IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4608~4618
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4608-4618

Detecting student attention through electroencephalography signals: a comparative analysis of deep learning models

Eng Lye Lim[1,3], Raja Kumar Murugesan[2,4], Sumathi Balakrishnan[2,3]
[1]School of Diploma and Professional Studies, Taylor's College, Subang Jaya, Malaysia
[2]School of Computer Science, Taylor's University, Subang Jaya, Malaysia
[3]Digital Health and Innovations Impact Lab, Taylor's University, Subang Jaya, Malaysia
[4]Digital Innovation and Smart Society Impact Lab, Taylor's University, Subang Jaya, Malaysia

Article Info

Article history:

Received Feb 16, 2024
Revised Apr 24, 2024
Accepted Jun 14, 2024

Keywords:

Attention detection
Deep learning
Electroencephalography
Long short-term memory
Machine learning

ABSTRACT

In the landscape of educational technology, understanding and optimizing student attention is important to enhance student's learning experience. This study explores the potential of using electroencephalography (EEG) signals for discerning students' attention levels during educational tasks. With a cohort of 30 participants, EEG data were meticulously collected and subjected to robust preprocessing techniques, including independent component analysis (ICA) and principal component analysis (PCA). The research then employed different deep learning algorithm such as long short-term memory (LSTM), recurrent neural network (RNN), gated recurrent unit (GRU), multi-layer perceptron (MLP), and convolutional neural network (CNN) classifiers to predict students' attention. The results reveal notable variations in the classifiers' predictive performance. Our finding revealed that the LSTM model emerged as the top performer and achieved 96% of the accuracy. This study not only contributes to the advancement of attention detection in educational technology but also underscores the importance of preprocessing methodologies, such as ICA and PCA, in optimizing the performance of deep learning models for EEG-based applications.

Corresponding Author:

Raja Kumar Murugesan
School of Computer Science, Taylor's University
Subang Jaya, Selangor 47500, Malaysia
Email: rajakumar.murugesan@taylors.edu.my

1. INTRODUCTION

The prevalence and significance of distant online learning have soared in recent times, propelled by technological advancements and widespread internet accessibility. This mode of education offers unparalleled convenience and flexibility, allowing learners to engage with educational content from virtually any location. Its popularity continues to rise steadily, driven by its capacity to accommodate diverse learner preferences and adapt to the changing educational landscape. Effective learning necessitates the deliberate allocation of attention to comprehend, anticipate, oversee, and regulate students' cognitive processes and behavioral responses. In order to increase student engagement and learning quality, it is crucial to identify and keep track of the students' attention [1].

Despite the importance of maintaining attention throughout a learning activity for successful learning, it is extremely difficult to determine if students are able to focus on a learning activity due to the lack of supervised processes to track their attention states. To address this, studies have explored attention-aware e-learning systems employing attention-aware models to assess and bolster students' attention levels in online learning environments [2]. Research indicates that activity levels within different frequency bands of EEG data correlate with attentional processing, reflecting its multidimensional nature. However, due to their weak electrical impulses, electroencephalography (EEG) signals require amplification for detection. Consequently, accurately measuring learners' attention levels based on EEG data remains a challenge, necessitating innovative engineering solutions. Addressing this challenge entails developing robust methodologies to effectively process EEG signals and extract meaningful information to gauge attention levels in learners [3].

Over the last 20 years, researchers have conducted a significant amount of emotion recognition research using EEG [4]. EEG is a neuroimaging technology that is often used to investigate the dynamics of neural information processing in the brain, identify brain illnesses, and assess cognitive functions. Visual assessment of large volumes of EEG data is impractical, necessitating the extraction of relevant information for accurate understanding of cognitive processes. Hence, preprocessing, feature extraction, and classification constitute vital stages in seperating the irrelevant data from EEG recordings. Among these, feature extraction plays an important role in EEG data pre processing which would directly impacting the classification performance. Selecting appropriate features is crucial to enhancing classification accuracy, underscoring the importance of employing robust feature extraction techniques tailored to EEG data characteristics. Effective feature extraction ensures optimal utilization of EEG data for accurate cognitive state assessment and classification.

Machine learning techniques have revolutionized the analysis of EGG data, unlocking new possibilities for understanding brain activity and cognitive processes. By leveraging algorithms and statistical models, machine learning can extract meaningful patterns, classify EEG signals, and make predictions based on the data [2]. Support vector machine (SVM) classifiers have been used for this purpose in the past, and there is room of improvement for the high accuracy performance [5].

Recent development of machine learning in deep learning algorithm that strive to emulate the intricate information processing mechanisms of the human brain. Deep learning models achieve this by integrating multiple layers of non-linear transformations applied to the data. Through this process, deep learning architectures generate abstract representations of the input data, ultimately yielding more comprehensive and practical insights. By simulating the brain's ability to discern complex patterns and extract meaningful features from data, deep learning algorithms enhance various tasks such as image recognition, natural language processing (NLP), and predictive analytics [6].

The objective of this research paper are: i) to review the related work of using deep learning algorithm in EEG signal analysis in the different domain, ii) to propose attention detection with EEG data experiment procedures, and iii) to conduct a comparative analysis of various deep learning classifiers to identify the one yielding the highest accuracy. Through these objectives, the paper aims to contribute to the advancement of EEG-based analysis and classification techniques in the area of student's attention detection in the online learning domain. There's increased of interest in detecting attention span using EEG data, which provide affordable, transportable, and trustworthy methods. In this part, we looked at several deep learning techniques for analyzing EGG data from different domains. Research has demonstrated that deep learning algorithms offer structural advantages by utilizing more than two hidden layers or neurons. This architecture combined with innovative training platforms has proven beneficial in various fields [7], [8]. As the volume of available data increases, deep learning algorithms excel in filling in knowledge gaps where human capabilities may be limited. Deep learning leverages its ability to discover intricate patterns and relationships within complex datasets, enabling it to uncover insights that might elude human analysis. This capability makes deep learning an invaluable tool for handling vast amounts of data and unlocking hidden knowledge across diverse domains especially in EEG data analysis [9]. Deep learning incorporates both unsupervised and supervised learning approaches. In supervised learning, data from both the input and output streams are used to train a model for making predictions. As opposed to unsupervised learning, which merely groups and interprets data depending on input.

Amin *et al*. [4] proposed a pattern recognition-based method for categorizing cognitive processes. To decompose EEG signals, the discrete wavelet transform (DWT) was used. The model applied DWT, functional data analysis (FDA), and principal component analysis (PCA) algorithms to classify the cognitive activities. Results have shown the improvement from their previous studies and this approach yields a reliable feature extraction method.

A convolutional neural network (CNN)-E classification model was created by Wen *et al*. [10] based on CNN. The model may be used to categorize EEG signals with various sample rates and can be adjusted to signals of various durations. Their research also examined potential issues with the conventional feature extraction-based classification approach when used to classify EEG data with various sampling rates.

