Computational and Structural Biotechnology Journal 23 (2024) 1450–1468

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

# Depression assessment using integrated multi-featured EEG bands deep neural network models: Leveraging ensemble learning techniques

Kuo-Hsuan Chung [a,b], Yue-Shan Chang [c,*], Wei-Ting Yen [c], Linen Lin [d], Satheesh Abimannan [e]

[a] Department of Psychiatry and Psychiatric Research Center, Taipei Medical University Hospital, Taipei, Taiwan
[b] Department of Psychiatry, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[c] National Taipei University, Sanxia District, New Taipei City 237, Taiwan
[d] Department of Psychiatry, En Chu Kong Hospital, Taiwan
[e] Amity School of Engineering and Technology, Amity University Maharashtra, Mumbai, India

A R T I C L E  I N F O

*Keywords:*
Mental status assessment
Ensemble learning
Multi featured deep learning
EEG signal
Deep neural network

A B S T R A C T

Mental Status Assessment (MSA) holds significant importance in psychiatry. In recent years, several studies have leveraged Electroencephalogram (EEG) technology to gauge an individual's mental state or level of depression. This study introduces a novel multi-tier ensemble learning approach to integrate multiple EEG bands for conducting mental state or depression assessments. Initially, the EEG signal is divided into eight sub-bands, and then a Long Short-Term Memory (LSTM)-based Deep Neural Network (DNN) model is trained for each band. Subsequently, the integration of multi-band EEG frequency models and the evaluation of mental state or depression level are facilitated through a two-tier ensemble learning approach based on Multiple Linear Regression (MLR). The authors conducted numerous experiments to validate the performance of the proposed method under different evaluation metrics. For clarity and conciseness, the research employs the simplest commercialized one-channel EEG sensor, positioned at FP1, to collect data from 57 subjects (49 depressed and 18 healthy subjects). The obtained results, including an accuracy of 0.897, F1-score of 0.921, precision of 0.935, negative predictive value of 0.829, recall of 0.908, specificity of 0.875, and AUC of 0.8917, provide evidence of the superior performance of the proposed method compared to other ensemble learning techniques. This method not only proves effective but also holds the potential to significantly enhance the accuracy of depression assessment.

## 1. Introduction

Depression is categorized as a mental disorder characterized by a persistent low mood, behavioral alterations, negative thoughts, and decreased motivation, leading to challenges in carrying out daily activities. A study conducted by Brown University and Boston University revealed that depression among American adults persisted and worsened during the early stages of the COVID-19 outbreak. A supplementary report [1] further confirmed this trend, indicating that the prevalence of severe depression symptoms among American adults rose from 27.8% in 2020 to 32.8% in 2021.

Earlier studies [2–5] have addressed this issue by examining the detection of biofeedback using biological signals, without employing objective measurements, Electroencephalography (EEG) has been identified in recent research studies [6–8] as a potential tool to measure emotional state or depression. The device was utilized in various studies [9–11] to explore medical subjects and verify their accuracy in detecting the brainwave signals [12].

In the field of Deep Learning applications, various network model architectures have been developed earlier [13], such as RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), etc. These models, especially for EEG-related applications, are widely applied in classification and prediction processes in different fields. The researchers conducted several studies using Deep Learning (DL) methods to train EEG signals and build predictive models [11,14–16]. EEG signal is characterized as a time series wave, where each EEG frequency band implies a certain feature. The signals are typically presented in 5-dimensional EEG such as *delta wave, theta wave, alpha wave, beta, and gamma wave*. Each one has its own characteristics regarding emotional or mental state [17,18]. Therefore, it may be insufficient to assess an individual's mental state based solely on a specific frequency band.

* Corresponding author.
*E-mail address:* ysc@mail.ntpu.edu.tw (Y.-S. Chang).

https://doi.org/10.1016/j.csbj.2024.03.022
Received 8 November 2023; Received in revised form 24 March 2024; Accepted 25 March 2024
Available online 1 April 2024

Ensemble learning is a well-known and commonly used method in machine learning. It involves integrating multiple weak learners into a strong learner in order to improve classification accuracy [19,20]. It has been used in various machine learning (ML) and deep learning techniques to integrate different types of models. Previous literature [21–24] introduced ensemble methods as a means to incorporate multiple features into machine learning models. The results achieved in these studies confirmed that the prediction accuracy can be enhanced significantly.

In the current study, a multi-tier ensemble learning technology is proposed to integrate multiple EEG bands and assess the mental state or depression of an individual. In this study, the EEG is subdivided into 8 distinct sub-bands, with each sub-band showcasing unique characteristics related to emotion and mental state. Therefore, we train the model using an LSTM network, and subsequently employ ensemble learning to integrate models across equivalent numbers of sub-bands, ranging from 1 to 8 bands, for enhanced accuracy. These combinations may also be relevant to the mental state and emotion. The results of these models are ensemble together, representing the outcome of initial tier ensemble learning. Finally, the results of these multiple same sub-band models are integrated using ensemble learning to form second-tier ensemble learning.

The major goal of this research is to determine ways for interpreting EEG data using a neural network and determining whether the outcomes indicate depression in an individual. In order to achieve the objectives, the EEG frequency bands are collected as per the literature [23]. The EEG signals are participated into eight sub-bands such as delta, theta, low-alpha, high-alpha, low-beta, high-beta, low-gamma, and high-gamma respectively. Afterwards, the frequency domain is arranged in an increasing order and the signals are subsequently categorized into 1-band signal, 2-bands signal, …, and 8-bands signal. These signals are used to combine $N$ adjacent sub-bands into $N$-bands signals. Following that, the model is trained individually for each band using a LSTM-based Deep Neural Network (DNN). Multiple Linear Regression (MLR)-based two-tiers' ensemble learning technology is utilized herewith to integrate the multi-bands' EEG frequency models and assist in the evaluation of mental state or depression. The primary objective is to combine LSTM models representing a range of bands (1–8 bands) in the first tier, followed by integrating the results obtained from the first tier in the second tier.

To demonstrate the effectiveness of the proposed method, the authors conducted the experiments and used different measures to validate the outcomes. In this study, various ensemble learning methods are used for comparison purposes. These include *simple average approach*, *ACC weight* method, *MAE weight* method, and *hard voting priority* method. The results revealed that the proposed method outperformed other ensemble learning methods.

The contributions of the current study are as follows.

1. A multi-band and multi-featured Deep Learning model is proposed for the evaluation of mental status or depression.
2. A multi-tier ensemble learning method is applied to integrate a variety of feature-associated Deep Learning models.
3. Based on the comparison among different ensemble learning methods, it can be concluded that the proposed multiple learning regression method is superior to other techniques.

The organization of the rest of the paper is as follows. Section II presents the background and the works related to the domain. In section III, the information on how to collect and process the EEG signals such as splitting, combining, and preprocessing is presented. Section IV explains the design of N-Band LSTM-based base classifiers. Section V depicts the architectural design of two-tier ensemble learning. The experimental procedures and the key findings of the research are presented in section VI. In section VII, the concluding remarks and the scope for future work are discussed.

## 2. Background and related work

In this section, we will first present some background knowledge. It consists of the EEG frequency band relevant to emotional mental status and the concept of ensemble learning. Subsequently, we will conduct a comprehensive review of prior research related to the proposed study.

### 2.1. EEG frequency band relevant to emotional and mental status

In the field of psychiatry, it is important to determine the mental state of an individual before initiating treatment. Several studies [21,25,53] suggested methods to measure the emotional state using EEG waves. In previous literature [26], the authors provided a thorough explanation of the methodology for calculating the scores that instantaneously evaluate the mental state of an individual using a simple EEG headband. In [27], the authors used features such as intra-channel information and inter-channel information for the classification of different workloads based on electroencephalogram signals.

It is widely recognized that EEG measurements obtain five different kinds of frequency bands such as *delta, theta, alpha, beta,* and *gamma*. The signals of these frequency bands usually correspond to certain emotional characteristics. For instance, *theta* signal [28] is associated with the inhibition of elicited responses whereas a high range of *beta* signal is associated with anxiety [29]. In the previous study [30], the authors investigated the impaired functional connectivity at *alpha* and *theta* frequency bands of EEG during a major depression episode. The analysis results confirmed the study hypothesis, indicating an observe in brain functional connectivity during a major depression. The authors [18] revealed that the mean percent *delta* power/min is associated with reduced anxiety and hostility, and increased energy. In literature [31], the authors found that the patients who suffer from Generalized Anxiety Disorder (GAD) express highly negative emotions when they are worried. In such instances, EEG *gamma* band is helpful in monitoring the fluctuations in pathological worry that are expected to follow a successful treatment.

According to the authors [32], the baseline *gamma* power is significantly associated with a change in depression severity, whereas the *gamma* power can be utilized as a potential biomarker to predict antidepressant response. In literature [33], the authors established that the global EEG coherence of patients with depression was significantly higher than that of the healthy controls in both *gamma* bands, especially in *high gamma* band. Therefore, it can be concluded that these frequency bands have varying degrees of relationship with mental state or depression.

To summarize the aforementioned points, different frequency bands characterize different psychiatric disorders. In such a case, if only one frequency band is used for the assessment of mental status or depression, it may prove insufficient and insignificant.

Additionally, the bandwidths of the five frequency bands differ from each other. As depicted in Table 1, the frequency range of the delta wave is 0.1–3 Hz, while the gamma wave has a frequency range of 25–60 Hz. In the current study, certain bands, including alpha, beta, and gamma bands, are characterized by a wide range. Consequently, to mitigate the impact of this wide range of waves, some of these bands have been

**Table 1**
The 8-bands EEG frequency domain.

| Notation of Brain Waves | Band | Frequency (Hz) |
|---|---|---|
| $\boldsymbol{\delta}$ | delta | 0.1 ~ 3 Hz |
| $\boldsymbol{\theta}$ | theta | 4 ~ 7 Hz |
| ***Low-α*** | Low-alpha | 8 ~ 9 Hz |
| ***High-α*** | High-alpha | 12 ~ 14 Hz |
| ***Low-β*** | Low-beta | 12.5 ~ 16 Hz |
| ***High-β*** | High-beta | 20.5 ~ 28 Hz |
| ***Low-γ*** | Low-gamma | 25 ~40 Hz |
| ***Mid- γ*** | Mid-gamma | 40~ 60 Hz |

divided into two sub-bands. It is assumed in this study that the frequency bands of adjacent EEGs have a certain relationship with each other, thereby increasing the selection of training features. In literature [53], the authors collected the EEG signals with five bands, whereas in the current study, the signals were collected and split into eight bands as given; *delta* wave, *theta* wave, *low-$\alpha$* wave, *high-$\alpha$* wave, *low-$\beta$* wave, *high-$\beta$* wave, *low-$\gamma$* wave, and *mid-$\gamma$* wave, as shown in Table 1.

### 2.2. Ensemble learning

Ensemble learning occurs when different types of independent models are integrated together to enhance the prediction accuracy. The fundamental premise of ensemble learning is that multiple learners tend to learn with high efficiency compared to a single learner [31]. The following is an overview of some of the frequently employed techniques in this domain.

Simple Average, otherwise known as 'Soft Voting' [32], is used to aggregate the outcomes of multiple ML classifiers and add the probabilities of multiple prediction outcomes in order to select the probability and the largest category after learning. Though *Hard Voting* [33,34] is similar to soft voting, it is decided mainly based on majority voting with equal weights for every classifier. The result can be obtained if a major number of certain objects are present. The *MAE (Mean Absolute Error) weighted method* is another indicator that is employed to evaluate the results [35]. MAE refers to the average value between the predicted value and the actual value. The definition of MAE involves the selection of weighting components in ensemble learning weighting formulas to achieve an error-minimizing solution. The model with high accuracy is given higher weight to optimize the overall model. The ACC (Accuracy) weighted method is an alternative approach, akin to the MAE weighted method The ACC weighted method is learned from the SWE (Self-Weighted Ensemble) method [36] which is based on the F1-score to calculate the performance. The aim of the ACC weighted method is to reduce the weight of inaccurate learners and boost the reasonably-accurate weak learners so as to increase the recognition rate of the overall model. Therefore, the accuracy of the weak learner is evaluated to assign the desired amount of weight.

### 2.3. Related work

In literature [21], the authors proposed an Ensemble Learning (EL) technique to integrate multi-featured Deep Learning (DL) models to predict air pollution. In the study conducted earlier [22], an ensemble learning-like model was utilized to combine the air quality prediction models with spatiotemporal features of three different regions. The studies mentioned above used ensemble learning techniques to predict the air quality.

In literature [37], the author employed three traditional ML algorithms to construct the base classifier for EEG signals. Subsequently, these classifiers were combined using ensemble learning techniques, resulting in excellent model performance. This study utilized traditional ML models in mental imagery tasks. In their research [38], the authors first analyzed various features and employed various traditional ML methods such as SVM and KNN to construct classification models. They then utilized F-test and ReliefF methods, followed by comparing accuracy values with various conventional ML methods.

In a previous study [39], two modalities, namely EEG signals and sound, were used as features to detect depression. Each modality provided training to six classifiers, and these classifiers were combined using fusion technology, showing potential for improving recognition accuracy. In literature [40], the authors proposed a new Deep Learning network architecture called FLDNet, which initially designed three networks and integrated them later with an ensemble layer. This network can automatically learn to perform emotion recognition with high-level features.

In another study [41], the authors proposed a multi-domain ensemble CNN called ensCNN-MD to enhance the accuracy of fatigue recognition. This work utilized a combination of seven different features, and the classification models of these CNNs were integrated using the bootstrap aggregation technique. Additionally, the authors [42] proposed a hybrid model for recognizing human emotions using CNN and LSTM networks. In this model, five frequency bands of EEG were fed into three developed models initially, and the outcomes of these base models were integrated using an ensemble model.

