides valuable insights into optimizing machine learning models for accurate cancer prediction and diagnosis.

Santamaría-Vázquez et al. [10] propose a robust asynchronous control method for ERP-based brain-computer interfaces (BCIs) using deep learning. Their research focuses on enhancing the usability and effectiveness of BCIs by leveraging deep learning techniques. The study demonstrates the potential of deep learning for achieving reliable and efficient asynchronous control of BCIs, facilitating improved communication between humans and machines.Wu et al. [11], the authors propose an automated approach for hyperparameter tuning in deep learning-based side-channel analysis. By optimizing the hyperparameters, the authors demonstrate improved performance in analyzing side-channel vulnerabilities. This research contributes to the advancement of automated techniques for enhancing the effectiveness of deep learning-based side-channel analysis.

Men and Li [12] investigate the use of EEG signals to detect student confusion in Massive Open Online Courses (MOOCs). Their study demonstrates the potential of EEG-based analysis in monitoring and understanding the cognitive states of students during online learning. By detecting confusion, this research contributes to developing effective interventions and personalized learning strategies in MOOCs, ultimately enhancing the learning experience. Kunang et al. [13], the authors investigate the application of deep learning and hyperparameter optimization techniques for classifying attacks in an intrusion detection system (IDS). The study demonstrates the effectiveness of deep learning models in accurately classifying different types of attacks, and the importance of hyperparameter optimization in improving the performance of the IDS. This research contributes to the field of cybersecurity by showcasing the potential of deep learning and hyperparameter optimization for robust attack classification in IDS.

Cooney et al. [14] evaluate the impact of hyperparameter optimization in machine and deep learning methods for decoding imagined speech using EEG signals. Their research emphasizes the importance of optimizing hyperparameters for achieving accurate and reliable decoding of imagined speech from EEG data. The study provides valuable insights into optimizing machine learning models for understanding speech-related brain activity. Reñosa et al. [15] focus on classifying confusion levels using EEG data and artificial neural networks (ANNs). Their study demonstrates the potential of EEG-based classification in assessing cognitive states, such as confusion. By leveraging ANNs, the authors provide insights into using EEG data for real-time cognitive assessment, with potential applications in various fields, including education and human-computer interaction.

Miah et al. [16] develop a real-time EEG classification system for predicting voluntary hand movement directions using brain-machine interfaces (BMIs). Their study demonstrates the feasibility of using EEG signals to accurately classify hand movement directions in real-time, contributing to the advancement of BMIs. The research showcases the potential of EEG-based BMIs for applications in prosthetics, rehabilitation, and assistive technology. Thomas et al. [17] explore the application of deep learning techniques in brain-computer interfaces (BCIs). Their research demonstrates that deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), outperform traditional methods in accurately classifying brain signals. This highlights the potential of deep learning to enhance the performance and usability of BCIs, paving the way for more effective brain-computer communication.

Above existing papers have contributed to the field of Brain-Computer Interfaces (BCIs). However, there are notable gaps that our paper addresses. First, many existing approaches lack a focused effort on hyperparameter optimization, relying on traditional methods that might limit BCI model optimization. Our paper introduces more efficient Bayesian optimization for hyperparameter tuning. Second, there is a shortage of specialized models for classifying cognitive states, like concentration levels, using EEG signals. Our paper bridges this gap by introducing ODL-BCI, designed explicitly for personalized education and cognitive state monitoring. Third, the efficiency required for practical BCI deployment in real-world scenarios is often overlooked. Our paper emphasizes both effectiveness and efficiency. Lastly, while previous studies primarily report accuracy, our comprehensive analysis in terms of sensitivity, precision, and F-score demonstrates the practical advantages of ODL-BCI over existing algorithms, highlighting its novelty in these key aspects.

III. Experimental Paradigm and Signal Acquisitions

A. *NeuroSky MindSet*

The NeuroSky MindSet is an innovative brain-computer interface (BCI) device engineered for the measurement and recording of electrical brain activity through electroencephalography (EEG)[18]. It is primarily marketed as a brain-computer interface (BCI) device, allowing users to interact with various applications and devices using their brainwave activity[19]. The main feature of the NeuroSky MindSet headset is equipped with an EEG sensor, which is its primary component for measuring the brain's electrical activity[20]. This single dry electrode is strategically positioned on the FP1 area of the forehead to capture brainwave data. Complementing this, the headset also employs reference and ground sensors located on the ear clip. These are essential for establishing a baseline for the EEG measurements, ensuring the captured data remains accurate and devoid of external interference[21]. It records numeric values, determined by proprietary algorithms, reflecting the user's mental states like attention and meditation. The device also logs numeric values for frequency bands,