Aliyu and Lim [11] suggest to use minimal features in the model in order to achieve higher accuracy. Hence, the study applies DWT transformation to the EEG dataset to remove noise and 20 eigenvalues features were extracted. The study applied long short-term memory (LSTM) networks as classification method. Correlation-consistent pfrojection (CCP) and PCA were then used to decrease the amount of features, making the LSTM model simpler. Feature reduction revealed that a model with just three characteristics could accurately categorize EEG data from people with epilepsy.

2. METHOD

In order to accomplish the major goals of our study, the procedure of data gathering and execution is briefly described in this part. This work aims to build the attention prediction model with the maximum accuracy. Figure 1 summarize the method and procedures of EEG data processing using deep learning model involved in this experiement.

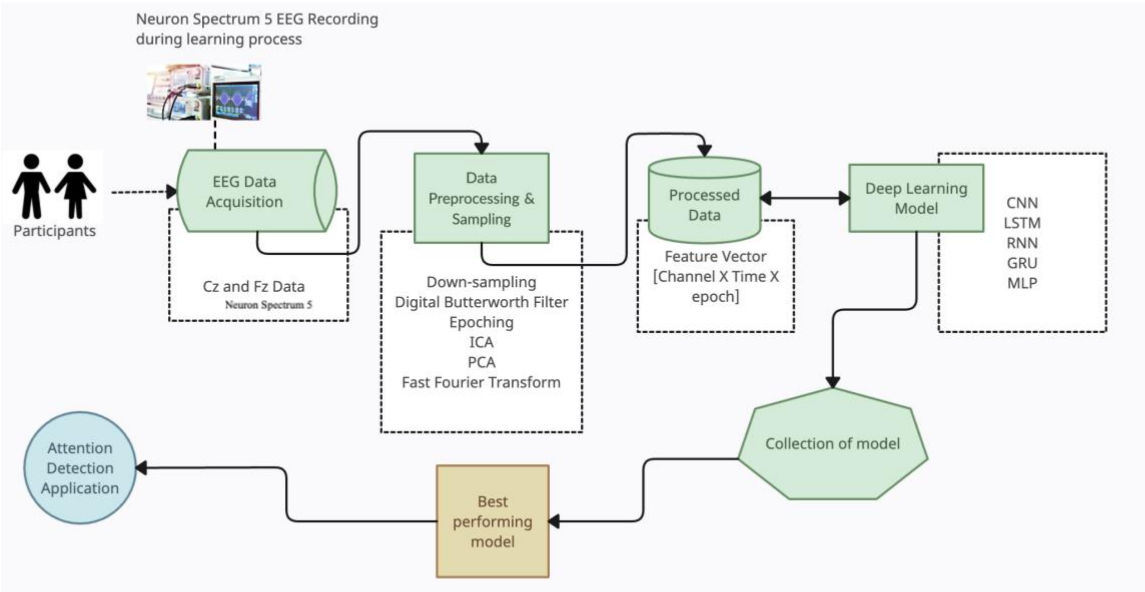

Figure 1. EEG data processing using deep learning model

2.1. Experiment setup and protocol

Neuron spectrum 5 was the equipment being used in the data collection. It is a 32-channel EEG system equipped with long-term monitoring capabilities, making it suitable for collecting EEG data over extended periods. This system enables the recording of up to 32 EEG channels, along with additional channels for recording electromyography (EMG), electrocardiogram (ECG), electrooculogram (EOG), and other physiological. The advanced signal amplification & digitization as well as the distinctive hardware design of the earlier neurosoft digital EEG devices were carried over to the neuron spectrum 5 device. The 32-channel digital EEG systems neuron spectrum 5 is flexible and extendable and has a lot of excellent polygraphic channel to capture all physiological data, from EOG to short-latency evoked potential (EP).

The participants were asked to fill up their particular at the begining of the experiment. All the participants were brief on the processes involved during the experiement and agree to the data collection. Then, the each participant will take turn to go into a quiet room for the EEG data aquisition. This is to minimize the external noise during the EEG data collection. An EEG therapist would assist during the data aquisition.

2.1.1. Data description

There were 15 males and 15 females involved in the EEG data collection. Data from Fz and Cz channels (with delta theta alpha hi-beta beta frequency) were collected from the participants. Fz data is particularly useful for monitoring cognitive processes such as attention, decision-making, and executive control, as these functions are primarily associated with the frontal lobe. Cz is often used as a reference electrode due to its central location on the scalp, providing a stable reference point for measuring electrical activity in other brain regions [12]. The data recorded with 256 Hz of the sample rate. The data collected were divided into 5 periods, which consist of eye-closed, eye-opened, watching the instructional video, eye-closed, and eye-opened. Each period consists of 3 minutes EEG recording. The eye opened state represents the participant's brain is responding to the cognitive processes that associated with the attention. The eye closed state represents the resting state. The data from period 1, 2, 4, and 5 were used for traning and testing the model and the period-3 was used to predict the attentive level of the participants.

2.2. Preprocessing

Pre-processing in machine learning refers to the modifications made to data before training a model, aiming to enhance model's performance. Precise data formatting is important for optimal model training. Different algorithms may have different in the data requirements; for instance, random forest models cannot handle null values, hence null values need to be removed from the dataset. Additionally, organizing data to allow for the simultaneous execution of various machine learning and deep learning algorithms enables selecting the best-performing model. Figure 2 depict the sample raw data collected from the data acquisition.