Following existing work, all are utilizing EEG signals to construct depression prediction models. In the work [43], authors adopted CNN as the technique to train the depression prediction model. The EEG data set was 4348 records collected from 15 normal and 15 depressed patients. In the literature [44], the authors did not exploit deep learning techniques for model training. Instead, they employed spectral asymmetry index and nonlinear, detrended fluctuation analysis to differentiate the depressed and healthy subjects. The EGG data set was collected from 17 depressed patients and 17 control subjects. The work [45] utilized CNN and 1DCNN+LSTM network model to train the depression prediction model. The EEG data were collected from 33 depressed and 30 healthy subjects. In the literature [54], the authors exploited CNN to train the model for depression discrimination, named HybridEEGNet. This model consists of two parallel lines designed to learn the synchronous and regional EEG features. In the work [46], the authors combined CNN and LSTM for depression detection, named DepHNN, in which CNN is used for temporal learning while LSTM is used for the sequence learning process. The EEG signals were obtained from 45 subjects. The work [47] proposed a CNN based depression detection model. The EEG signals are mainly 4 frequency bands using 19 electrodes from 46 normal and 46 depressed subjects. In the literature [48], a CNN-based SparNet depression discrimination model was proposed to learn EEG space-frequency domain characteristics. The study involved the collection of EEG signals from 48 subjects, with 24 normal and 24 depressed subjects. The work [49], a CNN-based depression detection model, named DeprNet, was proposed. The EEG signals were collected from 33 subjects with 19 electrodes. In the literature [50], the authors proposed a CNN-based Inception Time model for depression detection via 19-channel raw EEG signals. The EEG signals were collected from 30 healthy and 34 depressed subjects.

In the literature [23], the authors proposed MSFBEL (Multi-Scale Frequency Bands Ensemble Learning) method to identify four types of emotions such as *happy*, *normal*, *sad*, and *fear* from EEG signals. At first, five frequency bands of the EEG signals were divided into multiple data sets after which it was rearranged using multiple scales. Then, a basic classifier was constructed for each scale to recognize the emotions. The output of all the classifiers were combined using simple majority voting-based ensemble learning in order to integrate the classifiers.

As stated in the earlier research, a few common problems have been identified: (1) some studies utilize traditional machine learning methods to build EEG classifier instead of utilizing Deep Learning neural network; (2) some researchers used ML techniques to identify different types of emotional states rather than assess the level of depression; (3) and some investigations did not exploit ensemble learning method to integrate multiple classifiers; and (4) some studies used ensemble learning though not for assessing the depression levels. In light of the research gap, the current study proposes a multi-tier ensemble learning method that involves the integration of multiple LSTM-based base classifiers, for training various N-bands EEG signals and assessing depression.

## 3. System flow and signal preprocessing

This section describes the system flow and shows the signal preprocessing employed, including the splitting, combining and preprocessing of EEG frequency bands.

### 3.1. System flow

The step-by-step operations is shown in Fig. 1 and is explained in the following.

1. First, we use a single-channel EEG to collect the patient's brain waves, and transmit the EEG to the back-end server through the APP. The collection of EEG by this APP can be explained in Section 6.1.
2. The backend server will process the collected data using the following steps:
   a. Receive signals.
   b. Outlier detection and missing value padding. Details please refer to Section 3.2.
   c. Signals splitting and combining. Details please refer to Section 3.3.
   d. Store signals into the backend database.
3. Training all N-bands LSTM classifier models using LSTM neural network. The network architecture of LSTM will be explained in Section 4.1, and training all N-bands models, the details are presented in Section 4.2.
4. Training first-tier ensemble learning models using various techniques. Please refer to Section 5.1.1 for these detailed operations.
5. Training second-tier ensemble learning models using various techniques. Details please refer to Section 5.1.2.
6. Evaluating the performance of models. Please refer to Section 5.2 for various evaluation indicators and results.

### 3.2. Data pre-processing

Data processing is an important step to collect the signals from EEG devices. It deals with how to feed the processed data into a neural network.

The data collected by the front-end device is then stored in the database. It has a total of 600 s of raw data under 8 bands.

The outliers of the observed target may come from sensor noise, static human body, communication interference and other factors that cannot be grasped. In the current work, the well-known statistical method, i.e., IQR (interquartile range) [52] is exploited. To replace the value of the outlier, a linear interpolation method is used.

The EEG signal obtained using the general EEG sensor demonstrates a considerable amount of noise. The aim of this work is to propose an ensemble multi-tiers multi-featured EEG-bands deep learning models for depression assessment. To effectively get the research results, a commercially available EEG sensor (NeuroSky, Mindwave Mobile 2 headset) was acquired.

It also offers an API (Application Program Interface) to developers to retrieve raw data of sensed brainwave bands directly. It is likely that the commercial device can effectively reduce most of the noise. However, in this study, steps were taken to minimize the issue arising from any remaining noise. These noises are regarded as outliers. The outliers of the observed target may come from sensor noise, static human body, communication interference and other factors that cannot be grasped. Fig. 3(a) shows the collected signals without outlier detection and processing. It is obviously that there are some outliers are appeared in the collected signals. In the current work, the well-known statistical method, i.e., IQR (interquartile range) [52] is exploited. To replace the value of the outlier, a linear interpolation method is used. The research [56] made a comparison between IQR and Local Outlier Factor (LOF), and result revealed that both methods have identical performance in terms of accuracy, precision, recall, and f1-score.

The interquartile range (IQR) method has been proved that it is a simple and robust technique for detecting outliers with following advantages:

1. Simplicity: Grasping and employing the IQR method is uncomplicated.
2. Non-parametric: It applies to distributions, whether skewed or symmetric, rendering it adaptable to diverse dataset types.
3. Robustness: It is less influenced by extreme values or outliers compared to other statistical measures like the mean and standard deviation.
4. Flexibility: The outlier detection threshold can be customized to suit the specific analysis requirements.

But it also has some limitations as follows:

1. Insensitivity: Despite its robustness against extreme values, the IQR method may struggle to effectively identify mild outliers.
2. Threshold Dependence: The efficacy of the IQR method heavily relies on the choice of the threshold multiplier.
3. Disregarding Data Distribution: The IQR method disregards the underlying data distribution beyond quartile boundaries. Consequently, in datasets with intricate distributions, this approach might inaccurately flag outliers.
4. Information Loss: By focusing solely on the central 50% of the data, the IQR method risks overlooking valuable insights from extreme values.

In this method, the IQR is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of the entire sample. The multiplier for determining the maximum and minimum values is set at 1.5, yielding the following formulas:

$$\text{Maximum} = Q3 + 1{\cdot}5 * IQR$$

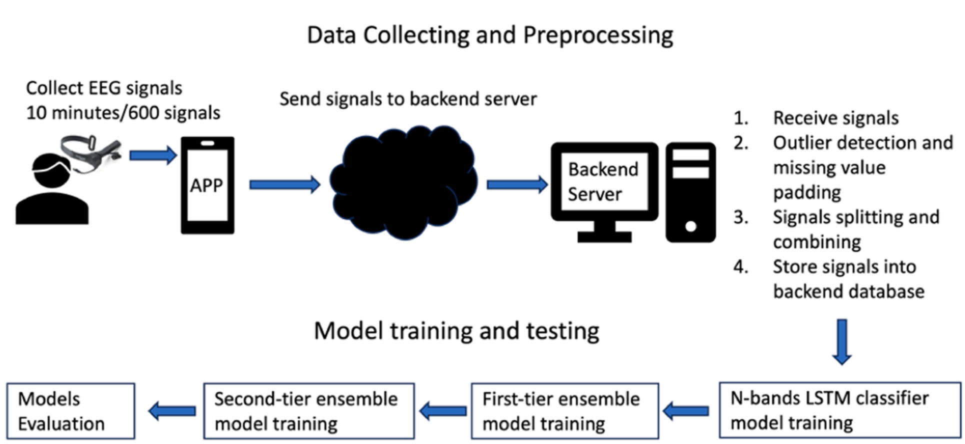



**Fig. 1.** System Flow.

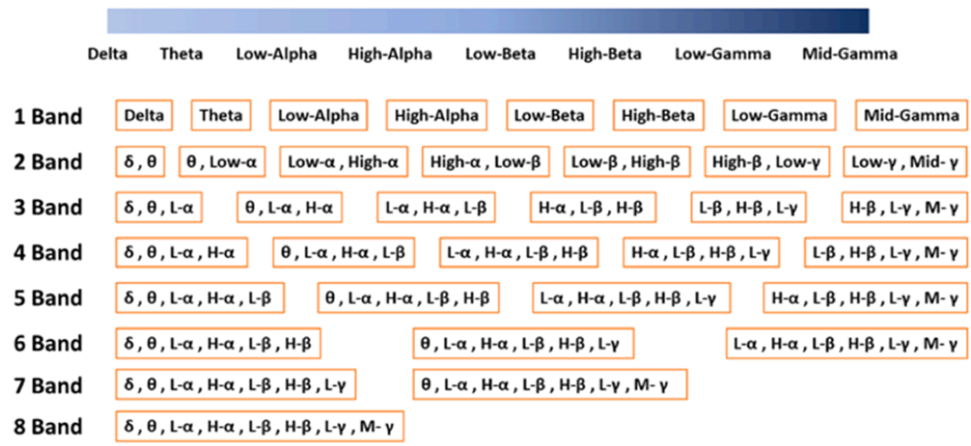



**Fig. 2.** EEG band splitting and combining.

Minimum = Q1 - 1·5 * IQR

This adjustment aims to mitigate the issue of Threshold Dependence. Additionally, our method samples data at a frequency of once per second, which helps minimize the impact of disregarding data distribution. Fig. 3(b) depicts the processed signals using the IQR method. The results demonstrate that signals filtered through IQR exhibit reduced extreme values and are less susceptible to various environmental noises, thereby enhancing signal integrity. In addition, given that the EEG signal is a time series data, errors in the field within the deep learning network might adversely impact the construction of prediction models and subsequent analysis results. Consequently, upon detecting outliers, we employ statistical techniques such as linear interpolation to substitute these outliers. This approach aims to mitigate the influence of outliers on the analysis.

In essence, the interquartile range method provides a straightforward and resilient means of identifying outliers, especially in datasets featuring skewed distributions or outliers. Nevertheless, it is not exempt from limitations and warrants careful application, taking into account the unique attributes of the data and the analytical objectives. We are optimistic that mitigating the impact of certain noise-related issues is achievable.

### 3.3. *Frequency band splitting and combining*

As discussed earlier, brain wave bands may correspond to distinct features of mental state. Although the EEG signal is divided into eight sub-bands, it is not possible to correlate the influence of every individual's EEG signal on emotion. So, it is assumed that the adjacent bands are still relevant. In this regard, as per the literature [21], the separated frequency bands are combined into a 1-band set, two adjacent bands are combined into a 2-bands set, …, 7 adjacent bands are combined into a 7-bands set and all 8 bands are combined into a set of 8-bands as shown in Fig. 2. For example, in a 2-bands set, there exists *((δ, θ), (θ, low-α), (low-α, high-α), (high- α, low-β), (low-β, high-β), (high-β, low-γ), (low-γ, mid-γ))*.

## 4. N-Band LSTM-based classifiers for depression assessment

This section is to present the proposed method for assessing depression. In Section 4.1 depicts the LSTM model, which is suitable for the time series data. Subsequently, we will demonstrate the procedure for constructing the N-bands LSTM classifiers. This classifier is the base of the multi-featured models.

### 4.1. *LSTM model for time series data*

LSTM is an appropriate and a suitable method for time series data classification or prediction tasks (Chang et al., 2020). Being a kind of RNN model, LSTM is widely applied in different fields, especially Deep Learning network architectures, to build the classification and prediction models. In order to ensure the entirety for the article, the network model of LSTM and the function of each gate are briefed in this section.

Fig. 4 shows the LSTM network and its internal gate. The LSTM cell has three gates, such as the forget gate, input gate, and the output gate. Among the three, the forget gate is the first gate, which is determined using the Eq. (1) to identify the information that needs to be discarded. Here, $h_{t-1}$ denotes the output at the previous time unit (t-1) and $X_t$ corresponds to the input of the current time unit.

$$f_t = \sigma\left(W_f \bullet \left[h_{t-1}, x_t\right] + b_f\right) \quad (1)$$

$$S(t) = \frac{1}{1 + e^{-t}} \quad (2)$$

Where $W_f$ and $b_f$ are two parameters in which $W_f$ denotes the weight matrix and $b_f$ is the bias vector These two parameters are learned at the time of training. And $\sigma$ is the *S(t)* that adapts the *Sigmoid* function in current work.

The second gate is the input gate, which is used to determine the information that should be remembered for every cell state. There are three operations in this gate as shown in the Eqs. (3)–(5). Here, $C_t$ is the output to the next cell.

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\widetilde{C}_t = \tanh\left(W_C \bullet \left[h_{t-1}, x_t\right] + b_C\right) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (5)$$

In these equations, $W_i$, $W_c$, $b_i$, and $b_c$ correspond to the same value with forget gate and are learnt at the time of training.

The third gate is the output gate. It is used to determine which information will be output in the cell state. In the gate, the $o_t$ and the $h_t$ functions are obtained using the Eqs. (6) and (7).