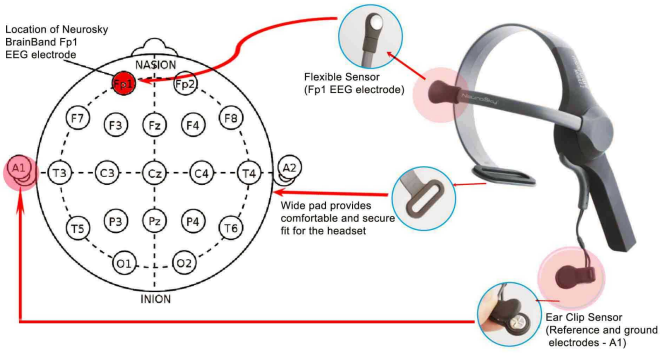

Fig. 1: NeuroSky headset electrodes distribution according to the 10-20 international system. Source: NeuroSky.

capturing data from 0 to 60Hz every half-second[15]. It has 14 metallic electrodes mounted on a plastic base, meticulously positioned on the scalp in accordance with the internationally recognized 10-20 system [22]. Refer to Figure 1 for a visual representation of the electrode layout, aligning with the 10-20 system [23]. The 10-20 system's meticulous electrode placement ensures consistency and standardization by distributing electrodes based on precise percentages of the scalp's left-right and front-back dimensions [24]. This adherence to the 10-20 system simplifies integration with existing EEG research and data analysis methodologies.

B. EEG Dataset Background and Description

The primary objective of the EEG dataset was to investigate the relationship between students' confusion levels and EEG signals while watching Massive Open Online Courses (MOOC). The dataset was inspired by a pilot study conducted by Wang et al. [25], where college students' EEG signals were collected to determine their confusion levels when exposed to MOOC content.

TABLE I: Features extracted from MindSet.

Features	Description	Sampling Rate
Attention	Proprietary measure of mental focus	1 Hz
Meditation	Proprietary measure of calmness	1 Hz
Raw	Raw EEG signal	512 Hz
Delta (δ)	1-3 Hz of power spectrum	8 Hz
Theta (θ)	4-7 Hz of power spectrum	8 Hz
Alpha1 (α)	Lower 8-11 Hz of power spectrum	8 Hz
Alpha2 (α)	Higher 8-11 Hz of power spectrum	8 Hz
Beta1 (β)	Lower 12-29 Hz of power spectrum	8 Hz
Beta2 (β)	Higher 12-29 Hz of power spectrum	8 Hz
Gamma1 (γ)	Lower 30-100 Hz of power spectrum	8 Hz
Gamma2 (γ)	Higher 30-100 Hz of power spectrum	8 Hz

Ten college students were enlisted in this dataset to participate in the study. Each student was equipped with a wireless single-channel Neurosky MindSet EEG headset. [26], which precisely measured cerebral activity over the frontal lobe. These students were tasked with watching a set of ten 2-minute-long videos. The MindSet EEG device adeptly extracted a comprehensive set of features, as outlined in Table I, to capture various aspects of cognitive responses. Notable features included "Attention," serving as a metric for the students' mental focus, and "Meditation," quantifying levels of calmness.

Additionally, the "Raw" feature represented the average of the original EEG signals. The dataset also encompassed features derived from different frequency regions of the power spectrum. These features were sampled at a frequency of 2 Hz, providing a detailed insight into students' cognitive states. It has a total of 12811 data samples. These data samples are divided into an average value of 120 depending on the length of the material that was shown to each student. This, in overall, produces a total of 100 data points, ten from each student. Each visual material used for the experiment where the dataset was gathered has a rough length of more or less than two minutes, but to represent about 60 seconds of data samples in each data point, the researchers decided to remove the first and last 30 seconds on each, collecting only the middle section of the recording of each of the visual materials are shown to each of the students involved.

Figure 2 depicts the correlation matrix of the extracted features with a colour bar, which highlights the probability of the correlation of each of the extracted features with each other. The diagonal has the possibility of one (off-white). As shown, to illustrate the correlation matrix, the hierarchical clustering method was employed. Positive correlations are displayed in light colours, and negative correlations in deep shades. The correlation coefficient values, positive or negative, are displayed on the box. On the right side, the colours are mapped with different coefficient values. As seen, attention and meditation are inversely correlated with the rest of the features, and Beta and Gamma signals showed the highest linear correlation among all the extracted features.

IV. Classification Techniques

Classification techniques play a pivotal role in the realm of brain-computer interfaces (BCIs) to analyze electroencephalogram (EEG) data [8]. The application of these techniques aids in classifying different mental states based on the acquired data [16], [27]. In this study, we utilized several baseline models, including Decision Tree, AdaBoost, Bagging, Multilayer Perceptron (MLP), Naïve Bayes, Random Forest, Support Vector Machines (SVM), XG Boost, and a specially designed deep learning model. Each of these classifiers offers different strengths, and their performance tends to vary depending on the nature of the dataset and problem at hand [28]. Our objective was to explore and compare the efficiency of these classifiers on the EEG data, with particular emphasis on our proposed deep learning model.