# mean_0_a	mean_1_a	mean_2_a	mean_3_a	mean_4_a	mean_d_0_a	mean_d_1_a	mean_d_2_a	mean_d_3_a	mean_d_4_a	mean_d_0_a2	mean_d_1_a2	mean_d_2_a2	mean_d_3_a2	mean_d_4_a2	mean_d_5_a	mean_d_6_a	mean_d_7_a	m
4.62E+00	3.03E+01	-3.56E+02	1.56E+01	2.63E+01	1.07E+00	4.11E-01	-1.57E+01	2.06E+00	3.15E+00	2.15E+00	2.95E+01	-3.53E+02	1.44E+01	2.15E+01	5.98E+00	3.07E+01	-3.43E+02	
2.88E+01	3.31E+01	3.20E+01	2.58E+01	2.28E+01	6.55E+00	1.68E+00	2.88E+00	3.83E+00	-4.82E+00	2.56E+01	3.28E+01	2.96E+01	2.15E+01	1.74E+01	2.55E+01	3.17E+01	3.15E+01	
8.90E+00	2.94E+01	-4.16E+02	1.67E+01	2.37E+01	7.99E+01	3.36E+00	9.02E+01	8.99E+01	2.03E+00	7.75E+00	3.01E+01	-4.41E+02	9.89E+00	2.53E+01	-6.89E+01	2.53E+01	-4.81E+02	
1.49E+01	3.16E+01	-1.43E+02	1.98E+01	2.43E+01	-5.84E-01	-2.84E-01	8.82E+00	2.30E+00	-1.97E+00	1.73E+01	3.20E+01	-1.48E+02	2.04E+01	2.28E+01	1.32E+01	3.15E+01	-1.47E+02	
2.83E+01	3.13E+01	4.52E+01	2.73E+01	2.45E+01	3.48E+01	-5.79E+00	3.06E+00	4.14E+01	5.52E+00	2.61E+01	3.43E+01	4.37E+01	2.37E+01	2.06E+01	-3.87E+00	3.41E+01	4.37E+01	
3.10E+01	3.09E+01	2.96E+01	2.85E+01	2.40E+01	1.65E+00	1.54E+00	3.83E+00	1.87E+00	-1.21E+00	2.80E+01	3.17E+01	2.90E+01	2.51E+01	1.75E+01	3.23E+01	2.85E+01	2.64E+01	
1.08E+01	2.10E+01	4.47E+01	4.87E+00	2.81E+01	2.14E+00	1.02E+00	1.32E+01	1.16E+00	-4.39E+00	1.07E+01	2.00E+01	2.47E+01	2.32E+00	3.33E+01	8.83E+00	2.10E+01	5.11E+01	
1.78E+01	2.78E+01	-1.02E+02	1.69E+01	2.69E+01	-3.21E+00	-1.95E+00	9.80E+00	-3.24E+00	-9.55E-01	1.96E+01	2.58E+01	-1.04E+02	1.84E+01	3.20E+01	1.92E+01	3.16E+01	-1.11E+02	
1.15E+01	2.97E+01	3.49E+01	1.02E+01	2.69E+01	-3.80E+01	-1.65E+00	3.89E+00	-3.35E+01	-3.30E+00	2.86E+01	2.86E+01	3.88E+01	2.69E+01	2.35E+01	3.24E+01	3.23E+01	2.73E+01	
8.91E+00	2.92E+01	-3.14E+02	6.51E+00	3.09E+01	-1.88E+00	1.90E+00	1.19E+01	-3.60E+00	5.70E+00	9.85E+00	2.98E+01	-3.22E+02	8.43E+00	2.93E+01	9.85E+00	2.67E+01	-3.19E+02	
5.21E+00	2.84E+01	1.85E+01	3.66E+00	2.26E+01	-1.19E-01	-2.35E+00	5.07E+00	4.61E-01	-4.26E+00	6.18E+00	3.15E+01	1.52E+01	3.11E+00	1.98E+01	4.39E+00	2.78E+01	1.68E+01	
1.33E+01	3.04E+01	-1.49E+02	1.18E+01	2.83E+01	3.03E+00	8.95E-01	-4.52E+00	1.72E+00	-6.33E-01	1.14E+01	2.94E+01	-1.58E+02	1.17E+01	2.78E+01	1.21E+01	3.06E+01	-1.35E+02	
3.01E+01	3.27E+01	2.94E+01	2.83E+01	2.43E+01	-6.05E+00	-3.94E+00	1.69E+00	-3.57E+00	6.02E+00	3.37E+01	3.51E+01	3.08E+01	3.03E+01	1.77E+01	3.26E+01	3.42E+01	2.64E+01	
1.93E+01	3.17E+01	-4.56E+00	2.38E+01	3.29E+01	-3.41E+00	6.77E-01	4.66E+01	6.80E+00	3.89E+00	8.92E+00	3.05E+01	-6.95E+01	1.76E+01	2.60E+01	3.28E+01	3.21E+01	1.27E+01	
2.88E+01	3.18E+01	3.01E+01	2.69E+01	2.18E+01	-3.55E-01	4.65E-01	3.82E+00	4.42E-01	-1.83E+00	2.64E+01	3.07E+01	3.05E+01	2.37E+01	2.10E+01	3.15E+01	3.25E+01	2.60E+01	
1.27E+01	2.38E+01	-1.93E+02	1.11E+01	2.76E+01	-3.72E+00	2.05E+00	1.98E+01	-8.04E-01	-2.87E+00	1.49E+01	2.07E+01	-1.84E+02	1.27E+01	2.56E+01	1.42E+01	2.48E+01	-2.21E+02	
4.25E-01	3.49E+01	5.39E+01	-1.19E+01	2.26E+01	1.24E+01	-2.72E+00	-4.86E+00	5.29E+00	-1.04E+00	-1.29E+01	3.86E+01	6.36E+01	-1.45E+01	2.35E+01	1.07E+00	3.40E+01	4.93E+01	
2.89E+01	3.08E+01	2.94E+01	2.74E+01	2.51E+01	-7.86E+00	1.02E+00	4.99E-01	-8.50E+00	-1.26E+00	3.31E+01	3.11E+01	2.96E+01	3.08E+01	3.23E+01	3.27E+01	2.94E+01	2.86E+01	
-1.41E+01	2.19E+01	-9.46E+02	8.93E+00	2.26E+01	2.89E-02	1.28E+00	1.64E+01	9.20E+00	4.26E+00	-1.66E+01	2.16E+01	-9.62E+02	3.29E+00	1.53E+01	-1.17E+01	2.09E+01	-9.46E+02	
8.74E+00	2.76E+01	-3.99E+02	1.50E+01	2.81E+01	-5.86E+00	-1.88E+00	9.10E+00	2.07E+00	-6.29E+00	3.62E+01	2.94E+01	-4.09E+02	1.71E+01	2.41E+01	-1.22E+01	2.78E+01	-3.99E+02	
1.03E+01	3.32E+01	-2.17E+02	1.89E+01	2.75E+01	-3.88E-01	4.90E-01	1.20E+01	2.88E+00	7.89E-01	6.15E+00	3.35E+01	-2.52E+02	1.42E+01	3.00E+01	1.46E+01	3.24E+01	-1.94E+02	
2.97E+01	3.12E+01	-8.70E+00	2.74E+01	2.66E+01	-7.23E+00	1.94E+00	-9.29E+00	-5.24E+00	3.50E+00	3.75E+01	3.11E+01	8.36E+00	3.38E+01	2.29E+01	2.92E+01	2.94E+01	-1.61E+01	
1.41E+01	2.85E+01	-1.37E+02	1.44E+01	2.74E+01	2.33E+01	-1.13E+01	1.87E+01	2.92E+01	-3.27E+00	1.59E+01	3.05E+01	-1.52E+02	1.36E+01	2.41E+01	-1.07E+01	3.78E+01	-1.40E+02	
8.53E+00	2.54E+01	-2.95E+02	6.78E+00	2.96E+01	3.20E+00	-1.84E-01	7.33E+00	3.47E+00	8.70E-01	5.21E+00	2.67E+01	-3.01E+02	4.82E+00	2.52E+01	8.61E+00	2.42E+01	-2.95E+02	
4.56E+00	3.00E+01	2.38E+01	-9.91E+00	2.15E+01	-5.14E-02	3.34E+00	3.38E+00	-1.28E+00	-3.19E+00	6.88E+00	3.05E+01	2.27E+01	-7.20E+00	2.42E+01	2.36E+00	2.62E+01	2.17E+01	
2.88E+01	3.01E+01	2.88E+01	2.78E+01	2.18E+01	3.02E+00	1.71E+00	3.57E-01	3.57E-01	3.36E+00	2.46E+01	2.89E+01	2.56E+01	2.48E+01	1.74E+01	3.00E+01	2.97E+01	3.16E+01	
1.25E+01	2.59E+01	-2.88E+02	8.16E+00	2.68E+01	-3.47E-01	6.05E-01	-2.86E+00	4.69E+00	3.98E+00	1.07E+01	2.51E+01	-2.95E+02	2.57E+00	2.08E+01	1.46E+01	2.61E+01	-2.79E+02	
2.92E+01	3.10E+01	2.92E+01	2.83E+01	2.52E+01	1.50E+01	-4.13E+00	1.88E+00	1.63E+01	-3.60E+00	-1.70E-01	4.01E+01	3.37E+01	-2.39E+00	2.78E+01	4.30E+01	2.62E+01	2.29E+01	
1.30E+01	2.92E+01	-3.07E+02	1.78E+01	2.63E+01	2.27E+00	-3.85E-01	9.82E+00	-2.04E+00	4.13E+00	1.41E+01	3.18E+01	-2.96E+02	1.85E+01	2.83E+01	9.67E+00	2.70E+01	-3.27E+02	