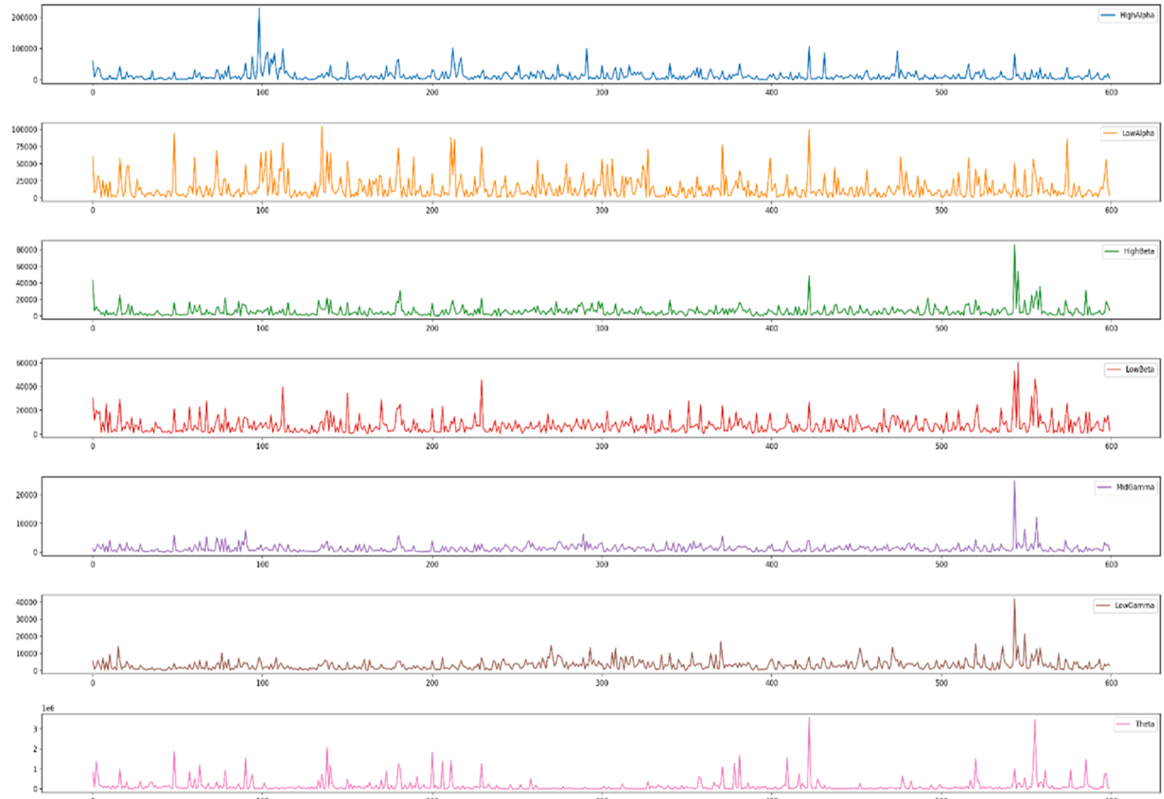
$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

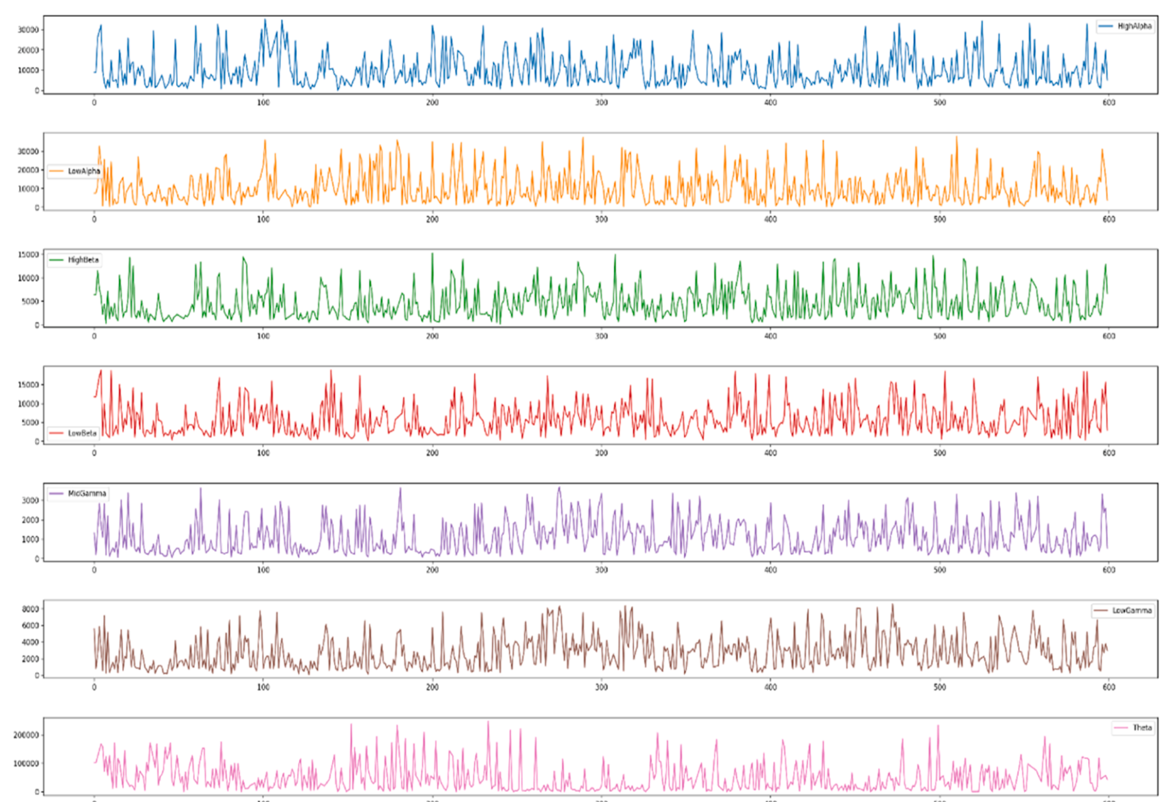
where $W_o$, and $b_o$ are also the same with forget gate and are learnt at the time of training.

### 4.2. *N-bands LSTM classifiers*

The LSTM network model is exploited to design the LSTM classifiers for each N-band set. As shown in Fig. 4, the classification model is




(a) The collected signals without outlier detection and processing




(b) The signals after outlier processing

**Fig. 3.** The EEG signals before and after outlier processing.

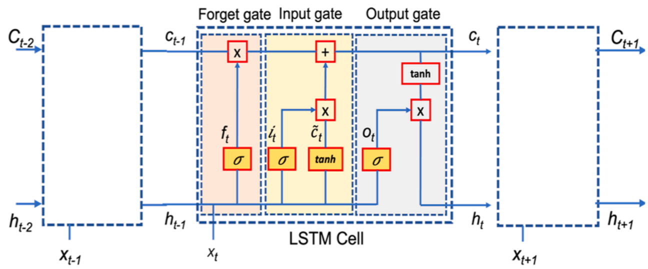



**Fig. 4.** The internal gates of LSTM.

trained with an LSTM for each N-band set in which N is obtained from 1 to 8. After preprocessing, the data set is divided into 70% for training, 20% for validation, and 10% for testing. The architecture of the 1-band LSTM classifier is shown in Fig. 5 whereas the 2-band LSTM classifier is shown in Fig. 6. The following section provides a simple description for the designed LSTM classifier. The input layer uses Dense with RELU activation function. The second layer of the LSTM uses *tanh* activation function, and the third layer is connected to the *Dropout* layer. The rate is set at 0.5 i.e., 50% of the data is randomly filtered. Then, 32, 16 and 8 nodes of the dense layer are used for the RELU excitation function.

The number of output nodes is the same as the number of classifications. *Softmax* function (normalized exponential function) is used to fix the probability value of the output range of the node, between 0 and 1, as shown in Eq. (8). When the input is X, the predicted category of the probability of j, is P whereas $w_j$ represents the probability of category, j. The sum of the probabilities of all the categories is calculated using *softmax* whereas the result is finally achieved and portrayed using the output layer.

$$P(X) = \frac{e^{w_j}}{\sum_{k=1}^{K} e^{w_k}} \tag{8}$$

Algorithm 1 shows the complete algorithm of N-bands LSTM classifier. In this algorithm, the signals of N-bands are regarded as input data, and the LSTM model of all N-bands will be output. Here N ranges from 1–8. When N = 1, one signal is input into the LSTM network to train the model. As depicted in Fig. 5, a total of eight 1-band LSTM models will be acquired at this time. When N = 2, two adjacent 2-bands signals will be input into the LSTM network to train the model. Similarly, as shown in Fig. 6, there are 7 LSTM models obtained.

The LSTM network starts with 10 nodes *dense* layer (Line 2), followed by the addition of an *LSTM* layer with 32 nodes (line 3). Subsequently, a *dropout* layer is incorporated, and then followed by 4 *dense* layers with 32, 16, 8, 2 nodes (Line 5–8), respectively. The practical implementation of this algorithm can be found in the Supplementary section. Apart from code explanations, it provides detailed descriptions of parameters. It's worth noting that variations in LSTM network designs may influence parameter configurations and subsequently affect the outcomes.

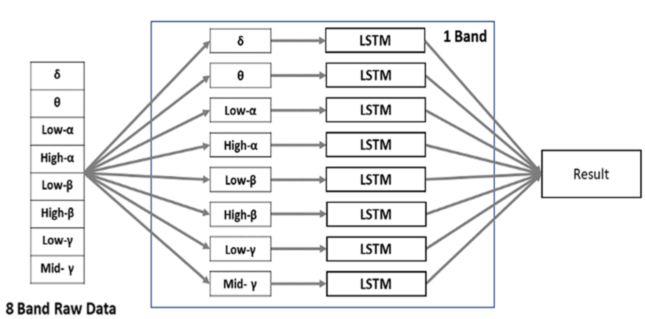



**Fig. 5.** The 1-band LSTM classifiers.

**Algorithm 1.** The procedure for N-bands LSTM Classifier.

```
Input:   N-bands dimension data
Output:  LSTM model
1:   model = Sequential()
2:   model.add(Dense ( 10 ,
           activation = 'relu',
           input_shape = (600, N)))
3:   model.add(LSTM(32, activation = 'tanh'))
4:   model.add(Dropout(0.5))
5:   model.add(Dense(32, activation = 'relu'))
6:   model.add(Dense(16, activation = 'relu'))
7:   model.add(Dense(8, activation = 'relu'))
8:   model.add(Dense(2, activation = 'softmax'))
9:   model.compile(optimizer = 'Adam',
         loss = 'categorical_crossentropy',
         metrics = ['accuracy'])
10:  Return model
```

## 5. Two-tiers ensemble learning for integrating multiple EEG bands

The ensemble learning approach, for integrating various EEG bands, is presented in this section. A two-tier ensemble learning technique is proposed to integrate all the classifiers of EEG bands.

### 5.1. Architecture design

A two-tier ensemble learning method is proposed to integrate all the EEG bands. The first-tier ensemble learning is to integrate all the LSTM classifiers of each N-band set. For example, Fig. 5 shows the 1-band classification model of first tier ensemble learning. Here, each LSTM classifier classifies and identifies the target initially, according to the input of adjacent two-band brain waves. Once the 2-band LSTM classifiers have produced the results, these outcomes are sent to the first-tier ensemble learning module. After all the N-band ensemble modules of the first tier have undergone the ensemble learning approach, the results from each module are sent to the ensemble module of the second-tier ensemble module in order to consolidate the whole results. The following section provides an overview of the architecture and the method of two-tier ensemble learning method.

#### 5.1.1. First tier ensemble learning

As shown in Fig. 6, a 2-band ensemble learning module is used to integrate seven 2-band LSTM classifiers. Similarly, other N-band ensemble learning modules are used in the first-tier ensemble learning to integrate eight N-band LSTM classifiers based on N adjacent bands. The ensemble learning module for N-band LSTM classifiers is called the N-band ensemble. Therefore, the first-tier ensemble learning module consists of eight N-band ensembles.

As mentioned in Section 2.2, various ensemble learning methods are in practice such as (i) *simple average method*, (ii) *ACC weight method, (iii) MAE weight method*, and *(iv) hard voting priority method.* Multiple linear regression methods are exploited here to calculate the intercept and weight of all the models. The results are obtained after the models are substituted in multiple linear regression.

It is evident that the 1-band ensemble integrates eight 1-band LSTM modules, the 2-band ensemble integrates seven 2-band LSTM modules, and the 8-band ensemble integrates one 8-band LSTM module. Therefore, a total of 36 LSTM classifiers are integrated in the first-tier ensemble.