A. Baseline Models

This section presents an overview of the baseline models employed in our investigation. These models serve as essential benchmarks for assessing the performance of our proposed deep learning model, $ODL - BCI$, in classifying students' concentration levels from EEG data.

1) Decision Tree: Decision trees are renowned for their interpretability and simplicity [16]. They partition the dataset based on the most relevant features, constructing a tree-like structure that aids decision-making [8]. Here, we implement the $C4.5$ [29] algorithm.

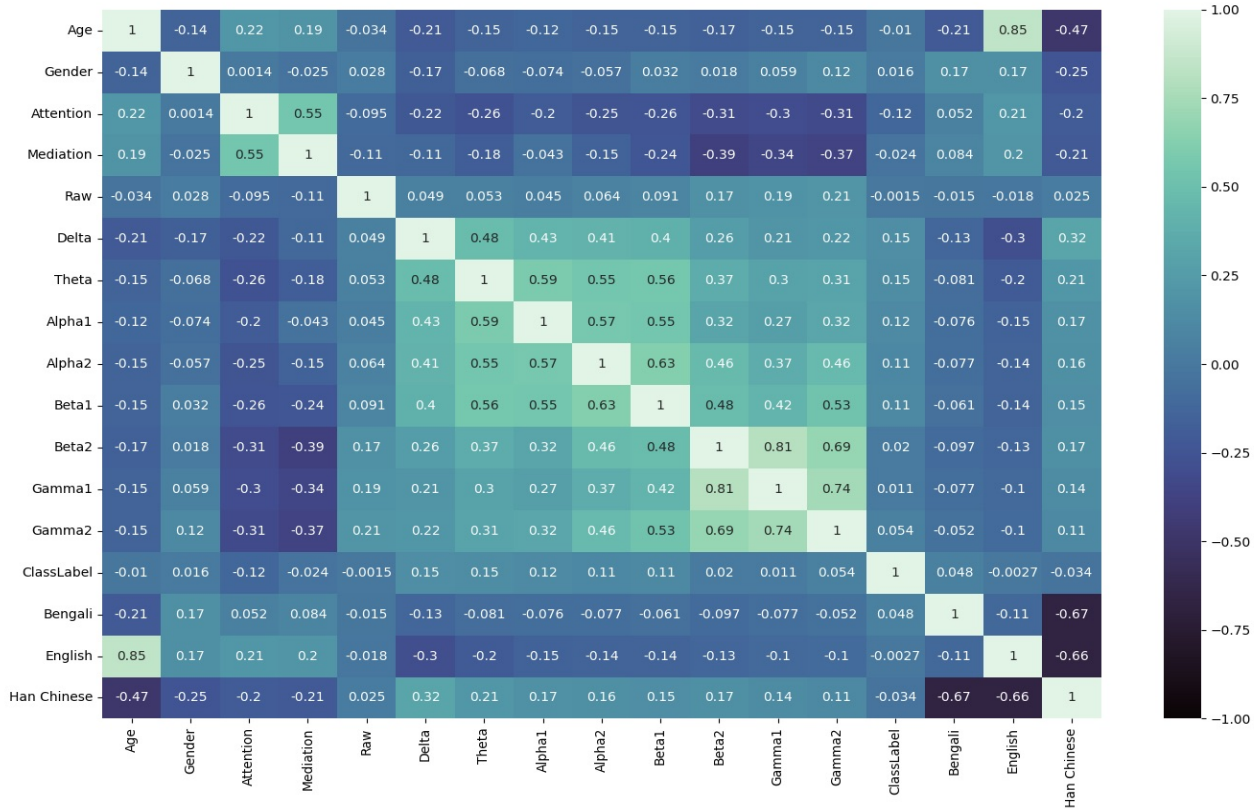

Fig. 2: Correlation matrix plot of the extracted features.

2) AdaBoost: AdaBoost is an ensemble learning technique that amalgamates the predictions of multiple weak classifiers to create a robust classifier [8]. It focuses on rectifying the errors of prior classifiers, ultimately enhancing classification accuracy [28]. We used the $C4.5$ algorithm and 100 learners to train and build a robust classifier.

3) Bagging: Bagging, or Bootstrap Aggregating, is another ensemble method that crafts numerous base models using random subsets of the dataset [27]. The amalgamation of these models results in better predictions, reducing overfitting [8]. We used $DecisionTreeClassifier$ as the base estimator.

4) Multilayer Perceptron (MLP): MLP, an artificial neural network, consists of multiple interconnected layers [8]. It excels at capturing complex patterns in data [8]. Here, we deploy MLP to gauge its performance relative to our deep learning approach. We used 100 hidden layers, activation function $relu$, $adam$ for weight optimization, and learning rate 0.0001.

5) Naïve Bayes: The Naïve Bayes classifier relies on Bayes' theorem and feature independence assumption [27]. Despite its simplicity, it demonstrates surprising effectiveness, particularly in text classification tasks [28]. We used the $MultinomialNB$ classifier to construct the model.