Figure 2. Sample of raw EEG data collected in the experiment

To create a more balanced data collection, down-sampling involves removing entries from majority classes. The most straightforward method of downsampling is to remove records from a majority class at random. A band-pass filter is a type of filter that accepts frequencies within a certain range while rejecting frequencies outside of that range. The frequency band used in this band-pass is 2 – 40 Hz. Butterworth filters are a type of signal processing filters that are designed to have a pass band frequency response that is as linear as possible. In this study, we investigate the magnitude, phase, and impulse response of the digital butterworth filter in accordance with certain requirements [13]. We explore the digital butterworth filter's performance as a band-pass filter by analyzing its magnitude, phase, and impulse response. When working with EEG data, the electrical potential levels at each electrode are relative measures. The difference in electrical potential between each electrode and a reference electrode is measured in microvolts (V) at each electrode. Electrical potential measurements always function as a relative measure, which is one of their core characteristics [14].

Before averaging, it is essential to perform several pre-processing steps to ensure that artefacts and poor channels do not affect the interpretation of event-related potential (ERP). One of the pre-processing steps involves interpolating data from good channels to substitute data from problematic channels. Another step is to reject and fix artefacts to reduce their influence. It is crucial to perform a visual assessment of a specific participant's EEG before carrying out these procedures to determine what sorts of issues are present in the dataset. While learning to do pre-processing, it is also important to spend considerable time looking at raw EEG information to learn how to recognize the most typical forms of issues [15].

The EEG signal will be undergoing epoching process. During EEG epoching, specific time-windows are extracted from the continuous EEG signal. These time-windows, also known as epochs, are usually time-locked relative to an event, such as a visual stimulus. If the EEG data are in the form of a matrix [channel x time], where time represents the entire continuous EEG signal, the resulting matrix after the epoching procedure should be [channel x time x epochs], where time is the duration of each epoch and epochs represent the number of segments extracted from the continuous EEG signal. It is important to identify the segments that need to be examined to identify epochs from the signals, such as a particular stimulus [16].

When rejecting noisy epochs, some of the artefacts that are rejected include line noise, temporal muscle activity, eye movements, and eye blinks. However, some kinds of artefacts, such as those connected to several unique scalp maps, may not be effectively rejected by independent component analysis (ICA). For example, if an individual scratches their EEG cap for a few seconds, the outcome would be a lengthy collection of slightly varied scalp maps linked to channel and wire movements. Hence, we eliminate certain kinds of "non-stereotyped" or "paroxysmal" noise before conducting ICA decomposition. Multichannel (EEG) epochs can be divided into linearly independent (temporally and spatially uncorrelated) components using (PCA).

ICA identifies paths in the feature set that correlate to high non-Gaussian projections. These paths are orthogonal in the whitersubspace even if not necessarily so in the initial space of features. On the other hand, PCA identifies orthogonal paths in the raw subspace that correspond to the directions that account for the greatest variation. ICA and PCA are comparable only if the only Gaussian processes in the data are present [17]. The graph in Figure 3 was produced using the labelled characteristics of the EEG signal from fast fourier transform (FFT) 0 b to FFT 749, which were obtained from the preprocessed dataset.