#### 5.1.2. Second tier ensemble learning

After the first-tier ensemble, eight outputs are obtained from the N-band ensemble. Consequently, the second-tier ensemble should be designed to integrate all the outputs. The second-tier ensemble learning

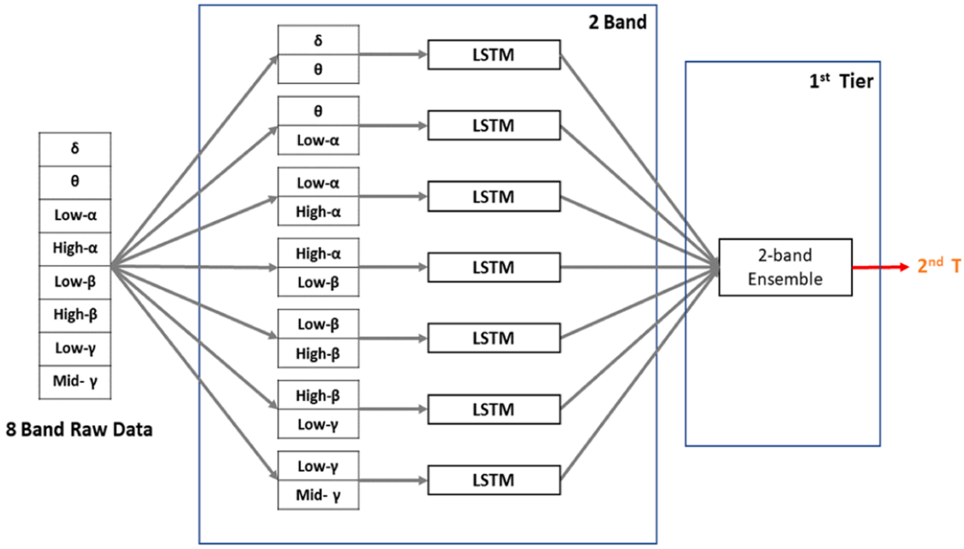



**Fig. 6.** 2-Band LSTM classification model with first tier ensemble learning.

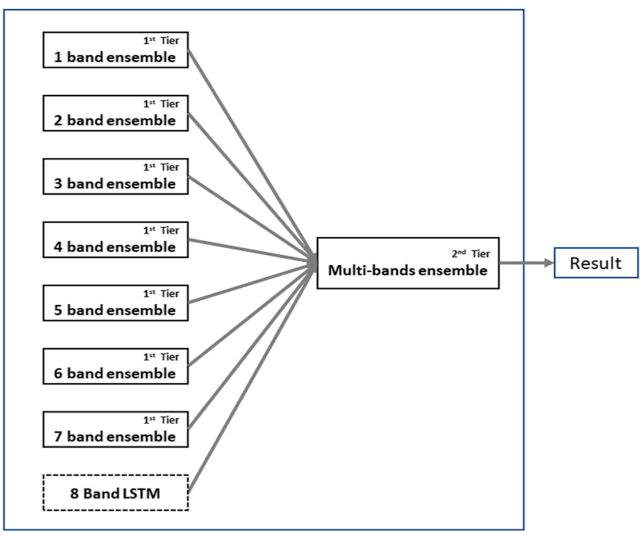



**Fig. 7.** The Second-Tier Ensemble Architecture.

module adopts five methods, discussed earlier, to integrate multiple N-band ensembles. Therefore, this module is designated as the multi-bands ensemble and is shown in Fig. 7.

### 5.2. Multiple linear regression method

In current research work, the Multiple Linear Regression (MLR) method is applied for ensemble learning. Assume that a set of n data is considered with an independent variable and a dependent variable. The formulae for multiple linear regression are denoted in Eqs. (9), (10) and (11) which are organized as Eq. (12). For example, there are seven classifiers, and each classifier predicts the values, such as $[x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}, x_5^{(1)}, x_6^{(1)}, x_7^{(1)}]$. The actual value is $[y_1]$, according to the formula $y_1 = \beta_0 + \beta_1 x_1^{(1)} + \beta_2 x_2^{(1)} + \beta_3 x_3^{(1)} + \beta_4 x_4^{(1)} + \beta_5 x_5^{(1)} + \beta_6 x_6^{(1)} + \beta_7 x_7^{(1)}$. Then calculate $[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7]$according to Y, $[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7]$. Here, $\beta_0$ represents the intercept of multiple regression while $[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7]$is $[x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}, x_5^{(1)}, x_6^{(1)}, x_7^{(1)}]$ of the *slope* denoting the weight. After calculating the above values, the predicted values calculated by seven classifiers are fed into the calculated Y.

$$(x_i, y_i), x_i \in R^{d+1}, \forall i = 1, \dots, n, x_i = \begin{bmatrix} 1 & x_i^{(1)} & \vdots & x_i^{(d)} \end{bmatrix} \tag{9}$$

$$Y_{n*1} = \beta_{(d+1)*1} X_{n*(d+1)} \tag{10}$$

$$Y = \begin{bmatrix} y_1 & y_2 \vdots y_n \end{bmatrix}_{n*1}, \beta = \begin{bmatrix} \beta_0 & \beta_1 \vdots \beta_d \end{bmatrix}_{(d+1)*1},$$

$$X = \begin{bmatrix} x_1^T & x_2^T \vdots x_n^T \end{bmatrix} = \begin{bmatrix} 1 x_1^{(1)} \cdots x_1^{(d)} \, 1 x_2^{(1)} \cdots x_2^{(d)} \vdots \vdots \ddots \vdots \, 1 x_n^{(1)} \cdots x_n^{(d)} \end{bmatrix}_{n*(d+1)} \tag{11}$$

$$y_i = \beta^T x_i = \begin{bmatrix} \beta_0 & \beta_1 \vdots \beta_d \end{bmatrix}^T \begin{bmatrix} 1 x_1^{(i)} \vdots x_d^{(i)} \end{bmatrix} = \beta_0 + \beta_1 x_1^{(1)} + \cdots + \beta_d x_d^{(i)} \tag{12}$$

Algorithm 2 shows the steps involved in MLR to calculate the target using the first- and the second-tier ensemble modules. In this algorithm, all first-tier output is treated as the input that are used to calculate all $\beta_d$ and the weights of $x_d^{(i)}$. Finally, the algorithm will return all $\beta_d$. The practical implementation of this algorithm also can be found in the Supplementary section. Apart from code explanations, it provides detailed descriptions of parameters.

In order to clearly understand the proposed architecture of the work, we merge two tiers of ensemble learning and show the overall architecture, as shown in Fig. 8. Initially, we measure and preprocess the 8 brainwave sub-bands using the EEG sensor. These sub-bands can be combined into N-sub-bands for further training using the LSTM model. After training individual band models (totally 36 LSTM models), we can merge N-bands LSTM models using the MLR method, namely 1st-tier ensemble. Lastly, we merge all 1st-tier MLR ensembles into 2nd-tier MLR ensembles.

**Algorithm 2.** The procedure for multiple linear regression method.

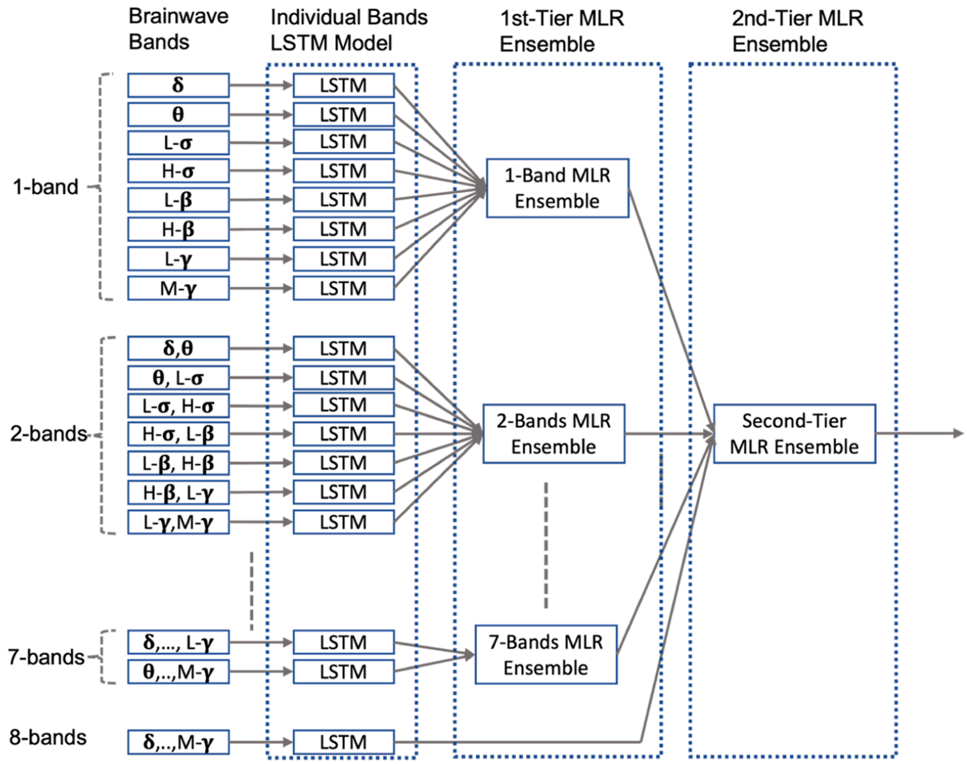



**Fig. 8.** Overall Architecture.

---

**Input:** $y_i$ is the actual value of the i-th data, and $x_d^{(i)}$ the predicted value given for the d-th classifier for the i-th item

**Output:** $\beta_0$ *is the intercept*, $\beta_d$ is weight of the *d-th classifier (Slope coefficient)*

$$y_i = \beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_d x_d^{(i)}$$

**for** i = n :

$y_i = dependent\ variable$ (Actual value)

$x_d^{(i)} = independent\ variable$ (Predictive value)

$\beta_0 = y\_intercept$ (intercept)

$\beta_d = slope\ coefficients\ for\ each\ explanatory\ variable$ (Slope coefficient, which can be used for each classifier weight)

**end for**

Compute $\beta_0$

Compute $\beta_d$

Return $\beta_0, \beta_d$

Note:
Through $\beta_0, \beta_d$ can be calculated by $x_d^{(i)}$ to calculate $y_i$

---

### *5.3. Performance metrics*

The current section discusses two methods, such as confusion matrix and ROC (AUC) which are used in the performance evaluation of the proposed work.

#### *5.3.1. Confusion matrix*

Confusion matrix is a common method used in the evaluation of a model's quality. In the confusion matrix, the predicted class is categorized as positive and the actual class is categorized as negative. In the current research work, positive represents depressed individuals and negative represents normal individuals. After classification, the following four situations are considered.

- *True Positive (TP)*: Predicts positive, while the subject is positive.
- *False Positive (FP)*: Predicts positive, while the subject is negative
- *False Negative (FN)*: Predicts negative, while the subject is positive.
- *True Negative (TN)*: Predicts negative, while the subject is negative.

To evaluate the validity of the proposed method, various measurements are used in data science analytical applications, such as *Precision*, *Recall*, *Sensitivity*, *Specificity*, *Accuracy* and *F1 score*. The definitions are as follows:

(i) *Positive Predictive Value* (also called Precision)

$$\mathrm{PPV}(Precision) = \frac{\sum TP}{\sum (TP + FP)} \tag{13}$$

(ii) *Negative Predictive Value* (NPV)

$$NPV = \frac{\sum TN}{\sum (FN + TN)} \tag{14}$$

(iii) *Sensitivity* (also called Recall)

**Table 2**
Subjects Data.

| Parameters | Depressed subjects | Control group |
|---|---|---|
| Average age | 38.96 ± 14.02 | 24.66 ± 0.88 |
| Age range | 19–70 | 23–28 |
| Average age (male) | 40.73 ± 14.38 | 25 ± 0.76 |
| Average age (female) | 37.31 ± 13.26 | 24 ± 0.5 |
| Gender (male/female) | (18/31) | (14/4) |
| Data collection duration | 10 min | 10 min |
| No. of EEG records | 142 | 72 |

$$Recall = \frac{\sum TP}{\sum (TP + TN)} \tag{15}$$

(iv) *Specificity*

$$Specificity = \frac{\sum TN}{\sum FP + TN} \tag{16}$$

(v) *Accuracy*

$$Accuracy = \frac{\sum (TP + TN)}{\sum (TP + FP + FN + TN)} \tag{17}$$

(vi) *F1 Score*: This value is calculated combining *Precision* and *Recall*.

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{18}$$

#### 5.3.2. Receiver Operator Characteristic (ROC) curve and Area Under the Curve (AUC)

ROC curve and AUC [51] are comprehensive indicators in performance evaluation procedure. ROC is drawn as a curve for the comparison of true positive ratios (TPR) among all the samples. The true positive and the ratio of false positives (FPR), among all the samples, are actually negative. FPR can be determined based on the inclusion of both true value and the predicted value of all the positive and negative samples. These values are calculated as a coordinate point (FPR, TPR) and are mapped on the graph. Theoretically, the AUC of a model follows four cases: (i) AUC = 1, a perfect classifier or a perfect prediction result. (ii) 0.5 < AUC < 1, the outcome is better than random guessing and has a good prediction value. (iii) AUC = 0.5 denotes that the model has no predictive value and mimics random guessing. (iv) and AUC < 0.5 which is worse than random guessing.

## 6. Experiments, results, and performance evaluation

This study provides a comprehensive description of the experimental environment technology that was employed. The brain wave sensor (Mindwave Mobile 2 headset) uses a brain-computer interface (Brain-computer Interface, NeuroSky).

### 6.1. Data collection

In current study, 8 sub-bands of the EEG brainwaves were adopted to train the LSTM models. However, these models are irrelevant and there are no useful EEG signals available. EEG signals were collected from both individuals diagnose with depressed and individuals without any mental health conditions. For this research work, all the EEG signals were gathered under IRBs permit from Taipei Medical University Hospital and En Chu Kong Hospital. Table 2 contains the relevant details of the dataset. The authors gathered 142 EEG recordings from depressed people and 72 EEG recordings from healthy participants, which represent different frequency bands of the EEG signal. For each participant, 600 s of eight sub-band EEG signals were collected.

We have implemented an APP and a web to collect the data from the NeuroSky device, as shown in Fig. 9. Fig. 9(a) shows the APP interface and it receives the 8 sub-band of EEG data from the NeuroSky device, such as *hAlpha, lAlpha, hBeta, lBeta, mGamma, lGamma, Delta and Theta*. Fig. 9(b) shows the login interface[1] of the web site. After entering the system, user can choose a de-identified user name, as shown in the (1) bottom of Fig. 9(c), and a specified daytime record, as shown in the (2) bottom of Fig. 9(c). And then by dragging down the web page, users will be able to see the original raw data of specified de-identified user and daytime, as shown in Fig. 9(d). In Fig. 9(d), block (1) represents the results identified by the model, while block (2) represents the results of clinical diagnosis by doctors. We are sorry that other information may not be provided due to research ethics regulations.