6) Random Forest: an ensemble method, Random Forest constructs numerous decision trees during training and combines their outputs [28]. Its robustness and suitability for high-dimensional data make it a valuable baseline [28]. We used 100 trees in the forest with $gini$ classifier and minimum samples 2 to split an internal node.

7) Support Vector Machines (SVM): SVMs are potent classifiers that aim to identify a hyperplane best suited to segregate data points from different classes [16]. They shine in high-dimensional spaces and complex datasets [28]. We implement SVM with $linearkernel$ and $predict_proba$ to enable probability estimation.

8) XG Boost: Renowned for its speed and performance, XG Boost is a gradient-boosting algorithm that builds an ensemble of decision trees iteratively, focusing on reducing predictive errors [30]. We used $binary : logistic$ for binary classification with logistic regression [30] and 100 decision trees in the ensemble with one random seed for reproducibility.

B. Proposed Deep Learning Model

Our proposed method incorporates an innovative deep-learning model with hyperparameters optimized by Bayesian optimization. It has been tailored explicitly for BCI data analysis. The below sections provide a detailed outline of the model structure, parameter selection, optimization technique, model construction, and training.

1) Model Structure and Parameter Selection: Our deep learning model architecture comprises an input layer, several hidden layers, and an output layer. The number of input features and output classes determines the nodes in the input and output layers. The architecture and parameters of the

hidden layers are meticulously selected through the Bayesian optimization process, as highlighted in Table II and III. The model features four hidden layers (H1, H2, H3, H4) with varying numbers of nodes (200, 100, 50, 16, respectively). The activation functions for these layers, selected based on their performance during the optimization process, are either Rectified Linear Units (RELU) or Parametric Rectified Linear Units (PRELU). The learning rates for each layer are set at 0.001 or 0.01, again based on the optimization process. Two nodes on the output layer represent the class label Confused and Not-Confused, which used the activation function Sigmoid to calculate the output of the nodes.

TABLE II: Optimized ODL-BCI hyperparameters.

Hyperparameters	List of Values (N)
No. of Layers	$1 \leq N \leq 6$
Nodes per Layer	$4 \leq N \leq 200$
Activation Function	Sigmoid, RELU, PRELU, ELU, Swish, Tanh
Learning Rate	0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1

TABLE III: Summary of Best Parameters.

	Grid Search	Bayesian Algorithm
Number of hidden Layer	3	4
Size of hidden Layers	150-60-20	200-100-50-16
Activation Function	PRELU, Sigmoid	RELU, PRELU, Sigmoid
Learning Rate	0.001	0.001
Time [h:m:s]	23:10:40	12:10:20
Number Of runs	42,552	21,200
Accuracy	68%	74%

2) Bayesian Optimization: The centerpiece of our hyperparameter optimization process is the implementation of Bayesian Optimization, as described in Algorithm 2. Initially, the hyperparameters are set to none. Each iteration includes sampling parameters from the search space, performing cross-validation, calculating an evaluation metric, and adding the parameters and evaluation metric to the hyperparameters set. This process continues until the top hyperparameters that yield the highest evaluation metric are obtained.

TABLE IV: The best performed hyperparameters for the deep structured model on students concentration EEG dataset.

Hidden Layers	Nodes per Layer	Activation Function	Learning Rate
Layer H_1	200	RELU	0.001
Layer H_2	100	RELU	0.001
Layer H_3	50	PRELU	0.01
Layer H_4	16	PRELU	0.001

3) Model Construction and Training: Our deep learning model's creation and training process are governed by the procedures listed in Algorithm 1. The model starts by dividing the BCI data into training and test datasets. Following this, the Bayesian optimization process generates a set of hyperparameters. The model is initialized using these parameters, trained using the training data, and evaluated on the test data. This cycle is repeated for each set of hyperparameters, with the most accurate model on the test data identified as the best model. Its hyperparameters are then stored.

The developed deep learning model, with its bespoke structure and optimization process, offers a potent approach to BCI data analysis. As confirmed by the results shown in Table V, the model surpasses traditional machine learning classifiers, demonstrating an accuracy of 74%. This underscores the viability and superiority of deep learning models optimized with Bayesian optimization in BCI applications.

Algorithm 1 Proposed deep learning model with optimized hyperparameters by Bayesian optimization

Input: BCI data D
Output: Optimized deep learning model
Method:

```
Split D into train_data and test_data;
hyperparams = BayesianOptimization();
best_accuracy = 0;
best_model = None;
best_hyperparams = None;
for params in hyperparams do
    nl = params['no_of_layers'];
    npl = params['nodes_per_layer'];
    af = params['activation_function'];
    lr = params['learning_rate'];
    model = initialize_model(nl, npl, af, lr);
    train model with train_data;
    accuracy = evaluate_model(test_data);
    if accuracy ¿ best_accuracy then
        best_accuracy = accuracy;
        best_model = model;
        best_hyperparams = params;
    end if
end for
return best_model with best_hyperparams;
```

Algorithm 2 Bayesian Optimization

Output: Optimized hyperparams
Method:

```
hyperparams = None;
for each iteration do
    Sample params from the search space;
    Perform cross-validation with params;
    Calculate evaluation_metric;
    hyperparams ∪ {params, evaluation_metric};
end for
Sort hyperparams in descending by evaluation_metric;
return top hyperparams;
```

V. Experiments

We performed a series of experiments to test the effectiveness of our proposed deep learning model. Our experimental setup, the optimized hyperparameters of the model, and the results obtained will be discussed in the following sections.

