ISSN 2355-3286

EEG-Based Depression Detection in the Prefrontal Cortex Lobe using mRMR Feature Selection and Bidirectional LSTM

Monica Pratiwi
Computer Engineering, Universitas Multimedia Nusantara, Tangerang, Indonesia
monica.pratiwi@umn.ac.id

Accepted on 04 December 2023
Approved on 27 December 2023

***Abstract*— Depression can induce significant anguish and impair one's ability to perform effectively in professional, academic, and familial settings. This condition has the potential to result in suicide. Annually, the number of deaths resulting from suicide exceeds 700,000. Among individuals aged 15-29, suicide has emerged as the fourth most prevalent cause of mortality. Challenges in treating depression include limited accessibility to mental health care in rural regions and misdiagnosis resulting from subjective evaluations, wherein insufficient expertise can contribute to inaccurate diagnoses. Electroencephalography (EEG) has gained popularity as a tool for the identification and study of a number of mental illnesses in the past years. Therefore, an automated technique is required to precisely distinguish between normal EEG signals and depression signals. This research focused on developing an EEG-based depression detection system in the prefrontal cortex lobe area (Fp1, Fpz, and Fp2). One of the developments carried out in this research is the implementation of Bidirectional Long Short-Term Memory (Bi-LSTM) as the model classification and minimum redundancy maximum relevance (mRMR) feature selection. The results suggest that the combination of mRMR feature selection with 25 features and the bidirectional LSTM obtained 92.83% accuracy.**

***Index Terms*—Bidirectional LSTM; Detection Depression; EEG Signals; mRMR Feature Selection.**

I. INTRODUCTION

The World Health Organization (WHO) recognized mental health and psychosocial wellbeing as essential components of health in 1978, and this definition has been discussed extensively as a UN resolution. In 2015, every nation has adopted the UN Sustainable Development Goal (SDG) point 3, which is to "ensure healthy lives and promote well-being for all at all ages." Substance misuse and mental health are directly addressed. Target 3.4 calls for countries to "promote mental health and well-being" through prevention and treatment in order to cut premature death by one-third[1].

However, since 2020, global anxiety and depression has become 25% more prevalent, with teenagers and women being the most affected. Of the 5,470 respondents, 40.9% reported having mental or behavioral health issues. TSRD (trauma- and stressor-related disorder) symptoms associated with COVID-19 (26.3%), increased drug usage as a result of COVID-19 (13.3%), and seriously considering suicide in the last 30 days (10.7%) are among the categories that include symptoms of anxiety or depressive illness (30.9%) [2]. Among these, symptoms of anxiety or depressive illness are the most prevalent behavioral health issues. Depression differs from mild mood fluctuations and short-term emotional responses against challenges in everyday life. Repeating instances of depression at a moderate or great scale may lead to the disorder becoming a serious health condition. Depression can cause great suffering and the affected may be unable to function well at work, at school, and in the family. The disorder can lead to suicide. More than 700,000 people die from suicide annually. Suicide has become the fourth leading cause of death for people ages 15-29.

The obstacles of effective mental health treatment include the lack of resources and the social stigma against mental disorders. It is a common occurrence that people from countries of all income levels to have undiagnosed depression or to be misdiagnosed with depression and prescribed antidepressants. Doctors and psychologists are able to diagnose depressive disorders through counselling sessions and ask relevant questions to the subject, despite being vulnerable to mistakes due to the examiner’s lack of experience.

For this reason, Electroencephalography (EEG) has gained popularity as a tool for the identification and study of a number of mental illnesses in the past several years, including autism, ADHD, Alzheimer's, dementia, alcoholism, and motor imagery. EEG captures electrical activity in the brain and shows how brain signals are used. In comparison to healthy persons, depression sufferers' synapses produce less neurotransmitters and have a reduced concentration of receptors due to cell malfunction. Compared to healthy individuals, this results in extremely low levels of brain activity in depressed persons. For psychiatrists, the processes of visual interpretation and complicated, nonlinear, nonstationary EEG signal analysis are challenging, time-consuming, and inefficient[3].