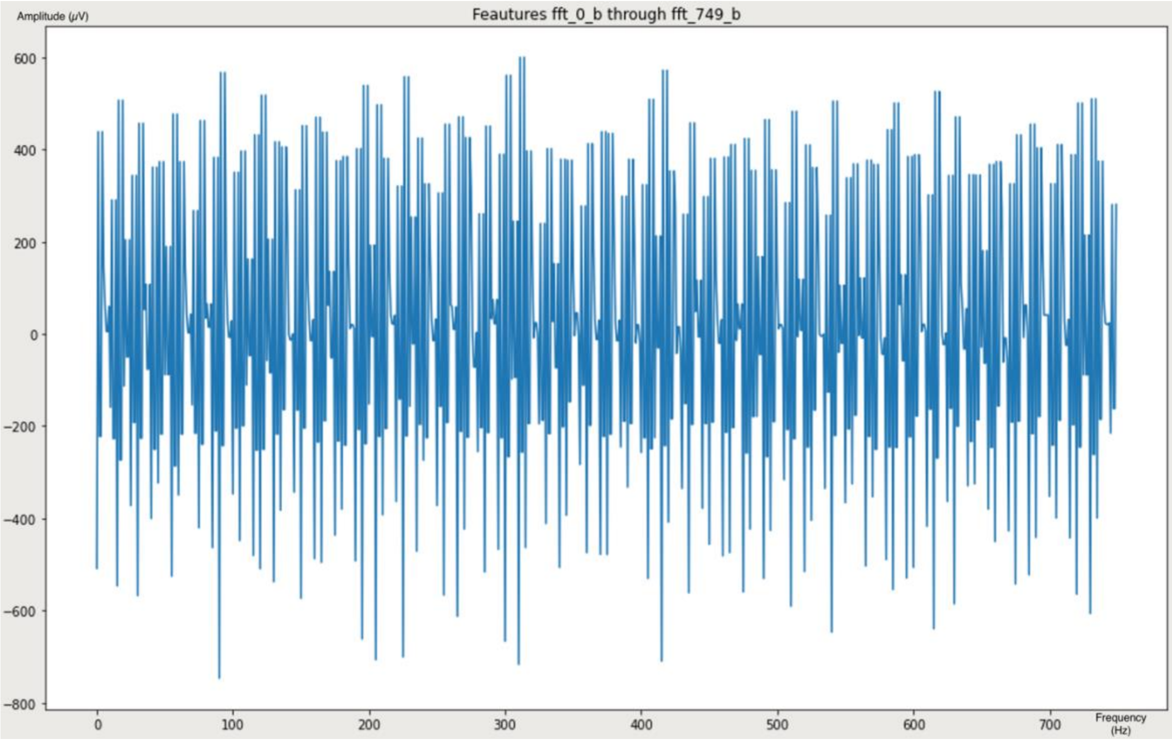

Figure 3. Plot of EEG data with time domain

2.3. Deep learning algorithm classification

The ICA-extracted features were used for classification using a variety of deep learning methods, including recurrent neural network (RNN), gated recurrent unit (GRU), LSTM, CNN, and multi-layer perceptron (MLP). 80% of the collected data is allocated for training the deep learning model. During training, the model learns the underlying patterns and relationships within the data, adjusting its parameters to minimize the prediction error. The remaining 20% of the data is used for testing model's performance. This new data helps evaluate how well the model generalizes to the new data. Testing ensures that the model has not memorized the training data but can make accurate predictions on new instances.

2.3.1. Long short-term memory classifier

LSTM networks are widely used for managing sequential input, such as sentences in NLP or prior power usage in the case of predicting electricity demand. LSTM cells have the capacity to recall both current and distant occurrences, which can have a significant impact on the predictions of dependent variables and may be utilized to gain insight into the circumstances that will lead to a subject's suitable reaction [18], [19]. The proposed LSTM model's configuration, detailed in Figure 4, was implemented using Python 3 and the

Google Colab platform, which offers GPU support for LSTM networks. The LSTM layer consists of 40 nodes, with ten nodes in the Dense 1 layer utilizing a 'tanh' activation function, while the one node in Dense 2 employs a 'sigmoid' activation function. To prevent overfitting, a dropout of 25% is applied between the LSTM layer and Dense 1 layer. Additionally, to optimize the binary-cross entropy loss function, we utilized the stochastic gradient descent (SGD) optimizer with specific parameters: a learning rate of 0.01, a learning rate decay constant of 1e-5, and a momentum constant of 0.9.

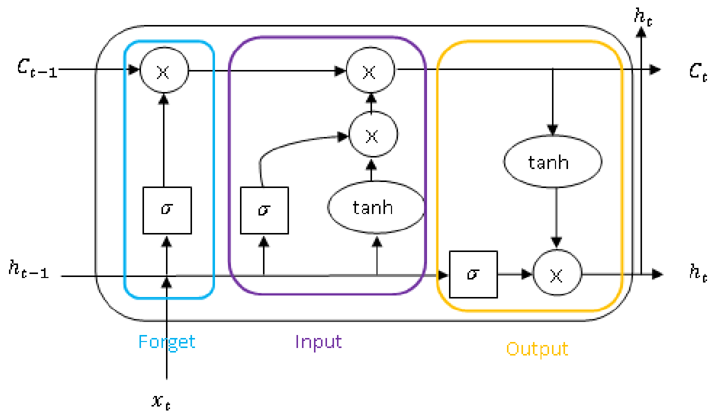

Figure 4. LSTM model

To improve the efficacy of the neural network, a four-layer neural network is constructed with rectified linear units (ReLU) activation functions in the first three layers and a sigmoid activation function in the final layer. Binary cross-entropy loss models and the Adam optimizer are employed for optimization. The neural network employs entirely interconnected layers throughout. In the process of optimization, calculated weights and threshold values serve as variables. Before deploying neural network classification design, the input is normalized. Four fully interconnected layers and defects are utilized to modify design parameters as required. Feature maps are derived using kernels. Our network performs better than LSTM-considered and standard neuro network since it just uses convolution layers and dropouts. The performance of neuro network may still be improved by pooling layers. Table 1 shows the layer type, output shape, and the param of using LSTM classification.

Table 1. The layer type, output shape, and param of LSTM

Layer (type)	Output shape	Param
Input Layer 6	2548	0
TFOpLambda	25481	0
LSTM	2548256	264192
Flatten_5	652288	0
Dense_5	3	1956867

2.3.2. Recurrent neural network classifier

The adopted RNN architecture comprises an LSTM layer, a dropout layer, a fully connected (FC) layer, and a SoftMax layer, serving as a two-category classifier. This architecture was chosen to address the limitations of ordinary RNNs, such as the vanishing gradient problem and the inability to capture long-term dependencies effectively. Hochreiter and Schmidhuber introduced the LSTM network to overcome these challenges by incorporating memory cells and gating mechanisms that enable better preservation of information over long sequences. By leveraging LSTM units within the network, the model can effectively learn and capture complex temporal patterns, making it suitable for tasks requiring memory retention and long-term dependency modeling in sequential data, such as NLP and time series prediction [20], [21]. Table 2 shows a 4-layer RNN is utilized to train the data set used for training. The total parameter count is 2,022,915, which is the sum of the RNN layer 4 parameters counts and the dense 3 parameter count. The trained data set parameters are chosen based on this.

Table 2. The layer type, output shape, and param of RNN

Layer (type)	Output shape	Param
Input Layer 6	2548	0
TFOpLambda	25481	0
simpleRNN	2548256	66048
Flatten	652288	0
Dense	3	1956867

2.3.3. Multilayer perceptron classifier

We developed a basic model mirroring the flattened input structure utilized in our previous classification model. We tested networks with both a single hidden layer and two hidden layers. Exploration involved the application of fully connected and dropout layers for each hidden layer, the latter employed to mitigate overfitting. To generate output probabilities, a softmax layer was utilized [22].

The MLP employed in this work included three layers: an input layer, a hidden layer that represented neuronal excitation via a sigmoid function, and an output layer. Each of the four states was assigned to one of the four neurons in the output layer, while the number of neurons in the input and hidden layers changed depending on the feature type used. The hidden layer's neuron count was calculated by summing the input and output layers' neuron counts, an empiricism-inducing method. For instance, when training using differential asymmetry (DASM) 12 features, the MLP's input layer included 12 neurons while the hidden layer contained 8. Back-propagation was utilized to modify the weight coefficients within the network layers based on the EEG feature vector or the associated affective label. The enhanced MLP predicted an affective designation for each EEG segment once the training procedure reached convergence. In this study, the MLP classification was performed using Weka [23], a collection of machine-learning algorithms designed for data mining. Table 3 shows the layer type, output shape, and param results of the MLP classification.