A. Experimental Setup

All experiments were performed on a MacBook Pro with an Intel Core i7 3.3 GHz Dual-Core processor and 16 GB of RAM. Python 3.8 was utilized as the primary language for coding, while the Scikit-learn 0.21.2 and TensorFlow 2.9 libraries facilitated the implementation of the machine-learning models. Spyder 3.3.1 (https://www.spyder-ide.org), a Python

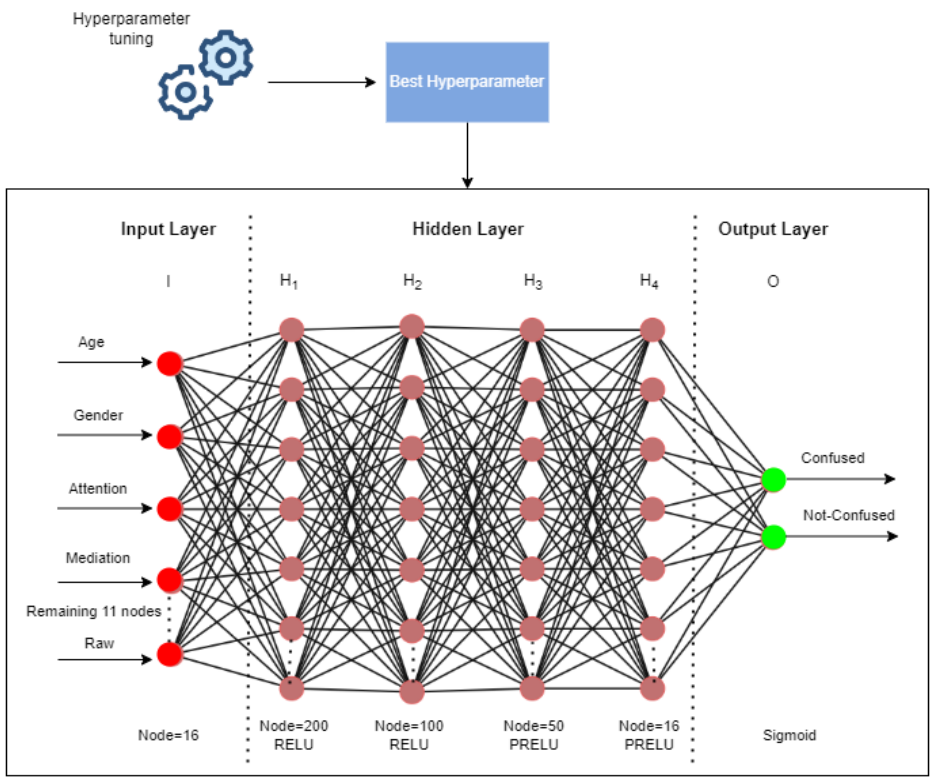

Fig. 3: The work-flow of the proposed optimal deep learning model.

development environment, was used for coding and running the models.

The performance of our model was compared against popular machine learning algorithms using classification metrics, including accuracy, precision, recall, and F-score. The accuracy is measured by Eq. 1 where $assess(x_i) = 1$ when x_i is correctly classified or $assess(x_i) = 0$ when x_i is misclassified [8]. The weighted average values of precision, recall, and F-score are considered, and the calculations are shown in Eq. 2 to 4 [28], [16]. ROC curves and AUC scores were used to visualize and quantify the trade-off between the classifiers' true positive rate (TPR) and false positive rate (FPR). In the ROC curve, the true positive rate (TPR) is plotted against the false positive rate (FPR), and the calculations are defined in Eq. 5 and 6 [16]. The lowest threshold is considered through a line, $y = x$ in the au-ROC curve where correctly classified data points represent 1 and misclassified instances reveal as 0 [28].

$$Accuracy = \frac{\sum_{i=1}^{|X|} assess(x_i)}{|X|}, x_i \in X \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - score = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

The notations TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative results, respectively [8].

B. Results

Our empirical results are embodied in Tables II through V, which jointly provide a comprehensive view of the optimal deep learning model's hyperparameter tuning, comparative performance, and preferred configuration.

We initially evaluated the performances of the ODL-BCI model with different learning parameters. These parameters

were tested using a comprehensive hyperparameter list, including the number of layers, nodes per layer, activation function types, and learning rate, as Table. II detailed.