In other words, an automated technique is required to precisely distinguish between normal EEG signals and depression signals as validation that helps psychiatrists or psychologists in diagnosing depression. Consequently, a number of researchers have put forth a computer-based detection system based on EEG data that use a classification technique to distinguish or identify whether the patient is in the normal or depressive category.

Wan et. al [4] proposed a machine learning technique to distinguish between Major Depressive Disorder (MDD) and normal control subjects. The EEG dataset is acquired using the Fp1 and Fp2 electrodes of a 32-channel EEG device. The findings indicate that the classification accuracy using EEG data from the Fp1 site is superior to that using EEG data from the Fp2 location. Moreover, the results suggest that analyzing single-channel EEG data can effectively differentiate Major Depressive Disorder (MDD) at a level comparable to analyzing multi-channel EEG data. In addition, a portable electroencephalogram (EEG) equipment is utilized to gather the signal specifically from the Fp1 region, resulting in the acquisition of the second dataset. The genetic algorithm (GA) integrated Classification and Regression Tree obtains an impressive accuracy of 86.67% by leave-one-participant-out cross validation. This result demonstrates the potential of the single-channel EEG-based machine learning technology in supporting the prescreening application for Major Depressive Disorder (MDD).

In a research conducted by Cai et. al [5] , a psychophysiological database was created, consisting of 213 people (92 depressive patients and 121 normal controls). The EEG signals of all subjects were recorded utilizing a prefrontal-lobe three-electrode EEG system at Fp1, Fp2, and Fpz electrode sites. The signals were collected during both resting state and sound stimulation. A total of 270 linear and nonlinear features were retrieved after using denoising techniques utilizing the Finite Impulse Response filter, which combined the Kalman derivation method, Discrete Wavelet Transformation, and an Adaptive Predictor Filter. Subsequently, the feature selection strategy known as minimal-redundancy-maximal-relevance was employed to decrease the number of dimensions in the feature space. The depressed individuals were differentiated from the normal controls using four classification methods: Support Vector Machine, K-Nearest Neighbor, Classification Trees, and Artificial Neural Network. The performance of the classifiers was assessed using 10-fold cross-validation. The findings indicated that the K-Nearest Neighbor (KNN) algorithm achieved the highest level of accuracy, reaching 79.27%.

Similar to the two previous studies mentioned, this research also focuses on developing an EEG-based depression detection system in the prefrontal cortex lobe area (Fp1, Fpz, and Fp2) to construct a system that is more user-friendly. There are several studies that have proven the correlation between depressive disorders and activity in the prefrontal cortex lobe[6]. This research aims to increase the accuracy value of the Major Depressive Disorder classification model with normal subjects. One of the developments carried out in this research is the implementation of Bidirectional Long Short-Term Memory (Bi-LSTM) for an EEG signal-based depressive disorder classification system. Apart from applying Bi-LSTM, time segmentation of the data was also carried out in the feature extraction process.

II. Research Methods

The research is constructed into several steps, namely dataset acquisition, signal pre-processing, feature extraction, feature selection, and classification.

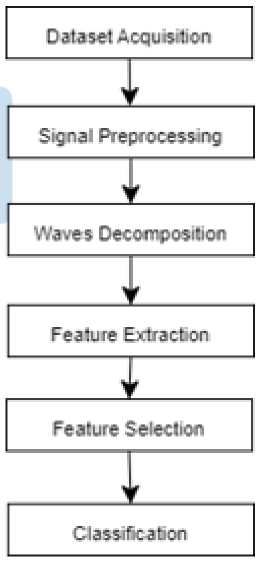

Fig. 1. Research Steps

A. Data Acquisition

This study used a publicly available dataset called MODMA, which stands for Multi-modal Open Dataset for Mental-disorder Analysis. This dataset is collected by Lanzhou University in China[7].