Table 3. The layer type, output shape, and param of MLP classifier

Layer (type)	Output shape	Param
Input Layer	2548	0
TFOpLambda	25481	0
Dense_20	2548256	512
Dense_21	2548128	32896
Dense_22	254864	8256
Flatten_9	163072	0
Dense_23	3	489219

2.3.4. Gated recurrent unit classifier

Nagabushanam *et al.* [24] proposed a bidirectional GRU network for EEG sequence encoding. The GRU model is a popular variant of LSTM that combines cell state or concealed state with the forgetting gate or input gate to produce a single update gate. As a result, the final GRU structure is faster and easier to implement than the conventional LSTM model. It may save a lot of time while only slightly detracting from the performance of a typical LSTM model, especially when training on large amounts of data. Both LSTM and GRU can retain essential characteristifcs through a variety of mechanisms and ensuring that these characteristics will not be lost over time. Table 4 shows the layer type, output shape, and param results of the GRU classification.

Table 4. The layer type, output shape, and param of GRU

Layer (type)	Output shape	Param
Input Layer	2548	0
TFOpLambda	25481	0
GRU	2548256	198912
Flatten_4	652288	0
Dense_4	3	1956867

2.3.5. Convolutional neural network classifier

Yin *et al.* [18] state that convolution procedures are useful for filtering out information in image processing, and CNN makes advantage of these properties to enable automated feature extraction from pictures. In this study, we implement electrode-frequency distribution maps (EFDM) with multi-channel

EEG data to feed it into a CNN for automating feature extraction and pattern detection in EEG-based emotion detection. These EFDMs may be used to do a two-dimensional convolution process as grayscale pictures [25].

For EEG emotion or attention identification, it is recommended that EFDMs be fed into a CNN containing four residual blocks. The network consists of one convolution layer, four residual blocks, four max pooling layers, two completely linked layers, and the Softmax layer. For consideration of over-fitting, the network additionally has 4 dropout layers and 5 batch normalisation layers. The maximum window size for pooling is 22 by 2, and the sliding window step size is 2. All the activation functions in the intermediate layers are ReLU. The blueprint of the discarded cube is seen in the dashed box. The residual block's convolution kernel has a size of 3, 3, 1, the slide step has a size of 1, and there is a batch normalising layer added after every convolution layer. Table 5 shows the layer type, output shape, and param results.

Table 5. The layer type, output shape, and param of CNN

Layer (type)	Output shape	Param
Input Layer_15	2548	0
TFOpLambda_14	25481	0
Conv 1D	254864	256
MaxPooling 1D	254864	0
Conv 1D	127132	6176
MaxPooling 1D	127132	0
Flatten	20320	0
Dense	3	60963

The block-based residual allows CNN to handle gradient disappearance and gradient eruption by using shortcut connections across layers. The network integrated with the max pooling layer has a certain degree of rotation and translation invariance for the input, and the max pooling layer values are shown in Table 5. Aggregation in the frequency directions may improve the neural network's capacity to extract emotional data from EFDMs since emotional aspects of EEG signals are mostly represented in their sub-frequency oscillations. Based on the properties retrieved in previous convolution layers, two fully connected layers are used at the end to classify emotions.

3. RESULTS AND DISCUSSION

Compared to previous research where the EEG data were trained based on the individual subject [26] which could led to a poor generalization of the algorithms, this study used a common-subject paradigm across subjects, assuming that the underlying patterns in the data are consistent enough across individuals to allow for shared model training. In machine learning, the loss function is used to measure how well a model is doing. During the training process, we expect the loss to decrease and accuracy to increase as the number of epochs increases. However, both loss and accuracy will be stabilized after some point [27]. To illustrate this the data loss in this study, Figures 5(a) to 5(e) show the data loss between training and test data using deep learning models in this study. The LSTM model in Figure 5(a) has a high data loss for the test data at epoch 6, but low data loss at epoch 7. The train data has low data loss at epoch 8. The RNN model in Figure 5(b) has low data loss for the test data at epoch 3 and 7, and low data loss for the train data at epoch 4 and 5. The GRU model in Figure 5(c) has a low data loss for the test data at epoch 2, 4, and 6. The train data shows low data loss at epoch 3 to 5. The MLP model in Figure 5(d) has a low data loss for the test data at epoch 2 and 5. The train data shows low data loss at epoch 3 and 5. Finally, the CNN model in Figure 5(e) has a low data loss for both the test and train data at epoch 3.

Table 6 shows the performance of different classifiers in terms of precision, recall, f score, and accuracy. These metrics are important for evaluating the performance of a classification model. Precision measures the number of positives were correctly predicted. Recall measures how many of the actual positive instances are correctly predicted as positive. F score is the harmonic mean of precision and recall. It is a measure of the balance between precision and recall. Accuracy is the number of correct predictions divided by the total number of predictions. It measures how well the model is able to correctly classify instances.

The LSTM classifier has the highest precision, recall, f score, and accuracy among all the classifiers. This indicates that the LSTM classifier is the best performing classifier among the five classifiers. The RNN and GRU classifiers also have high precision, recall, f score, and accuracy. However, the MLP and CNN classifiers have low precision, recall, f score, and accuracy. This indicates that the MLP and CNN classifiers are not performing well in this study. In summary, the table shows that the LSTM classifier is the best performing classifier among the five classifiers that achieved 96% of accuracy.