Table III tabulates the best parameters suggested by the grid search and Bayesian optimization algorithms. The total execution time taken by the Grid Search is more than 23 hours, almost double that of the Bayesian Algorithm. Also, the number of runs for the Bayesian Algorithm is half the number for Grid Search. A primary concern is the accuracy, and our observations indicate that Bayesian optimization outperforms grid search regarding model accuracy 74% and time efficiency.

The optimal hyperparameters for the deep learning model on students' concentration EEG dataset were selected based on the best performance metrics and are illustrated in Table IV. This table elucidates the most practical combination of hidden layers, nodes per layer, activation functions, and learning rate for the proposed model. A four-layer model is proposed, employing RELU and PRELU activation functions and adopting learning rates of 0.001 and 0.01.

TABLE V: Accuracy and Average Sensitivity/Recall, Precision and F-score on EEG confusion dataset.

Classifiers	Accuracy (%)	Precision	Recall	F-score
Decision Tree	60	0.60	0.60	0.60
AdaBoost	61	0.61	0.61	0.61
Bagging	64	0.65	0.64	0.64
MLP	66	0.66	0.66	0.66
Naïve Bayes	58	0.58	0.58	0.58
Random Forest	69	0.69	0.69	0.69
SVM	58	0.58	0.58	0.58
XG Boost	67	0.67	0.67	0.67
ODL-BCI	**74**	**0.74**	**0.74**	**0.74**

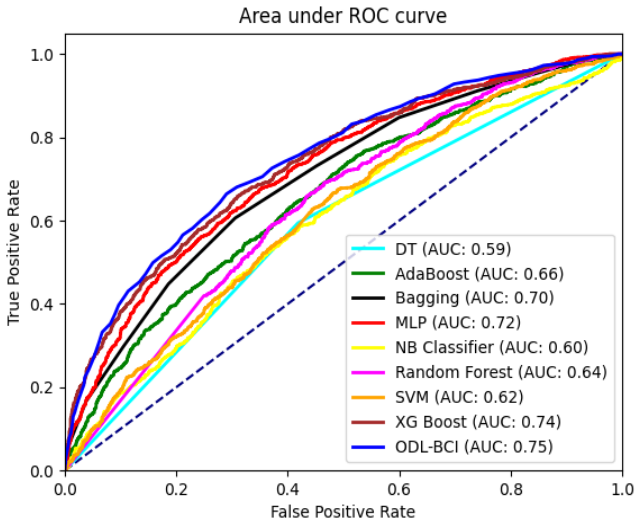

Fig. 4: ROC and AUC analysis of ODL-BCI model compared to popular classifiers on students concentration EEG dataset.

Next, the performance of the ODL-BCI model was compared with existing methods, such as Decision Tree, AdaBoost, Bagging, MLP, Naïve Bayes, Random Forest, SVM, and XG Boost. The comparative analysis of accuracy, precision, recall, and F-score of these classifiers on students' confusion level

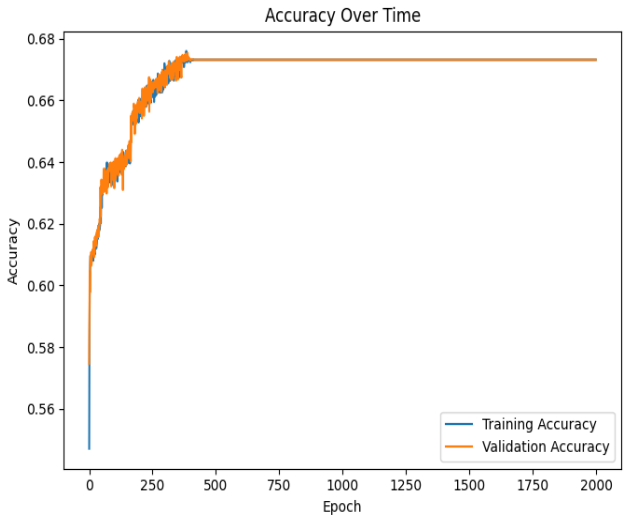

Fig. 5: Accuracy over time for Deep Learning model.

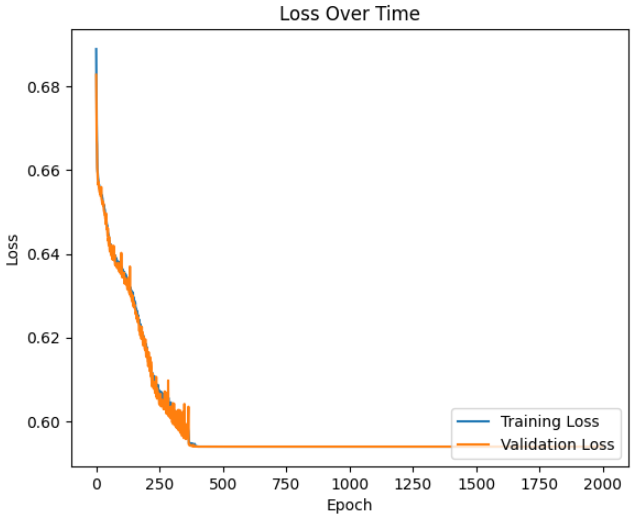

Fig. 6: Loss over time for Deep Learning model.