This study was conducted by the IRBs of two hospitals, and patients were asked about their willingness to participate in the admissions during outpatient visits. Each admission lasted about 10 min. Since the patients were outpatients, every participating patient could move freely. Therefore, the patients involved in this study exhibited either moderate or mild depression. The procedure of the data collection as described follows. Depressed subjects for collecting data were assigned by psychiatrists. All subjects are still in treatment when the psychiatrist obtains the patient's consent to collect data during the patient's return visit. The severity of symptoms may vary, but they indeed all be in treatment. In addition, the data collection procedure was discussed with the physicians. These steps are described as follows: First of all, the patient should be exposed to soft music for approximately 5 min to induce emotional relaxation prior to data collection to ensure the patient's state is in a resting state. Second, with the assistance of trained personnel, the patient wears a commercial EEG sensor (NeuroSky, Mindwave Mobile 2 headset) and then connects with our implemented APP. Upon reaching the 10-minute mark, data collection is initiated with the interval being discussed with a psychiatrist. Each subject can collect 600 EEG signal records. Table 3 shows a sample of the investigated dataset.

In addition, in a clinical setting, in order to reduce the pressure during subject data collection, and following the consent of the psychiatrist, a one-channel EEG sensor the NeuroSky, Mindwave Mobile 2 headset EEG sensor was employed, positioned at FP1 position. The rationale behind this is that the sensor remains visible to the patient during data collection, thereby reducing the likelihood of complications.

### 6.2. Experimental results and discussion

The current study authors conducted some experiments to evaluate the proposed method (*Multiple Linear Regression method*) and compared the output from the proposed method with that of four other ensemble learning approaches such as *simple average method*, *Accuracy* (*ACC*) *weight method*, *MAE weight method*, and (HV) *hard voting priority method*. At first, a comparison was conducted among different ensemble methods, in terms of accuracy of N-band LSTM and first-tier ensemble results. Then, the values of different indicators were compared among different ensemble learning approaches. Finally, ROC and AUC values were exploited to evaluate the ensemble learning method. The results achieved from these comparisons confirmed the superiority of the proposed method than other ensemble methods.

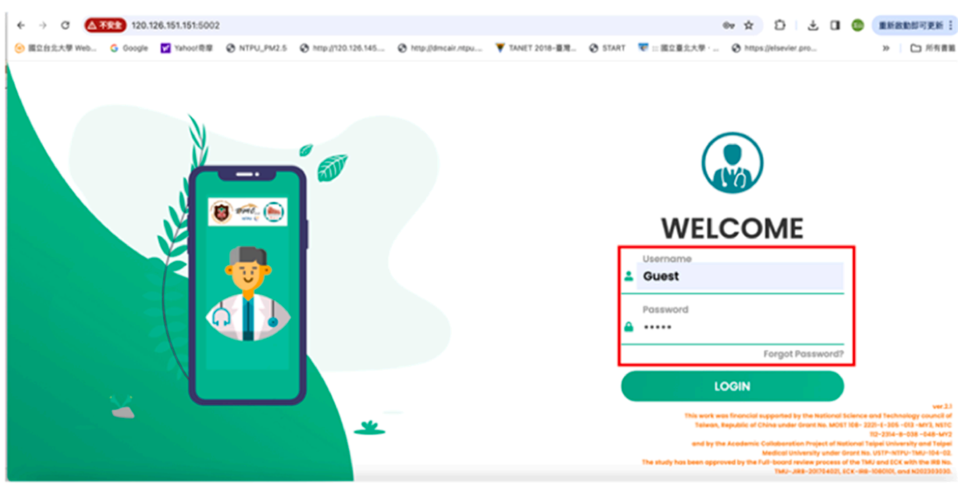
#### 6.2.1. Accuracy comparison of various first tier ensemble

This subsection firstly employs a *Simple Average (SA) method*, in the first tier, to ensemble all the N (where N is 1–8) band LSTM classifiers. In 1 band, each part is a band whereas it contains 8 parts such as delta, theta, …, mid-gamma. In 2 bands, each part is a combination of two bands (such as (delta, theta), (theta, low-alpha), …, (low-gamma, mid-gamma) with a total of 7 parts.

In the case of 8 bands, there is no need for an ensemble module, since all the bands are combined into a wave and fed as input to the LSTM classifier. Fig. 10(a) reveals that the SA method achieved the least output for all N band ensembles. In this result, the accuracy of 3-band

[1] http://120.126.151.151:5002/


(a) APP interface.




(b) The Login page of the work for accessing the collected data

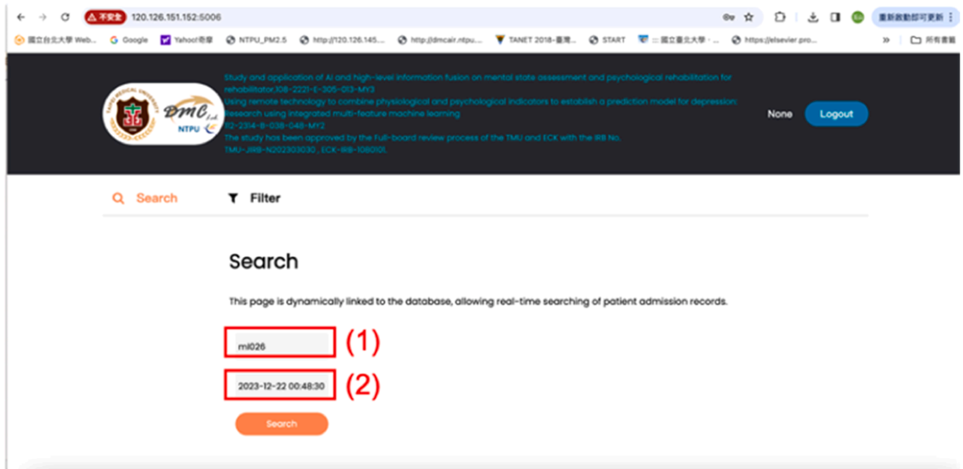
**Fig. 9.** Data collection interface of APP and Web.

classifiers was better than others. So, the result of the ensemble was also better. However, with a value of only 0.76, it remains insufficient.

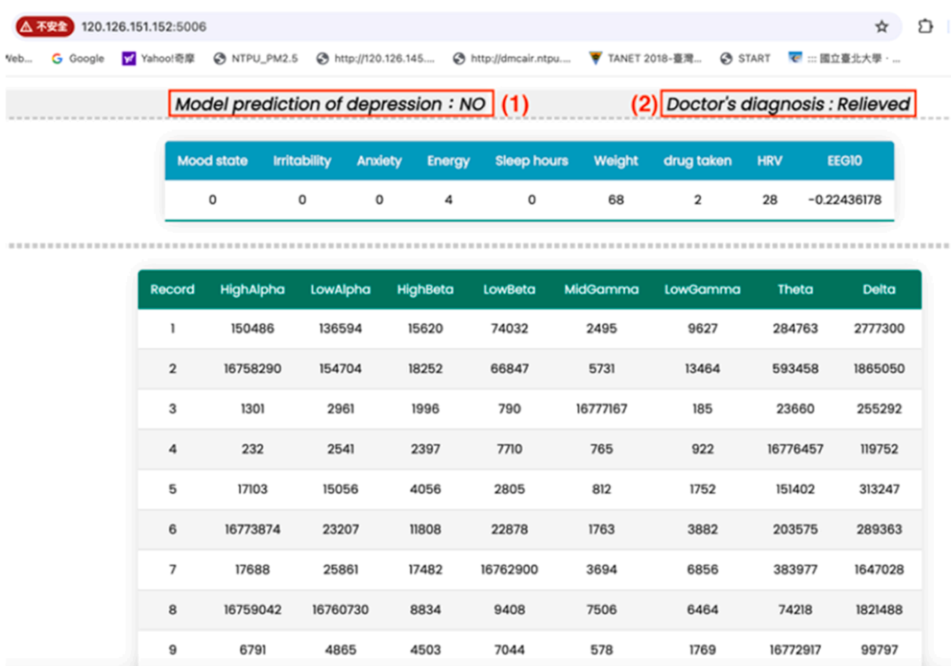
The purpose of the second ensemble approach is to exploit the ACC weight approach and the results are shown in Fig. 10(b). The figure infers that the result of the ACC weight method was slightly better than the SA method. This is attributed to the fact that SA uses only a simple average of weights, whereas ACC weight method achieved a high accuracy and provided a high weight. Though both are weighting methods, ACC weight was better. Further, the results also infer that the improvement of ensembles, in 1-band and 2-band, was better than the ensemble of 6-band and 7-band. Since the number of classifier outputs was high, more ACC weights got added to the ensemble module, which produced better results.

In this stage, the MAE weighting approach applies to the first-tier ensemble and the results are shown in Fig. 10(c). In comparison with the first two methods, the MAE weighting method performed better. Especially with the regard to the 6-band, the accuracy of the ensemble was higher than the rest of the parts. The maximum result, i.e., 0.77 among all the ensembles, was achieved in 3-band, whereas the rest of the results are still not good enough. The next step is to adopt the Hard Voting (HV) priority technique to the first-tier ensemble learning and the results are shown in Fig. 10(d). From the results, it can be inferred that the performance obtained in this approach was worse than the other three techniques. In this case, the best ensemble result was still in 3-band which stood at 0.74.

Finally, the authors evaluated the performance of the proposed approach, i.e., Multiple Learning Regression (MLR) approach and the results are shown in Fig. 10(e). The results infer that multiple linear regression can enhance the accuracy of most LSTM classification models when first-tier ensemble learning is employed. Obviously, the




(c) Please select one de-identified user name and specify the date-time of a record to search the 8-subbands data.

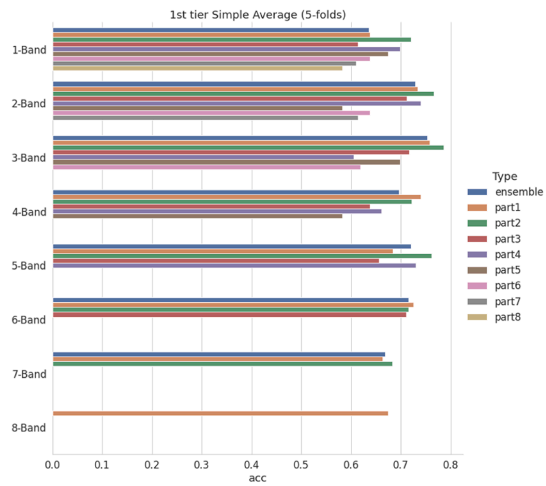



(d) 8-subbands raw data of specified user and daytime record
Fig. 9 Data collection interface of APP and Web

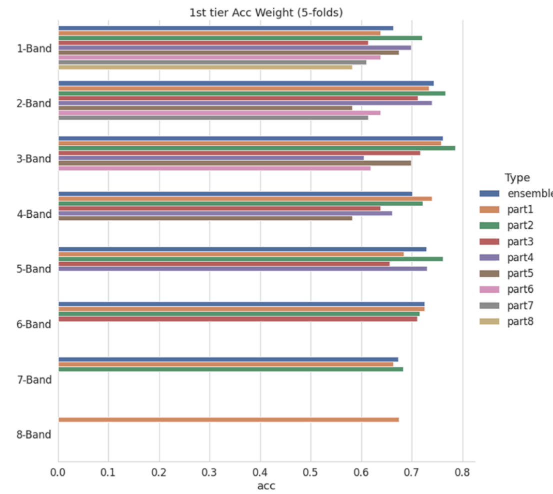
**Fig. 9.** (*continued*).

**Table 3**
A sample of the investigated dataset.

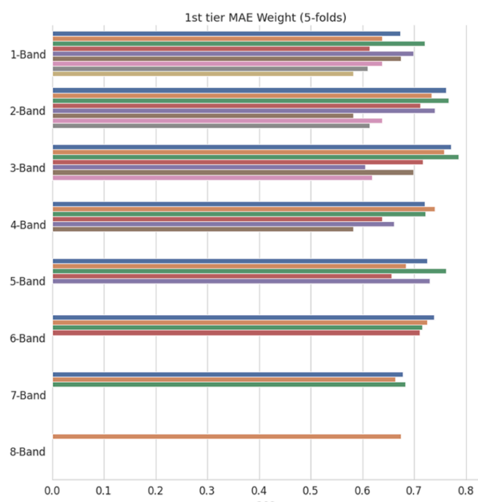
| Times | t1 | t2 | t3 | t4 | ... | t600 |
|---|---|---|---|---|---|---|
| **δ** | 919413 | 289987 | 75023 | 431546 | ... | 139594 |
| **θ** | 263542 | 29389 | 29258 | 114880 | ... | 24385 |
| **Low-α** | 120888 | 7633 | 17241 | 6424 | ... | 56920 |
| **High-α** | 10723 | 7094 | 26414 | 135055 | ... | 19139 |
| **Low-β** | 96408 | 2688 | 15101 | 9706 | ... | 16458 |
| **High-β** | 43795 | 2784 | 9546 | 9921 | ... | 15597 |
| **Low-γ** | 7077 | 6779 | 3357 | 5188 | ... | 9732 |
| **Mid- γ** | 7077 | 9664 | 6779 | 3357 | ... | 7584 |



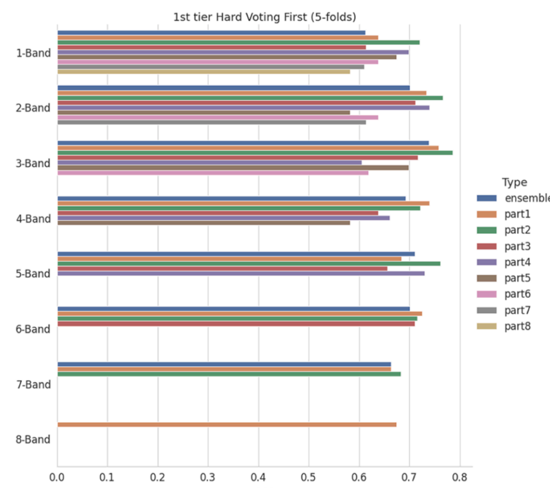

(a) Simple Average method.