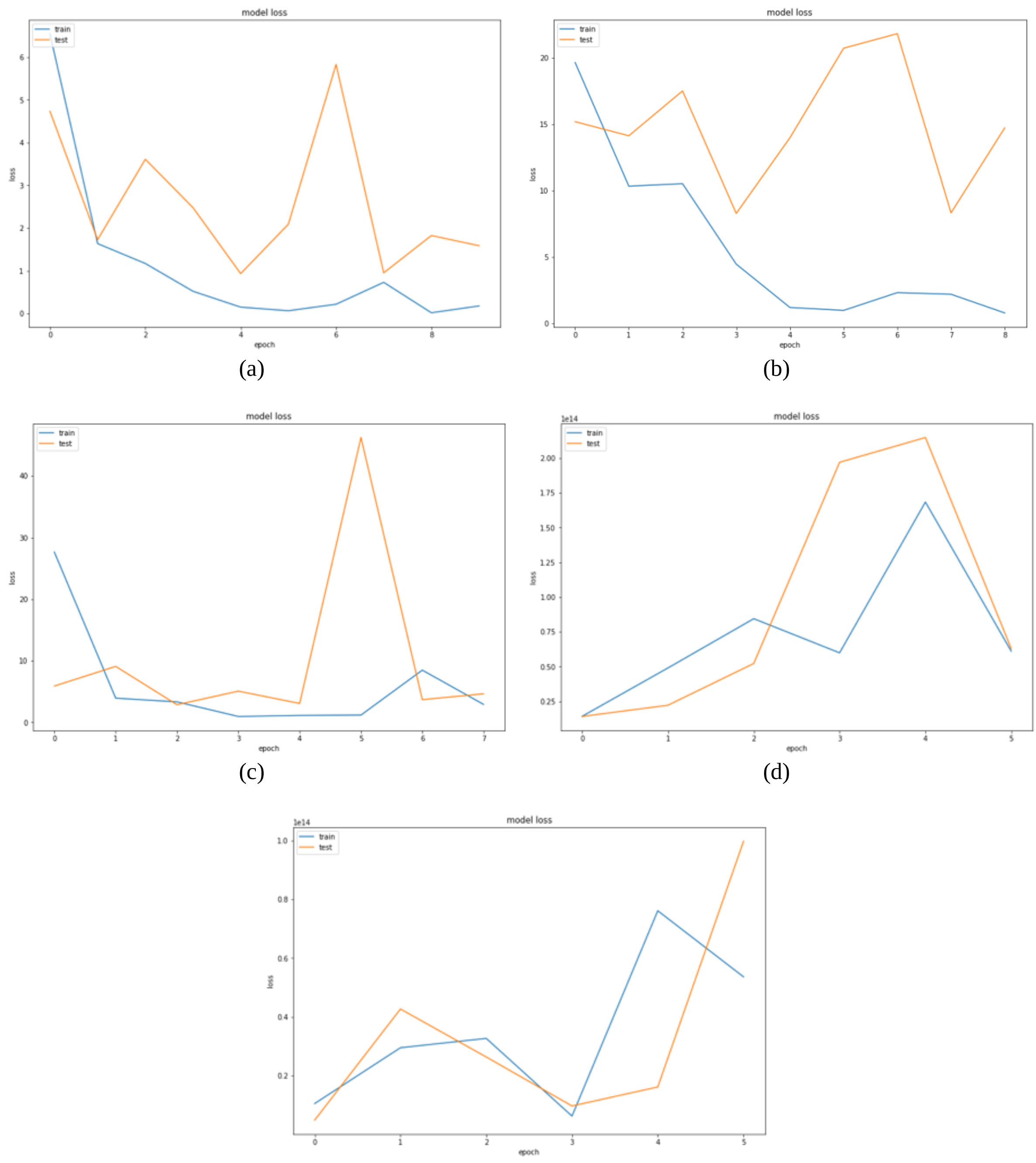

Figure 5. The test data and train data loss for, (a) LSTM, (b) RNN, (c) GRU, (d) MLP, and (e) CNN

Table 6. The performance of different classifiers

Classifier	Precision	Recall	f score	Accuracy
LSTM	0.96	0.96	0.95	0.96
RNN	0.93	0.94	0.93	0.94
GRU	0.92	0.94	0.93	0.92
MLP	0.41	0.33	0.25	0.33
CNN	0.41	0.33	0.25	0.33

To compare our model with other deep learning model which applied 14 channels in the EEG data collection, we are able to achieve similar accuracy with 96% and 96.4% [26] despite we only deloyed two channels in the EEG data collection. Besides that, we also found a similar result with the other research studies [28] that LSTM is out performed than the other deep learning model in attention detection using EEG data. Another attention detection using EEG data in the educational setting was conducted that applied e-SVM and linear discriminant analysis (LDA) approaches achieved the accuracy of 92.8% and 92.4% [29], our model using LSTM achieved higher accuracy over the traditional model.

4. CONCLUSION

The brain is the most complex and essential component of the human body. The EEG is a common electrophysiological test used to record activity in the brain's neural network under a range of stimuli. Several recent studies have employed deep learning techniques to classify various types of data and discover features. The number of research using these methods with EGG data is quite small, nevertheless. The performance of deep learning classifiers using various paradigms. This study is focused in attention detection in educational setting. RNN, GRU, LSTM, CNN, and MLP were among the other deep learning techniques we used for classification. In our study, LSTM model shows the high accuracy of 96% when compared with MLP, RNN, GRU, and CNN. Future research in attention detection using EEG data and deep learning could explore in real-time education applications and enhance in human-computer interaction domains. Moreover, interpretable models could provide valuable insights into attention mechanisms, aiding in understanding brain processes associated with attentional tasks.

ACKNOWLEDGEMENTS

The authors would like to extend their gratitude to the School of Diploma and Professional Studies, Taylor's College, and Taylor's University, Malaysia, for their unwavering support and encouragement throughout this research.

REFERENCES

[1] A. Al-Nafjan and M. Aldayel, “Predict students’ attention in online learning using EEG data,” *Sustainability*, vol. 14, no. 11, 2022, doi: 10.3390/su14116553.

[2] E. L. Lim, R. K. Murugesan, and S. Balakrishnan, “Leveraging machine learning techniques for student’s attention detection: a review,” *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 1195–1205, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1195-1205.

[3] C. Chen, J. Wang, and C. Yu, “Assessing the attention levels of students by using a novel attention aware system based on brainwave signals,” *British Journal of Educational Technology*, vol. 48, no. 2, pp. 348–369, Mar. 2017, doi: 10.1111/bjet.12359.

[4] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, “Classification of EEG signals based on pattern recognition approach,” *Frontiers in Computational Neuroscience*, vol. 11, Nov. 2017, doi: 10.3389/fncom.2017.00103.

[5] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, “Ensemble deep learning for automated visual classification using EEG signals,” *Pattern Recognition*, vol. 102, Jun. 2020, doi: 10.1016/j.patcog.2019.107147.

[6] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: A systematic review,” *Journal of Neural Engineering*, vol. 16, no. 5, Oct. 2019, doi: 10.1088/1741-2552/ab260c.

[7] H. Chen, Y. Song, and X. Li, “A deep learning framework for identifying children with ADHD using an EEG-based brain network,” *Neurocomputing*, vol. 356, pp. 83–96, Sep. 2019, doi: 10.1016/j.neucom.2019.04.058.

[8] Z. Gao *et al.*, “Complex networks and deep learning for EEG signal analysis,” *Cognitive Neurodynamics*, vol. 15, no. 3, pp. 369–388, Jun. 2021, doi: 10.1007/s11571-020-09626-1.

[9] A. Craik, Y. He, and J. L. C. -Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: A review,” *Journal of Neural Engineering*, vol. 16, no. 3, Jun. 2019, doi: 10.1088/1741-2552/ab0ab5.

[10] T. Wen, Y. Du, T. Pan, C. Huang, and Z. Zhang, “A deep learning-based classification method for different frequency EEG data,” *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–13, Oct. 2021, doi: 10.1155/2021/1972662.