The EEG signal was recorded using a 24-bit analog-to-digital converter with a sampling frequency of 250Hz. The total number of participants is 55, consisting of 26 patients diagnosed with major depressive disorder (MDD) and 29 individuals in the healthy control group. Among the patients, there are 15 males and 11 females, aged 16-56. In the healthy control group, there are 19 males and 10 females, aged 18-55.

EEG data was recorded using a three-electrode complete EEG collection equipment, as the prefrontal lobe exhibits a high correlation with emotional processes and mental disorders. The device is equipped with three electrodes that are strategically placed on the prefrontal lobe, specifically at locations Fp1, Fpz, and Fp2. A 90-second segment of EEG data was recorded while the participant was in a resting state. Subsequently, the participants were directed to remain seated with their eyes closed and minimize unnecessary physical movements for an extra minute.

Several additional questionnaires were used to validate the depression level of each patient, namely the Patient Health Questionnaire (PHQ-9) is used for diagnosing, screening, tracking, and gauging the severity of depression. The average PHQ-9 score for all depressed patients is 9.6 or rounded to 10. Thus based on Table 1, the patient's condition can be categorized as moderate depressed.

TABLE I. PHQ-9 SCORING STANDARD

Depression Score	Depressive Severity
1-4	Minimal Depression
5-9	Mild Depression
10-14	Moderate Depression
15-19	Moderately Severe Depression
20-27	Severe Depression

B. *Signal Preprocessing*

The data collected as an EEG signal was recorded includes both the EEG signal and noises called artifacts. The amplitude of the clean EEG signal is about ± 100 μV and artifacts can have an amplitude that is 10 to 100 time larger[8]. Therefore, signal preprocessing must be performed. The EEG data were filtered using Infinite Impulse Response (IIR) band pass filter. The IIR filter is recursive in nature and computes the output by incorporating current and past inputs as well as previous outputs, utilizing feedback in its structure based on the pulse transfer function to meet specified filter requirements[9].

Next, artifact data were removed using Independent Component Analysis (ICA) in order to obtain more accurate EEG data that only shows brain activity. This includes non-EEG signals such as pulse signals, muscular activity, and eye-blinking components. The ICA stage and band pass filter were conducted using Matlab R2015b with the EEGLab plugin.

C. *Waves Decomposition*

EEG signal decomposition is the process of converting signals into its simpler form. This process takes place after clean signals are obtained from the artifact removal process. EEG signals are divided into several frequency subbands such as delta (δ), theta (θ), alpha (α), beta (β), gamma (γ), alpha low, alpha high, beta low, and beta high.

A method commonly used in the decomposition process is the Butterworth Filter. This filter is first described by Stephen Butterworth in 1930. Butterworth is a type of signal processing filter that is designed to have as flat frequency response as possible, also called the maximum average magnitude filter. This filter has a better time of domain and a more stable output (with no peaks) as a result, with a better balance between smoothness and accuracy than the Chebyshev Filter. In addition to having a flatter response and no ripples in the bandpass, the Butterworth Filter also rolls toward zero as the band stops[10].

D. *Feature Extraction*

The goal of feature extraction is to identify important information from the signal to classify it accurately. This process is required to reduce the amount of the data, while retaining the essential details inside the signal. It is crucial to determine the fundamental characteristics that distinguish the dataset. Two types of features can be used for EEG signal analysis, namely linear features and nonlinear features.

Linear features in EEG data refer to patterns that can be studied through applying linear mathematical techniques. This study utilizes linear features such as the Hjorth activity parameter and statistical parameters including mean absolute value, maximum, and standard deviation. The Hjorth activity parameter is a method to describe spectral characteristics of EEG data within the time domain. The activity refers to how the signal varies[11]. There is a notable increase in activity during seizures, which means the signal deviates significantly in amplitude from its average value. Meanwhile, statistical parameters are used to define the spread of biological signals and can be calculated using the following formulas:

- Mean Absolute Value

$$MAV = \frac{\sum_{i=1}^{n} |xi|}{n} \quad (1)$$

- Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma(xi - \mu)^2}{n}} \qquad (2)$$

Traditional methods such as time-domain analysis or Fourier Transform, which are often employed for signal analysis, are inadequate for a full study of EEG signals due to their non-stationary and complicated nature. The study of the brain's dynamic nature can be explored by employing nonlinear analysis, which is based on the mathematical theory of dynamical systems[12].