EEG data is presented in Table V. The proposed deep learning model significantly outperforms the existing classifiers. The ODL-BCI model achieves 74% accuracy, precision, recall, and F-score, demonstrating superior performance to the other tested classifiers. In this dataset, single classifiers failed to achieve more than 60% accuracy, whereas existing ensemble approaches can reach up to 69.0% accuracy. This table shows that ODL-BCI can achieve over 5.0% better performance than other ensemble classifiers used for this task. The decision boundaries of the proposed method compared to employed classifiers for this dataset are shown in Fig. 4.

For a more in-depth understanding of our deep learning model's behaviour, we monitored its training progress over 2000 epochs. Figures 5 and 6 depict the accuracy and loss over time. Notably, we observed that after approximately 400

epochs, accuracy and loss reached stable levels. The accuracy plateaued at around 0.67, while the loss remained below 0.60. These observations suggest that our model effectively captured the underlying patterns in the data, achieving remarkable performance.

C. Comparison with State-of-the-Art Methods

To evaluate the prowess of our proposed ODL-BCI model, we compared its performance with several state-of-the-art methods previously employed on the students' concentration EEG dataset. This comparison sheds light on the effectiveness of our novel approach in pushing the boundaries of classification accuracy.

TABLE VI: Compare with state-of-the-art methods used students concentration EEG dataset.

Classication Methods	Accuracy (%)
SVM (linear kernel) (Ni et al. [31])	67.2
Bidirectional LSTM (Ni et al. [31])	73.3
RNN-LSTM (Ni et al. [31])	69.0
Pre-defined Confusion Level (Haohan et al. [25])	67.0
User-defined Confusion Level (Haohan et al. [25])	56.0
Proposed ODL-BCI	**74.0**

As illustrated in Table VI, our ODL-BCI model achieved a remarkable accuracy of 74.0%. This accomplishment surpasses the performance of other notable methods. For instance, SVM with a linear kernel, as documented by Ni et al. [31], attained an accuracy of 67.2%. Meanwhile, Bidirectional LSTM and RNN-LSTM, also explored by Ni et al. [31], achieved accuracies of 73.3% and 69.0%, respectively. Haohan et al. [25] presented two approaches—Pre-defined Confusion Level and User-defined Confusion Level—with accuracies of 67.0% and 56.0%, respectively.

Our ODL-BCI model's substantial superiority over these state-of-the-art methods is evident, marking a significant advancement in EEG-based concentration level classification.

VI. Conclusions & Future Work

This study introduced Bayesian optimization as a robust methodology for fine-tuning hyperparameters in deep learning-based brain-computer interface (BCI) models, aiming to classify students' concentration levels using EEG data. We developed a tailored framework that combines the power of deep learning with Bayesian optimization techniques, creating a model referred to as ODL-BCI.

Our findings demonstrate that Bayesian optimization is exceptionally effective for optimizing hyperparameters in EEG-based cognitive state classification. The ODL-BCI model, enriched with Bayesian optimization, outperformed conventional machine learning classifiers and even state-of-the-art methods on the "Confused student EEG brainwave data" dataset. This model achieved an impressive accuracy of 74 percent, underscoring its potential as a valuable tool in the educational sector for real-time concentration level assessment.

Although the proposed model shows promising results, future work can further enhance its capabilities. For instance, the model can be tested with a larger dataset to verify its scalability. Incorporating other types of neural networks and more advanced optimization algorithms can also be explored to improve the model's performance. Furthermore, real-world deployment of the model in the educational sector can be examined to assess its practical utility in enhancing student concentration and learning outcomes.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] X. Zhang, Z. Ma, H. Zheng, T. Li, K. Chen, X. Wang, C. Liu, L. Xu, X. Wu, D. Lin *et al.*, "The combination of brain-computer interfaces and artificial intelligence: applications and challenges," *Annals of translational medicine*, vol. 8, no. 11, 2020.

[2] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, A. H. Khandoker, and M. Baumert, "Progress in brain computer interface: Challenges and opportunities," *Frontiers in Systems Neuroscience*, vol. 15, p. 578875, 2021.

[3] J. Jeon and H. Cai, "Multi-class classification of construction hazards via cognitive states assessment using wearable eeg," *Advanced Engineering Informatics*, vol. 53, p. 101646, 2022.

[4] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[5] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[6] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[7] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.

[8] M. O. Miah, R. Muhammod, K. A. A. Mamun, D. M. Farid, S. Kumar, A. Sharma, and A. Dehzangi, "Clusem: Accurate clustering-based ensemble method to predict motor imagery tasks from multi-channel eeg data," *Journal of Neuroscience Methods*, vol. 364, p. 109373, 2021.