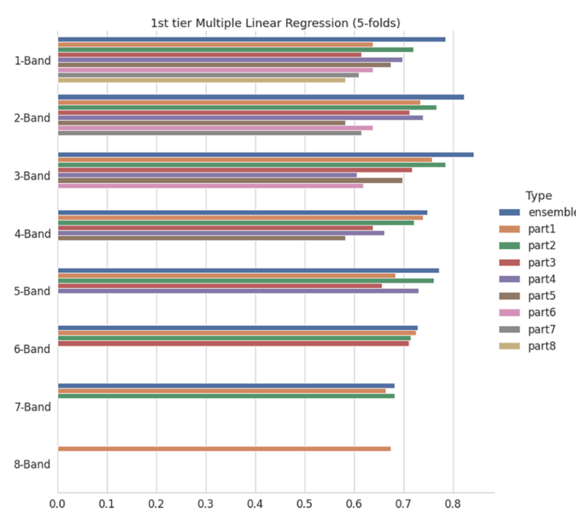

(b) Acc weight method.




(c) MAE weight method.




(d) HV method.




(e) MLR method

**Fig. 10.** First tier Accuracy Comparison of various ensemble methodsIn addition to using the above methods, traditional machine learning also has other ensemble learning methods, such as Adaboost and Random Forest (RF), which are used to combine some weak learners to form a strong learner. Here we also conducted experiments to evaluate the performance and make a comparison. In the experiments, first we exploit LSTM as the baseline models for the combination of various numbers of bands, and use the Adaboost to combine the LSTM models. In addition, we also utilize RF (ensemble multiple Decision Trees) to conduct the experiment and make a comparison. As shown in Fig. 12 the performance of Adaboost or RF is also superior to other methods in the combination of 1 to 8 bands. We also utilize the second-tier ensemble to integrate the results of the first-tier and compare the accuracy, as shown in Fig. 11. The results reveal that the results of Adaboost and RF are almost as close as MLR, and the accuracy of RF is 0.873, while Adaboost is 0.866.

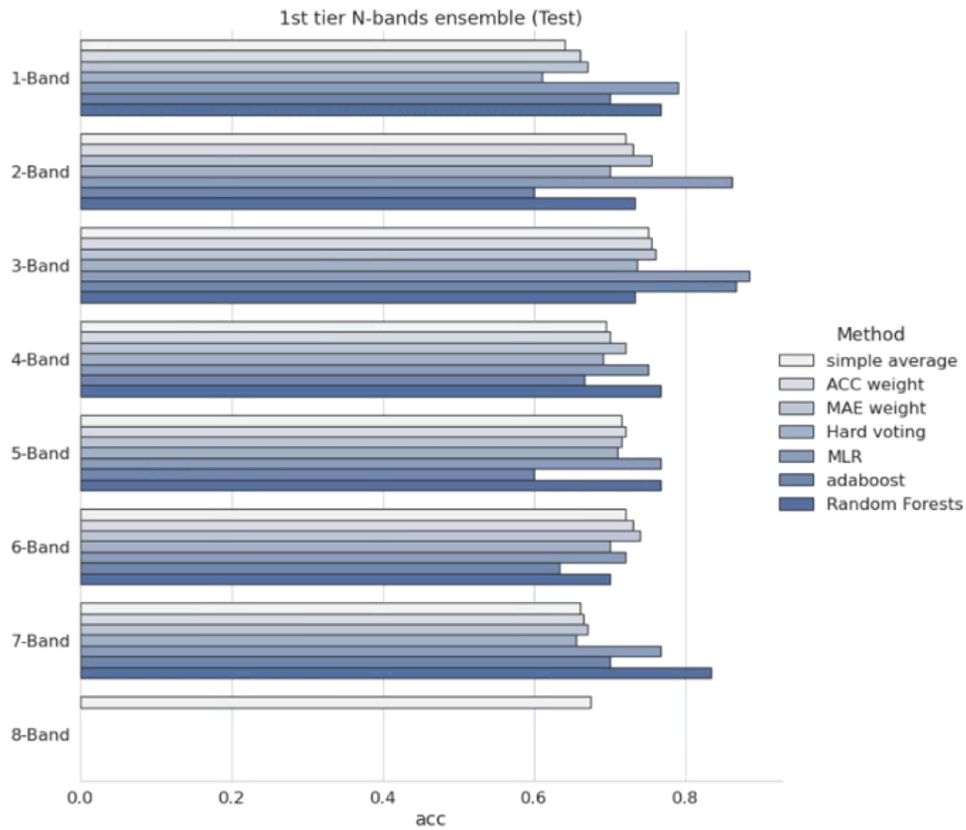



**Fig. 11.** Five methods of first-tier N-band ensemble learning.

performance outcomes of 1-band (about 0.78), 2-band (about 0.82) and 3-band (about 0.84) were significantly better than the input LSTM classification model. These outcomes demonstrate the superiority of the proposed multiple linear regression model. This may be attributed to the fact that MLR autonomously acquired the weight of each variable (part) based on the variable outcomes. Then, its offset is adjusted, as shown in Eq. (12).

A summary of the performance of five ensemble methods, in all N-bands, is shown in Fig. 11. From this result, it can be inferred that the results of MLR are indeed better than the other four methods in the first-tier ensemble method, except with 6-band. From 1-band to 5-band, the MLR exhibited significant improvement in terms of accuracy. However, no better improvement was found in 6-band and 7-band. On the other hand, only the MLR method achieved an accuracy of more than 0.8 in 2-band and 3-band. After MLR, MAE weighting was found to be the next best method.

From the results, two major findings are discussed herewith. First, with the exception of the MLR ensemble modules using 2-band and 3-band modules, the highest level of accuracy was attained in part 1 and part 2, regardless of whether first-tier ensemble learning occurred or not. Most of the modules achieved an accuracy in the range of 0.7 and 0.8. However, 2-band and 3-band of MLR exceeded 0.8. Second, the implementation of ensemble learning of MLR resulted in a substantial increase in accuracy, particularly in the 1-band, 2-band, and 3-band segments.

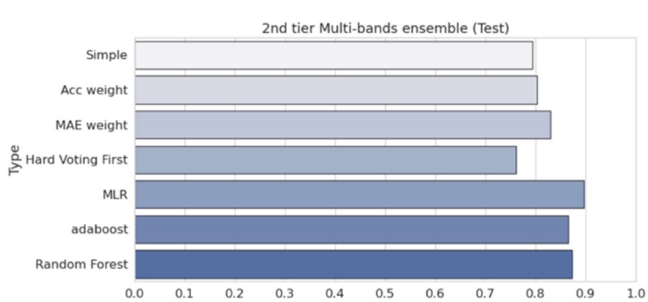



**Fig. 12.** Comparison of learning accuracy of the second-tier multi-bands ensemble.

Fig. 5 shows that the ensemble module of the first-tier of each band generated an ensemble result. Then, another ensemble module was exploited to integrate all the eight results together. These five methods can also be used in a second-tier ensemble learning method to combine the results of the eight first-tier ensemble modules. The above five methods were exploited to design the second-tier ensemble module and were validated through some experiments to evaluate the accuracy of these methods. When the output of the N-band ensemble flowed through the multi-band ensemble, the final forecast of the entire model was determined for which the results are shown in Fig. 12. From this result, it can be inferred that after using the above five methods like the second-tier ensemble module, the final result of the two-tier ensemble learning module, in terms of accuracy, is in the order of MLR > MAE > ACC > SA > HV.

### 6.2.2. *Various indicators comparison of second tier ensemble learning*

Multiple types of confusion matrix-related indicators are used in this study as an evaluation index. These indicators only present the final results of the second-tier ensemble, instead of intermediate results achieved from the first-tier ensemble. First, the confusion matrix was obtained after the experiments. The results obtained from the SA approach (TP=126, FP=28, FN=16, TN=44) were slightly lesser compared to that of the ACC weight method (TP=125, FP=25, FN=17, TN=47). It is also slightly lesser than that of the MAE weight method (TP=128; FP=21, FN=14, TN=51). On the contrary, the Hard Voting (HV) priority approach achieved the worst result (TP=126, FP=35,

**Table 4**
Multi-tiers five ensemble learning methods evaluation indicators.

| Various evaluation indicators | | | | | | |
|---|---|---|---|---|---|---|
| Index | Positive Predictive Value (Precision) | Negative Predictive Value | Sensitivity (Recall) | Specificity | Accuracy | F1 Score |
| **SA**[32] | 0.818 | 0.733 | 0.887 | 0.611 | 0.794 | 0.851 |
| **ACC (modified from**[55]**)** | 0.833 | 0.734 | 0.88 | 0.653 | 0.803 | 0.856 |
| **MAE [35]** | 0.859 | 0.785 | 0.901 | 0.708 | 0.837 | 0.88 |
| **HV**[34] | 0.783 | 0.698 | 0.887 | 0.514 | 0.761 | 0.832 |
| **LSTM + Adaboost** | 0.863 | 0.875 | 0.95 | 0.7 | 0.866 | 0.904 |
| **Random Forest** | 0.886 | 0.86 | 0.866 | 0.881 | 0.873 | 0.876 |
| **MLR** | **0.935** | **0.829** | **0.908** | **0.875** | **0.897** | **0.921** |

Note: Simple Average (SA)、Accuracy Weight (ACC)、MAE Weight (MAE)、Hard Voting priority (HV)、Multiple Linear Regression (MLR)

FN=16, TN=37). Finally, the Multiple Linear Regression (MLR) method (TP=129, FP=9, FN=13, TN=63) was found to be the best model since it significantly reduced the FP in comparison with other techniques.

Based on the confusion matrix, some well-known indicators can be calculated for which the results are shown in Table 4. Multiple Linear Regression approaches achieved the highest accuracy (0.897), followed by other methods such as MAE-weighted method (0.836), ACC-weighted method (0.803) and SA method (0.794). The HV priority method achieved an outcome of 0.761. Other indicators also confirmed that the MLR method is superior to MAE, ACC, SA, and HV, in terms of F1-score being 0.921, precision being 0.935, recall being 0.908 and the specificity being 0.875. From these results, it can be inferred that among various indicators of effectiveness, MLR obtained the maximum performance. In addition, we evaluated and compared MLR, Adaboost, and RF. Based on the findings presented in Table 4, MLR demonstrates favorable levels of Accuracy, Precision, and F1-Score. In contrast, the NPV and Sensitively of LSTM + Adaboost are relatively high, while the Specificity of RF is relatively high.

### 6.2.3. Summary of the evaluation

Next, we further analyze the relationship between this metrics and actual results shown in Table 4. According to the definition of Precision (Positive Predictive Value) in Eq. (13), the higher the Precision, the higher the true positive, which means the higher the accuracy of the prediction of depressed subjects. In this study, our precision reaches 0.935, representing a remarkably high accuracy in assessing depressed subjects. On the contrary, based on the definition of NPV (Negative Predictive Value) in Eq. (14), it can be found that the NPV is relatively low, as shown in Table 4. This implies a decrease in the accuracy of true negatives, indicating that our method is less accurate when assessing the health of subjects. From these two indicators, it means that the proposed method is suitable to evaluate depressed subjects.

In addition, according to the Accuracy definition of Eq. (17), this represents the overall assessment accuracy of whether it is depressed subjects or healthy subjects. The higher the Accuracy, the more accurate the depression assessment is. Among all methods in Table 4, our accuracy is nearly 0.9, which means that our evaluation accuracy is the highest among the methods, and is also higher than the two baseline methods of LSTM+Adaboost and Random Forest.

The definition of Sensitivity (Recall) in Eq. (15) is the number of correctly identified positive results divided by the number of all samples that should be identified as positive. Our result in recall is higher than 0.9. It means that the method can correctly identify the depressed subjects. Additionally, it is important to analyze the F1-Score. As defined by Eq. (18), F1-Score is the correct prediction percentage in classification applications. In this study, our MLR method reached the highest correct classification percentage of 0.921, which is higher than all methods. It revealed that our method can obtain the highest in the categories of depressed subjects and healthy subjects. It is better than other ensemble methods and two baseline methodologies (LSTM+Adaboost and Random Forest). It reveals that the proposed method can classify depressed subjects and healthy subjects more accurately.

### 6.2.4. ROC and AUC comparison

AUC (Area under the ROC (Receiver Operating characteristic Curve) Curve) is another useful and effective performance evaluation indicator [51]. In the current study, the authors evaluated the AUC values for the last results of five ensemble learning methods after performing two-tier ensemble learning. The results are shown in Fig. 13. The results show the AUC values of different ensemble learning methods are in the order of MLR (0.8917) > MAE (0.8049) > ACC (0.7665) > SA (0.7492) > HV (0.7006). The indicator also shows that MLR ensemble learning achieved the highest value, whereas the worst output was achieved by HV. This result is consistent with other previous performance indicators. The evaluation outcomes achieved above are summarized and presented in simple terms. Regardless of the results and irrespective of the processing techniques such as no ensemble learning, one-tier (only first-tier), or multi-tiers (two-tier), the proposed MLR ensemble learning method achieved superior outcomes in terms of all the indicators. Further, the results are consistent and can be trusted.

## 6.3. Discussion

### 6.3.1. The advantage and disadvantage using one-channel EEG

In this section, we will explain the advantages and disadvantages of using one-channel EEG to assess depression. The advantages are:

1). Easy to use. The utilization of multiple channel EEG sensors can be inconvenient due to the complexities involved in wearing and operating the devices.

2). One-channel equipment is relatively inexpensive, the user's price burden will be lighter.

3). In the collection of EEG signals, users can easily operate and collect signals at home without the help of professionals. After the model is trained, the signals can be retrieved remotely and then inferred.

4). Reduce the patient's psychological pressure when collecting data and reduce the user's fear of operations.