- Detrended Fluctuation Analysis (DFA)

$$F[n] = \sqrt{\frac{1}{N}\Sigma_{t=0}^{N-1}(z[t] - \hat{z}[t])^2} \qquad (3)$$

- Sample Entropy

$$SampEn(X, m, r) = \log \phi^m(r) - \log \phi^{m+1}(r) \qquad (4)$$

- Correlation Dimension

The correlation dimension refers to the quantification of the dimensionality of the spatial extent encompassed by a collection of randomly distributed points. The estimation is obtained by calculating the slope of the correlation integral in relation to the range of radius of similarity. The utilization of correlation dimension as a characteristic metric is employed to differentiate between deterministic chaos and random noise, with the purpose of identifying potential defects.

In the feature extraction process, before being extracted, the signal was segmented into 20 miliseconds. The data collected has a duration of 90 seconds. Considering that the EEG instrument used has a sample frequency of 250Hz, the total data per participant equal to 22,500 lines of data (90 multiplied by 250).

Next, the 22,500 lines are divided into segments of 20ms or can be interpreted as taking every 5 lines for the purpose of calculating feature extraction. After applying a time segmentation of 20ms to the initial dataset of 22500 rows, the resulting dataset was reduced to 4500 data points per participant. Therefore, the total data obtained after the feature extraction process is 216,000 lines of data from a total of 48 participants' EEG signals.

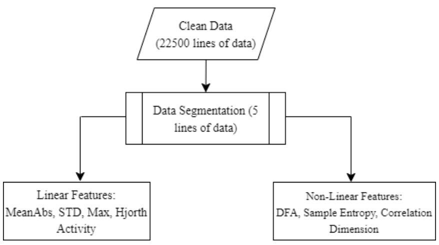

Fig. 2. Feature Extraction process using 20ms window data segmentation

E. Feature Selection

Feature selection is a crucial stage in large-scale machine learning systems. It allows developers to take advantage of the valuable feature store and address the issues and expenses that come with it. It enhances the machine learning application and system in various ways: (1) Enhanced computational efficiency: By utilizing a reduced set of features, the speed of model training and prediction is accelerated. (2) Improved prediction accuracy: This is accomplished through various methods, including the elimination of irrelevant features, prevention of overfitting, and the ability to fit a larger number of training samples into memory due to the reduced number of features. (3) Decreasing the number of features can substantially decrease the expenses associated with constructing, overseeing, and up keeping the model's feature pipeline, resulting in lower maintenance costs. (4) Simplified model interpretation and diagnosis: by exclusively using the essential feature set during the modeling process, it becomes more straightforward to comprehend the specific features and information upon which the model's prediction is founded[13].

The minimum redundancy maximum relevance (mRMR) is a feature selection approach that tends to select features with a high correlation with the class (output) and a low correlation between themselves. For continuous features, the F-statistic can be used to calculate correlation with the class (relevance) and the Pearson correlation coefficient can be used to calculate correlation between features (redundancy). Thereafter, features are selected one by one by applying a greedy search to maximize the objective function, which is a function of relevance and redundancy. Two commonly used types of the objective function are MID (Mutual Information Difference criterion) and MIQ (Mutual Information Quotient criterion) representing the difference or the quotient of relevance and redundancy, respectively[14].