[9] M. F. Kabir, T. Chen, and S. A. Ludwig, "A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction," *Healthcare Analytics*, vol. 3, p. 100125, 2023.

[10] E. Santamaría-Vázquez, a. S. P.-V. Víctor Martínez-Cagigal, D. Marcos-Martínez, and R. Hornero, "Robust asynchronous control of erp-based brain-computer interfaces using deep learning," *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106623, 2022.

[11] L. Wu, G. Perin, and S. Picek, "I choose you: Automated hyperparameter tuning for deep learning-based side-channel analysis," *IEEE Transactions on Emerging Topics in Computing*, 2022.

[12] X. Men and X. Li, "Detecting the confusion of students in massive open online courses using eeg," *International Journal of Education and Humanities*, vol. 4, no. 2, pp. 72–77, 2022.

[13] Y. N. Kunang, S. Nurmaini, D. Stiawan, and B. Y. Suprapto, "Attack classification of an intrusion detection system using deep learning and hyperparameter optimization," *Journal of Information Security and Applications*, vol. 58, p. 102804, 2021.

[14] C. Cooney, A. Korik, R. Folli, and D. Coyle, "Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech eeg," *Sensors*, vol. 20, no. 16, p. 4629, 2020.

[15] R. R. M. Reñosa, A. A. Bandala, and R. R. P. Vicerra, "Classification of confusion level using eeg data and artificial neural networks," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. IEEE, 2019, pp. 1–6.

[16] M. O. Miah, S. S. Khan, S. Shatabda, K. A. A. Mamun, and D. M. Farid, "Real-time eeg classification of voluntary hand movement directions using brain machine interface," in *The IEEE Region 10 Symposium (TENSYMP 2019) Symposium Theme: Technological Innovation for Humanity*, Kolkata, India, June 2019, pp. 534–539.

[17] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 234–239.

[18] A. Tahmassebi, A. H. Gandomi, and A. Meyer-Baese, "An evolutionary online framework for mooc performance using eeg data," in *2018 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2018, pp. 1–8.

[19] O.-A. Rușanu, "Python implementation for brain-computer interface research by acquiring and processing the neurosky eeg data for classifying multiple voluntary eye-blinks," in *5th International Conference on Nanotechnologies and Biomedical Engineering: Proceedings of ICNBME-2021, November 3-5, 2021, Chisinau, Moldova*. Springer, 2022, pp. 666–672.

[20] K. Sudarsanan and S. Sasipriya, "Controlling a robot using brain waves," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*. IEEE, 2014, pp. 1–4.

[21] O. A. RUȘANU, L. Cristea, and M. C. LUCULESCU, "The development of a bci prototype based on the integration between neurosky mindwave mobile eeg headset, matlab software environment and arduino nano 33 iot board for controlling the movement of an experimental motorcycle," 2021.

[22] K. Wang, Z. Wang, Y. Guo, F. He, H. Qi, M. Xu, and D. Ming, "An eeg study on hand force imagery for brain-computer interfaces," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2017, pp. 668–671.

[23] K. Holewa and A. Nawrocka, "Emotiv epoc neuroheadset in brain-computer interface," in *Proceedings of the 2014 15th International Carpathian Control Conference (ICCC)*. IEEE, 2014, pp. 149–152.

[24] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[25] W. Haohan, L. Yiwei, S. H. Xiaobo, Y. Yucong, M. Zhu, and K. C. Kai-min, "Using eeg to improve massive open online courses feedback interaction," in *International Conference on Artificial Intelligence in Education*, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:15915359

[26] M. Trigka, E. Dritsas, and P. Mylonas, "Mental confusion prediction in e-learning contexts with eeg and machine learning," in *Novel & Intelligent Digital Systems Conferences*. Springer, 2023, pp. 195–200.

[27] M. O. Miah, A. M. Hassan, K. A. A. Mamun, and D. M. Farid, "Brain-machine interface for developing virtual-ball movement controlling game," in *Proceedings of International Joint Conference on Computational Intelligence*. Springer, 2020, pp. 607–616.

[28] M. O. Miah, S. S. Khan, S. Shatabda, and D. M. Farid, "Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests," in *1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT 2019)*, Dhaka, Bangladesh, May 2019, pp. 1–5.

[29] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 234–239.

[30] F. Wang, Y.-C. Tian, X. Zhang, and F. Hu, "An ensemble of xgboost models for detecting disorders of consciousness in brain injuries through eeg connectivity," *Expert Systems with Applications*, vol. 198, p. 116778, 2022.

[31] Z. Ni, A. C. Yuksel, X. Ni, M. I. Mandel, and L. Xie, "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*, ser. ACM-BCB '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 241–246. [Online]. Available: https://doi.org/10.1145/3107411.3107513