5). The sensor nodes of EEG must be aligned with the position to be sensed to reduce variability during data collection. One-channel sensors are relatively simple in sensor settings and will not cause variability during data collection.

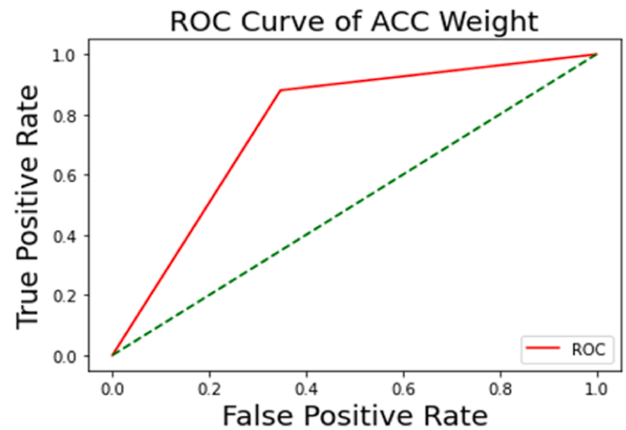
In addition, the disadvantages of using a single channel are as follows:

1). The data collected is less and may be less accurate than multi-channel research.

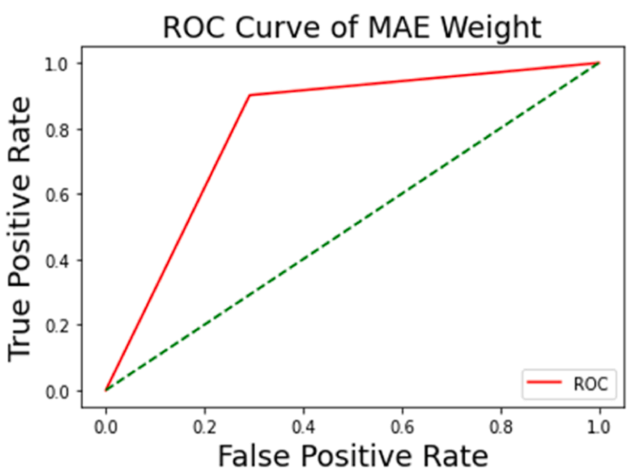
2). A more complex recognition model is required to achieve higher accuracy.

The main purpose of this study is to use advanced models that can increase the convenience of use, reduce prices, allow future users to operate by themselves, and reduce operational stress to achieve acceptable identification accuracy. This paper mainly proposes multi-featured EEG bands ensemble deep learning models. From the results shown in the above section, we can see that the method we proposed can increase the accuracy of assessment. It is believed that this method can also improve the accuracy of the overall system if used on multi-channels EEG sensors.

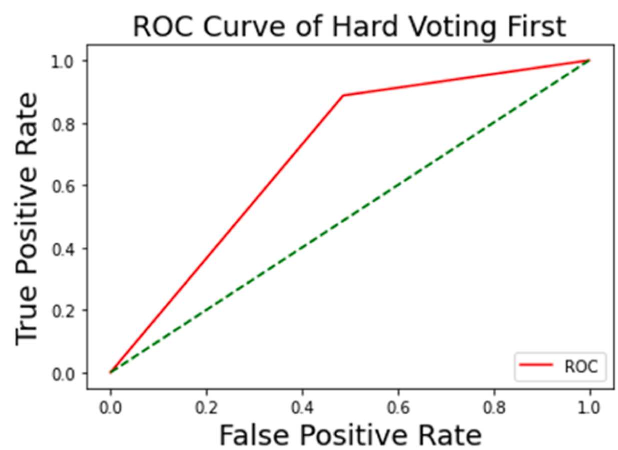

(a) The AUC ≈ 0.7492 for SA



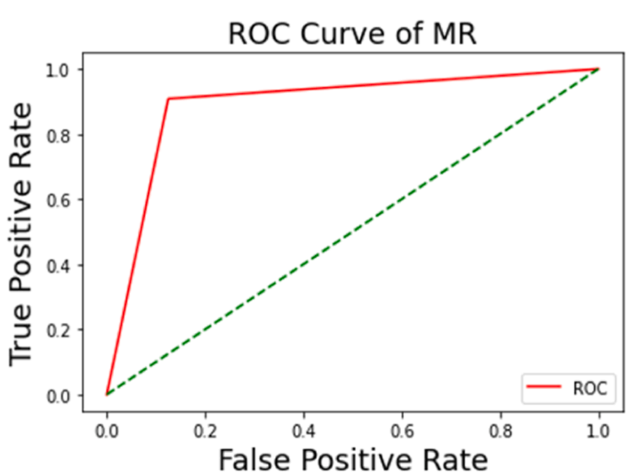

(b)The AUC ≈ 0.7665 for ACC




(c) The AUC ≈ 0.8049 for MAE




(d) The AUC ≈ 0.7006 for HV




(e) The AUC ≈ 0.8917 for MLR

**Fig. 13.** Ensemble learning method ROC curve and AUC.

*6.3.2. Extension to N-channels input*

In this current study, we solely utilize a single-channel EEG device to collect brainwave signals for training the model. Despite its advantages outlined in the previous section, this approach also imposes limitations, particularly in its restriction to single-channel data. Should future applications necessitate multi-channel EEG utilization, adjustments to the system architecture may be required.

This section elaborates on extending the current architecture to accommodate multi-channel data. Given that our architecture employs an LSTM model to predict each sub-band, we draw inspiration from a

**Table 5**
Summary and comparison of existing work relevant with deep learning network for depression assessment.

| | No. of Subjects | No. of Electrodes. | Baseline Model | Accuracy |
|---|---|---|---|---|
| Bachmann et al.,[44] | 34(17 normal, 17 depressed subject) | 19(18 +1) | spectral asymmetry index, and nonlinear, detrended fluctuation analysis | 91.2% |
| Rafiei et al., [50] | 64(30 normal, 34 depressed) | 19 | Convolution-based Inception | 91.67% |
| Wan et al., [54] | 35 (12 normal, 23 depressed) | 6 | CNN | 79.08% |
| Uyulan et al.,[47] | 92(46 normal, 46 depressed) | 19 | CNN * 3 | 90.22% |
| Proposed work | 67(18 normal, 49 depressed) | 1 | LSTM | |

study by [22], where an integrated LSTM model successfully predicted air quality. In this model, a merge layer combined three LSTM models into one, yielding favorable results. We propose extending this merge layer to integrate N LSTM models. Additionally, leveraging the fusion module proposed in [57], each sub-band signal can be inputted into the fusion module, where the inputs from N signals are fused into a single output. This output is then fed into the LSTM model within our architecture. Regardless of the above two methods, our system architecture can be easily applied to multi-channel EEG devices.

### 6.3.3. Comparison with other works

The comparison between existing work relevant with deep learning networks for depression assessment is shown in Table 5. It is widely recognized that the size of the training dataset has a great influence on the Deep Neural Network model. The larger the size of the training dataset, the better the prediction accuracy has. In addition, we use the basic network architecture of LSTM to recognize the pattern and to do the assessment. This is due to our belief that each signal is easily recognized as a 600-second time series data. The LSTM model is proven to outperform other kinds of DNN in the time series data.

Although the deep learning models and data sets used differ from existing studies which use multi-channels EEG to obtain more data, the work only uses one-channel EEG to collect and assess. With limited data sets, we retrieve the sub-bands of various features of EEG signals, combine the adjacent signals of features, and then integrate the results through ensemble learning. In the case of limited numbers of data using one-channel EEG, it can achieve performance similar to that of multi-channels. We believe that if we adopt multi-channels to do the work, it will have better results than the current performance. However, as listed in Section 6.2.1, one-channel EEG has some advantages. With more data sets in the future, this method will have good results.

To elaborate further on the aforementioned findings:

In reference [44], by employing 14 features and integrating SASI (spectral asymmetry index) and DFA (detrended fluctuation analysis) methods, an accuracy of 91.2% is achieved. However, when used individually, both methods yield accuracies below 80%.

In reference [47], three different methods were utilized, namely the CNN models ResNet-50, MobileNet, and Inception-v3, along with data from four EEG main frequency bands collected from 19 electrodes. It was observed that the delta frequency band outperforms others, achieving a predictive accuracy of 90.22% and an AUC value of 0.9 for the ResNet-50 architecture.

In reference [50], employing an InceptionTime-based CNN model with 19-channel raw EEG signals, a channel-selection strategy achieves an accuracy of 91.67%. However, when reducing the number of channels, the accuracy decreases to 87.5%.

In reference [54], a convolutional neural network named HybridEEGNet was employed. Six surface electrodes (Fp1, Fp2, F3, F4, P3, and P4) were placed on the scalp, and two parallel lines were used to learn synchronous and regional EEG features. The results demonstrate that HybridEEGNet achieves an accuracy of 79.08% in three-category classification.

Based on the findings from Table 5 and the preceding analysis, we can summarize as follows:

**Classification technology:** Most studies, except for [44], which employs a combination of SASI and DFA, utilize DNN-based models due to their proven success. DNNs eliminate the need for signal feature extraction, although they require a substantial amount of data for model training. This study also employs DNN technology.

**Network model:** Among the aforementioned studies employing DNNs, most utilize the CNN network model. They convert EEG frequency bands into images and then employ CNNs to recognize the spectrogram. Unlike [47], which examines the performance of four different frequencies, other studies do not analyze the impact of depression across various frequencies. In contrast, this study directly inputs raw data from different frequency bands with distinct mood characteristics into LSTM for model training. This approach bypasses the need to convert raw data into a spectrogram for image recognition.

**Number of electrodes:** With the exception of [54], most systems use 19 electrodes to capture brainwave signals. The results suggest that using more electrodes leads to improved accuracy. Since this study employs only one electrode for signal collection and model training, its performance is comparable to these methods. However, we believe that utilizing a more sophisticated EEG device with a greater number of electrodes would significantly enhance accuracy.

## 7. Conclusion and future work

In the current study, a multi-tier ensemble learning model was proposed to integrate various N-band LSTM classification models. These models are used in the classification of combined data of N-band EEG signals. The main purpose of the study is to evaluate how accurate the depression assessment is. In this work, at first, the signal collected by the EEG sensor is split into eight sub-bands. Then, the adjacent N sub-bands are combined into a set of N-band EEG signals. The LSTM network is exploited to train the model so as to evaluate the depression. Then, a two-tier ensemble learning module integrates all the LSTM models. When implementing the ensemble module, the Multiple Linear Regression method is utilized.

The research adopts the simplest EEG sensor, commercialized one channel EEG sensor, and the position of the sensor is placed at FP1 position to collect data with 57 subjects (49 depressed and 18 healthy subjects). The results are as follows, accuracy being 0.897, F1-score being 0.921, precision being 0.935, negative predictive value being 0.829, recall being 0.908, specificity being 0.875, and AUC being 0.8917, respectively. The performance of the proposed method was compared with four different ensemble learning methods, such as Simple Average (SA) method, Accuracy (ACC) weight method, MAE weight and Hard Voting (HV) priority method. These results revealed that irrespective of whether first tier or complete two-tier ensemble modules, the classification accuracy, achieved by MLR in case of depression assessment, was excellent compared to other four methods. In addition, we also evaluated and compared MLR, Adaboost, and RF. From Table 4, we can find that the Accuracy, Precision, and F1-Score of MLR are relatively high, while the NPV and Sensitively of LSTM + Adaboost are relatively high, and the Specificity of RF is relatively high. Therefore, it can be inferred that the proposed method is effective and can improve the accuracy of depression assessment.

Although the current study results showed good accuracy, the model did not classify the patients of different ages and different types of mental disorders. This is an important phenomenon to accurately classify and enhance the efficiency of the assessment process. In subsequent studies, it is suggested to incorporate a broader sample population, targeting different ages and mental disorders. Thus, accurate classification can be achieved which in turn helps the psychiatrists to make knowledge-based decisions for precise treatment. Besides, if similar features are extracted from each band and then exploits the proposed multi-tier ensemble learning method, potentially better results can be obtained. In the future, we aim to extract features from each band. Alternatively, we could utilize the attention mechanism to evaluate specific features. In addition, in order to make this system architecture suitable for multi-channel EEG devices, we also consider using integrated method or fusion module to integrate multi-channel signals into our system.

## Ethical approval

The study has been approved by the Full-board review process of the TMU and ECK with the IRB No. TMU-JIRB-201704021 and ECK-IRB-1080101.


## Funding

This work was partially supported by the Ministry of Science and Technology of Taiwan, Republic of China under Grant No. MOST 108-2221-E-305−013 -MY3, NSTC 112–2314-B-038-048 -MY2, NSTC 112-2425-H-305 -002 -, NSTC 111-2221- E-305 -011 -MY3; and by the Academic Collaboration Project of National Taipei University and Taipei Medical University under Grant USTP-NTPU-TMU-104–02, and by the National Taipei University under Grant 113-NTPU_ORD-F-004.


## CRediT authorship contribution statement

**Yue-Shan Chang:** Conceptualization, Formal analysis, Funding acquisition, Methodology. **Kuo-Hsuan Chung:** Conceptualization, Data curation, Funding acquisition, Validation. **Wei-Ting Yen:** Formal analysis, Methodology, Software, Visualization. **Linen Lin:** Data curation, Resources. **Satheesh Abimannan:** Visualization, Writing – original draft.

## Declaration of Competing Interest

No potential conflict of interest relevant to this article was reported.

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.


## Acknowledgement

We would like to express our deepest appreciation to anonymous reviewers for their hard review and providing valuable comments.


*Consent to participate*

The data collection was approved by the hospital's IRB, the risk of impact to participants is extremely low, and the public interest is likely to be high.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.03.022.

## References


[1] Ettman CK, Cohen GH, Abdalla SM, Sampson L. Persistent depressive symptoms during COVID-19: a national, population-representative, longitudinal study of U.S. adults. Lancet Reg Health 2022;5(100091):1–12. https://doi.org/10.1016/j.lana.2021.100091.