[11] I. Aliyu and C. G. Lim, “Selection of optimal wavelet features for epileptic EEG signal classification with LSTM,” *Neural Computing and Applications*, vol. 35, no. 2, pp. 1077–1097, Jan. 2023, doi: 10.1007/s00521-020-05666-0.

[12] A. Natalizio *et al.*, “Real-time estimation of EEG-based engagement in different tasks,” *Journal of Neural Engineering*, vol. 21, no. 1, Feb. 2024, doi: 10.1088/1741-2552/ad200d.

[13] Q. Zhong, Y. Zhu, D. Cai, L. Xiao, and H. Zhang, “Electroencephalogram access for emotion recognition based on a deep hybrid network,” *Frontiers in Human Neuroscience*, vol. 14, Dec. 2020, doi: 10.3389/fnhum.2020.589001.

[14] C. Y. Liao, R. C. Chen, and S. K. Tai, “Emotion stress detection using EEG signal and deep learning technologies,” in *4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018*, Apr. 2018, pp. 90–93, doi: 10.1109/ICASI.2018.8394414.

[15] T. Donoghue and B. Voytek, “Automated meta-analysis of the event-related potential (ERP) literature,” *Scientific Reports*, vol. 12, no. 1, Feb. 2022, doi: 10.1038/s41598-022-05939-9.

[16] Pawan and R. Dhiman, “Machine learning techniques for electroencephalogram based brain-computer interface: A systematic literature review,” *Measurement: Sensors*, vol. 28, Aug. 2023, doi: 10.1016/j.measen.2023.100823.

[17] X. Xu, X. Nie, J. Zhang, and T. Xu, “Multi-level attention recognition of EEG based on feature selection,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, Feb. 2023, doi: 10.3390/ijerph20043487.

[18] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, “EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM,” *Applied Soft Computing*, vol. 100, Mar. 2021, doi: 10.1016/j.asoc.2020.106954.

[19] J. X. Chen, D. M. Jiang, and Y. N. Zhang, “A hierarchical bidirectional GRU model with attention for EEG-based emotion classification,” *IEEE Access*, vol. 7, pp. 118530–118540, 2019, doi: 10.1109/ACCESS.2019.2936817.

[20] K. K. Dutta, "Multi-class time series classification of EEG signals with recurrent neural networks," in *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, Jan. 2019, pp. 337–341, doi: 10.1109/CONFLUENCE.2019.8776889.

[21] Q. Ma, M. Wang, L. Hu, L. Zhang, and Z. Hua, "A novel recurrent neural network to classify EEG signals for customers' decision-making behavior prediction in brand extension scenario," *Frontiers in Human Neuroscience*, vol. 15, 2021, doi: 10.3389/fnhum.2021.610890.

[22] R. Atangana, D. Tchiotsop, G. Kenne, and L. C. DjoufackNkengfac k, "EEG signal classification using LDA and MLP classifier," *Health Informatics - An International Journal*, vol. 9, no. 1, pp. 14–32, Feb. 2020, doi: 10.5121/hiij.2020.9102.

[23] M. Murtazina and T. Avdeenko, "Classification of brain activity patterns from EEG data using WEKA," in *7th IEEE International Conference on Information Technology and Nanotechnology*, Sep. 2021, pp. 1–5, doi: 10.1109/ITNT52450.2021.9649270.

[24] P. Nagabushanam, S. T. George, and S. Radha, "EEG signal classification using LSTM and improved neural network algorithms," *Soft Computing*, vol. 24, no. 13, pp. 9981–10003, Jul. 2020, doi: 10.1007/s00500-019-04515-0.

[25] F. Wang *et al.*, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, doi: 10.1016/j.neuropsychologia.2020.107506.

[26] D. Zhang, D. Cao, and H. Chen, "Deep learning decoding of mental state in non-invasive brain computer interface," *ACM International Conference Proceeding Series*, pp. 1-5, 2019, doi: 10.1145/3371425.3371441.

[27] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, vol. 9, no. 2, pp. 187–212, Apr. 2022, doi: 10.1007/s40745-020-00253-5.

[28] P. Kaushik, A. Moye, M. V. Vugt, and P. P. Roy, "Decoding the cognitive states of attention and distraction in a real-life setting using EEG," *Scientific Reports*, vol. 12, no. 1, Nov. 2022, doi: 10.1038/s41598-022-24417-w.

[29] M. Alirezaei and S. H. Sardouie, "Detection of human attention using EEG signals," in *2017 24th Iranian Conference on Biomedical Engineering and 2017 2nd International Iranian Conference on Biomedical Engineering, ICBME 2017*, Nov. 2018, pp. 1–5, doi: 10.1109/ICBME.2017.8430244.

BIOGRAPHIES OF AUTHORS

Eng Lye Lim received his B.Sc. in Computer Science from Universiti Teknologi Malaysia (UTM) in 2002 and his Master of Computer Science from Universiti Malaya (UM) in 2009. His research interests include EEG, machine learning, deep learning, natural language processing, and adaptive learning. His current research is in student's attention detection using EEG. He is currently a Ph.D. student in Taylor's University. He can be contacted at email: englye.lim@taylors.edu.my.

Raja Kumar Murugesan is an Associate Professor of Computer Science, and Head of Research for the Faculty of Innovation and Technology at Taylor's University, Malaysia. He has a Ph.D. in Advanced Computer Networks from the Universiti Sains Malaysia. His research interests include IPv6, future internet, internet governance, computer networks, network security, IoT, blockchain, machine learning, and deep learning algorithms. He is a senior member of the IEEE and IEEE Communications Society, member of the Internet Society (ISOC), member of the Malaysia Data Science Association and associated with the IPv6 Forum, Asia Pacific Advanced Network Group (APAN), Internet, and Malaysia Network Operator Group (MyNOG) member's community. Talks about IoT, Blockchain, AI, machine learning, IPv6, internet governance, digital transformation in the context of industry 4.0 (IR4.0), sustainable development goals (SDG's), and futures thinking at various international conferences, forums, and events. Obtained substantial funds locally and internationally, and actively engages in research, consultancy, and outreach activities in advanced computer networks. He can be contacted at email: rajakumar.murugesan@taylors.edu.my.

Sumathi Balakrishnan attained her Master of Software Engineering from University of Malaya, Malaysia, in 2009. She graduated with a Bachelor of Computer Science from the University Putra Malaysia. Currently, she has been working with Taylor's University, Malaysia as a senior lecturer. While teaching she is pursuing her research for her Ph.D. Her research areas involve activity recognition, depression detection using EEG, smart home and IoT, and applications for elderly activity monitoring. Many papers have been published based on her research interest. She can be contacted at email: sumathi.balakrishnan@taylors.edu.my.