[2] Lyons, G.M., Sharma, P., Baker, M., O'Malley, S., Shanahan, A. (2003). A computer game-based EMG biofeedback system for muscle rehabilitation. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)* , pp.17–21. DOI: 10.1109/IEMBS.2003.1279682.

[3] De Jonckheere J, Ibarissene I, Flocteil M, Logier R. A smartphone based cardiac coherence biofeedback system. *In Proc.* 36th Annu Int Conf IEEE Eng Med Biol Soc 2014:26–30.

[4] Vuillerme N, Chenu O, Pinsault N, Moreau-Gaudry A, Fleury A, Demongeot J, Payan Y. Pressure sensor-based tongue-placed electrotactile biofeedback for balance improvement - Biomedical application to prevent pressure sores formation and falls. *In Proc.* 29th Annu Int Conf IEEE Eng Med Biol Soc 2007:22–6. https://doi.org/10.1109/IEMBS.2007.4353744.

[5] Tai C-H, Chung K-H, Teng Y-W, Shu F-M, Chang Y-S. Inference of mood state indices by using a multimodal high-level information fusion technique. IEEE Access 2021;9:61256–68. https://doi.org/10.1109/ACCESS.2021.3073733.

[6] Li D, Tang J, Deng Y, Yang L. Classification of resting state EEG data in patients with depression. IEEE Int Conf E-Health Netw, Appl Serv (HEALTHCOM) 2020:1–2. https://doi.org/10.1109/HEALTHCOM49281.2021.9398920.

[7] Shen J, Zhao S, Yao Y, Wang Y, Feng L. A novel depression detection method based on pervasive EEG and EEG splitting criterion. IEEE Int Conf Bioinforma Biomed (BIBM) 2017:1879–86. https://doi.org/10.1109/BIBM.2017.8217946.

[8] Hajian M, Hassan Moradi M. Quantification of Depression Disorder Using EEG Signal. *24th Natl 2nd Int Iran Conf Biomed Eng (ICBME)* 2017:1–5. https://doi.org/10.1109/ICBME.2017.8430237.

[9] Al-Kaf H, Khandoker A, Khalaf K, Jelinek HF. NeuroSky Mindwave Mobile Headset 2 as an Intervention for Reduction of Stress and Anxiety Measured with Pulse Rate Variability. Comput Cardiol 2020;2020:1–4. https://doi.org/10.22489/CinC.2020.350.

[10] Dinesh Anton Raja P, Akash D, John Prem Kumar S, Harsha DS, Arunachalapperumal C. Feature extraction and classification of EEG signal-based anomaly detection and home automation for physically challenged/impaired people using neurosky mindwave headset. AIP Conf Proc 2020;2207(1):040006. https://doi.org/10.1063/5.0000064.

[11] Liao C-Y, Chen R-C, Tai S-K. Emotion stress detection using EEG signal and deep learning technologies. IEEE Int Conf Appl Syst Invent (ICASI) 2018:90–3. https://doi.org/10.1109/ICASI.2018.8394414.

[12] Morshad S, Mazumder R, Ahmed F. Analysis of Brain Wave Data Using Neurosky Mindwave Mobile II. *Proc Int Conf Comput Advancements* 2020;28:1–4. https://doi.org/10.1145/3377049.3377053.

[13] Shrestha A, Mahmood A. Review of Deep Learning Algorithms and Architectures. IEEE Access 2019;7:53040–65. https://doi.org/10.1109/ACCESS.2019.2912200.

[14] Liao C-Y, Tai S-K, Chen R-C, Hendry H. Using EEG and Deep Learning to Predict Motion Sickness Under Wearing a Virtual Reality Device. IEEE Access 2020;8:126784–96. https://doi.org/10.1109/ACCESS.2020.3008165.

[15] Gao Y, Lee HJ, Mehmood RM. Deep learning of EEG signals for emotion recognition. IEEE Int Conf Multimed Expo Workshops (ICMEW) 2015:1–5. https://doi.org/10.1109/ICMEW.2015.7169796.

[16] Islam MR, et al. Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques. IEEE Access 2021;9:94601–24. https://doi.org/10.1109/ACCESS.2021.3091487.

[17] Yean W, et al. An Emotion Assessment of Stroke Patients by Using Bispectrum Features of EEG Signals. Brain Sci 2020;10(10):672. https://doi.org/10.3390/brainsci10100672.

[18] Li Y, Cao D, Wei L, Tang Y, Wang J. Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing. Clin Neurophysiol 2015;126(11):2078–89. https://doi.org/10.1016/j.clinph.2014.12.026.

[19] Opitz D, Maclin R. Popular ensemble methods: An empirical study. J Artif Intell Res 1999;11:169–98. https://doi.org/10.1613/jair.614.

[20] Pagowski M, et al. Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts. Atmos Environ 2006;40(18):3240–50. https://doi.org/10.1016/j.atmosenv.2006.02.006.

[21] Lin C-Y, Chang Y-S, Abimannan S. Ensemble multifeatured deep learning models for air quality forecasting. Atmos Pollut Res 2021;12(5):101045. https://doi.org/10.1016/j.apr.2021.03.008.

[22] Chang Y-S, Chiao H-T, Abimannan S, Huang Y-P, Tsai Y-T, Lin K-M. An LSTM-based aggregated model for air pollution forecasting. Atmos Pollut Res 2020;11(8):1451–63. https://doi.org/10.1016/j.apr.2020.05.015.

[23] Shen F, Peng Y, Kong W, Dai G. Multi-Scale Frequency Bands Ensemble Learning for EEG-Based Emotion Recognition. Sensors 2021;21(4):1262. https://doi.org/10.3390/s21041262.

[24] Ganaie, M.A., Hu, Minghui, Malik, A.K., Tanveer, M., Suganthan, P.N. (2022). Ensemble deep learning: A review. 〈https://arxiv.org/abs/2104.02395〉.

[25] Kirmizi-Alsan E, Bayraktaroglu Z, Gurvit H, Keskin YH, Emre M, Demiralp T. Comparative analysis of event-related potentials during Go/NoGo and CPT: Decomposition of electrophysiological markers of response inhibition and sustained attention. Brain Res 2006;1104(1):114–28. https://doi.org/10.1016/j.brainres.2006.03.010.

[26] Michael AJ, Krishnaswamy S, Mohamed J. An open label study of the use of EEG biofeedback using beta training to reduce anxiety for patients with cardiac events. Neuropsychiatr Dis Treat 2005;1(4):357–63. PMID: 18568116.

[27] Fingelkurts AA, Fingelkurts AA, Rytsälä H, Suominen K, Isometsä E, Kähkönen S. Impaired functional connectivity at EEG alpha and theta frequency bands in major

depression. Hum Brain Mapp 2007;28(3):247–61. https://doi.org/10.1002/hbm.20275.
[28] Moldofsky H, Lue FA. The relationship of alpha and delta EEG frequencies to pain and mood in 'fibrositis' patients treated with chlorpromazine and l-tryptophan. Electroencephalogr Clin Neurophysiol 1980;50(1-2):71–80. https://doi.org/10.1016/0013-4694(80)90324-7.
[29] Oathes DJ, Ray WJ, Yamasaki AS, Borkovec TD, Castonguay LG, Newman MG, Nitschke J. Worry, generalized anxiety disorder, and emotion: Evidence from the EEG gamma band. Biol Psychol 2008;79(2):165–70. https://doi.org/10.1016/j.biopsycho.2008.04.005.
[30] Arikan MK, Metin B, Tarhan N. EEG gamma synchronization is associated with response to paroxetine treatment. J Affect Disord 2018;235:114–6. https://doi.org/10.1016/j.jad.2018.04.041.
[31] Zhou Z-H. Ensemble learning. Mach Learn 2021:181–210.
[32] Saqlain M, Jargalsaikhan B, Lee JY. A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. IEEE Trans Semicond Manuf 2019;32(2):171–82.
[33] Amin-Naji M, Aghagolzadeh A, Ezoji M. CNNs hard voting for multi-focus image fusion. J Ambient Intell Hum Comput 2020;11:1749–69. https://doi.org/10.1007/s12652-019-01199-0.
[34] Dagnew G, Shekar BH. Ensemble learning based classification of microarray cancer data on tree-based features. Cogn Comput Syst 2021;3:48–60. https://doi.org/10.1049/ccs2.12003.
[35] Acar E, Rais-Rohani M. Ensemble of metamodels with optimized weight factors. Struct Multidiscip Optim 2009;37(3):279–94. https://doi.org/10.1007/s00158-008-0230-y.
[36] Park Y, Lee J, Park W. Self-Weight Ensemble Method Adjust Influ Individ Models Based Reliab 2021. ArXiv, abs/2104.04120.
[37] Sun S, Zhang C, Zhang D. An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognit Lett 2007;28(15):2157–63. https://doi.org/10.1016/j.patrec.2007.06.018.
[38] Avots E, Jermakovs K, Bachmann M, Päeske L, Ozcinar C, Anbarjafari G. Ensemble Approach for Detection of Depression Using EEG Features. Entropy 2022;24(2):211. https://doi.org/10.3390/e24020211.
[39] Zhang X, Shen J, Din ZU, Liu J, Wang G, Hu B. Multimodal depression detection: fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble. IEEE J Biomed Health Inform 2019;23(6):2265–75. https://doi.org/10.1109/JBHI.2019.2938247.
[40] Wang Z, Gu T, Zhu Y, Li D, Yang H, Du W. FLDNet: Frame-Level Distilling Neural Network for EEG Emotion Recognition. IEEE J Biomed Health Inform 2021;25(7):2533–44. https://doi.org/10.1109/JBHI.2021.3049119.
[41] Ye C, Yin Z, Zhao M, Tian Y, Sun Z. Identification of mental fatigue levels in a language understanding task based on multi-domain EEG features and an ensemble convolutional neural network. Biomed Signal Process Control 2022;72(Part-B):103360. https://doi.org/10.1016/j.bspc.2021.103360.
[42] Iyer A, Das SS, Teotia R, et al. CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings. Multimed Tools Appl 2022. https://doi.org/10.1007/s11042-022-12310-7.
[43] Rajendra Acharya U, Oh SL, Hagiwara Y, Tan JH, Adeli H, Subha DP. Automated EEG-based screening of depression using deep convolutional neural network. Comput Methods Prog Biomed 2018;Volume 161(Pages):103–13. https://doi.org/10.1016/j.cmpb.2018.04.012. DOI: 10.1109/TSM.2019.2904306.
[44] Bachmann M, Lass J, Hinrikus H. Single channel EEG analysis for detection of depression. Biomed Signal Process Control 2017;Volume 31(2017, Pages):391–7. https://doi.org/10.1016/j.bspc.2016.09.010.
[45] Mumtaz W, Qayyum A. A deep learning framework for automatic diagnosis of unipolar depression. Int J Med Inform 2019;Volume 132:103983. https://doi.org/10.1016/j.ijmedinf.2019.103983.
[46] Sharma G, Parashar A, Joshi AM. DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. Biomed Signal Process Control 2021;Volume 66:102393. https://doi.org/10.1016/j.bspc.2020.102393.
[47] Uyulan C, et al. Major depressive disorder classification based on different convolutional neural network models: deep learning approach. Clin EEG Neurosci 2021;Vol. 52(1):38–51. https://doi.org/10.1177/1550059420916634.
[48] Deng X, Fan X, Lv X, Sun K. SparNet: A Convolutional Neural Network for EEG Space-Frequency Feature Learning and Depression Discrimination. Front Neuroinform 2022;16:914823. https://doi.org/10.3389/fninf.2022.914823.
[49] Seal A, Bajpai R, Agnihotri J, Yazidi A, Herrera-Viedma E, Krejcar O. DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG. 2021, Art no. 2505413 IEEE Trans Instrum Meas 2021;vol. 70:1–13. https://doi.org/10.1109/TIM.2021.3053999.
[50] Rafiei A, Zahedifar R, Sitaula C, Marzbanrad F. Automated Detection of Major Depressive Disorder with EEG Signals: A Time Series Classification Using Deep Learning. IEEE Access 2022;vol. 10:73804–17. https://doi.org/10.1109/ACCESS.2022.3190502.
[51] Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 2005;17(3):299–310. https://doi.org/10.1109/TKDE.2005.50.
[52] Shabbir J, Ahmed A. Estimation of interquartile range in stratified sampling under non-linear cost function. Commun Stat - Simul Comput 2022;51(4):1891–8. https://doi.org/10.1080/03610918.2019.1689267. DOI: 10.1109/EMBC.2014.6944695.
[53] Pei Z, Wang H, Bezerianos A, Li J. EEG-based multiclass workload identification using feature fusion and selection. IEEE Trans Instrum Meas 2021;70:1–8. https://doi.org/10.1109/TIM.2020.3019849.
[54] Wan Z, Huang J, Zhang H, Zhou H, Yang J, Zhong N. HybridEEGNet: A Convolutional Neural Network for EEG Feature Learning and Depression Discrimination. IEEE Access 2020;vol. 8:30332–42. https://doi.org/10.1109/ACCESS.2020.2971656.
[55] Park, YeongHyeon, JoonSung Lee, and Wonseok Park. Self-Weighted Ensemble Method to Adjust the Influence of Individual Models based on Reliability. arXiv preprint arXiv:2104.04120 (2021).
[56] Fredianto F, Putri DAP. Comparison of the interquartile range algorithm and local outlier factor on Australian weather data sets. AIP Conf Proc 2023;2727:040010. https://doi.org/10.1063/5.0141897.
[57] Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. Multi-channel fusion LSTM for medical event prediction using EHRs. J Biomed Inform 2022;Volume 127:104011. https://doi.org/10.1016/j.jbi.2022.104011.