This research involves three datasets that include varying numbers of selected features: 25, 50, and 75. Experiments were conducted to examine how the amount of features impacts the accuracy of the classification model. The details of the selected features are as follows:

TABLE II. NUMBER OF FEATURE SELECTED

Number of Features	Feature Details
25	'gamma2_corrdim', 'theta2_sampent', 'delta1_sampent', 'beta1_corrdim', 'theta1_dfa', 'theta1_corrdim',

	'gamma1_corrdim', 'theta3_corrdim', 'gamma3_corrdim', 'theta2_corrdim', 'beta2_dfa', 'theta1_sampent', 'alpha3_sampent', 'theta3_sampent', 'beta2_corrdim', 'beta1_sampent', 'delta3_sampent', 'beta3_sampent', 'theta2_dfa', 'gamma3_dfa', 'beta2_sampent', 'delta2_mean', 'gamma3_sampent', 'beta3_dfa', 'delta2_sampent’
50	'gamma2_corrdim', 'theta2_sampent', 'delta1_sampent', 'beta1_corrdim', 'theta1_dfa', 'theta1_corrdim', 'gamma1_corrdim', 'theta3_corrdim', 'gamma3_corrdim', 'theta2_corrdim', 'beta2_dfa', 'theta1_sampent', 'alpha3_sampent', 'theta3_sampent', 'beta2_corrdim', 'beta1_sampent', 'delta3_sampent', 'beta3_sampent', 'theta2_dfa', 'gamma3_dfa', 'beta2_sampent', 'delta2_mean', 'gamma3_sampent', 'beta3_dfa', 'delta2_sampent', 'gamma2_sampent', 'beta1_dfa', 'delta1_corrdim', 'delta3_mean', 'alpha3_corrdim', 'gamma1_dfa', 'delta1_mean', 'alpha1_corrdim', 'theta3_dfa', 'beta3_corrdim', 'alpha2_sampent', 'alpha2_dfa', 'gamma1_mean', 'delta2_corrdim', 'gamma2_dfa', 'delta2_std', 'delta3_corrdim', 'delta2_max', 'delta1_std', 'alpha2_corrdim', 'delta1_max', 'delta3_max', 'delta3_std', 'gamma2_mean', 'delta3_dfa’
75	'gamma2_corrdim', 'theta2_sampent', 'delta1_sampent', 'beta1_corrdim', 'theta1_dfa', 'theta1_corrdim', 'gamma1_corrdim', 'theta3_corrdim', 'gamma3_corrdim', 'theta2_corrdim', 'beta2_dfa', 'theta1_sampent', 'alpha3_sampent', 'theta3_sampent', 'beta2_corrdim', 'beta1_sampent', 'delta3_sampent', 'beta3_sampent', 'theta2_dfa', 'gamma3_dfa', 'beta2_sampent',
	'delta2_mean', 'gamma3_sampent', 'beta3_dfa', 'delta2_sampent', 'gamma2_sampent', 'beta1_dfa', 'delta1_corrdim', 'delta3_mean', 'alpha3_corrdim', 'gamma1_dfa', 'delta1_mean', 'alpha1_corrdim', 'theta3_dfa', 'beta3_corrdim', 'alpha2_sampent', 'alpha2_dfa', 'gamma1_mean', 'delta2_corrdim', 'gamma2_dfa', 'delta2_std', 'delta3_corrdim', 'delta2_max', 'delta1_std', 'alpha2_corrdim', 'delta1_max', 'delta3_max', 'delta3_std', 'gamma2_mean', 'delta3_dfa', 'delta2_hjorth', 'alpha3_dfa', 'gamma3_mean', 'delta1_hjorth', 'delta3_hjorth', 'gamma1_sampent', 'gamma1_std', 'theta2_hjorth', 'theta1_hjorth', 'theta3_hjorth', 'alpha1_sampent', 'gamma2_std', 'gamma3_std', 'alpha1_dfa', 'alpha1_hjorth', 'alpha2_hjorth', 'alpha3_hjorth', 'gamma1_hjorth', 'gamma2_hjorth', 'beta1_hjorth', 'beta2_hjorth', 'beta3_hjorth', 'gamma3_hjorth', 'delta2_dfa', 'alpha1_mean’

F. Classification

Following the data processing stage, the next stage is classification. The data is divided into two groups, the group diagnosed with major depressive disorder and the healthy control group. This method utilizes binary classification through machine learning. Due to the EEG data having a sampling rate of 250Hz, each second contains 250 data samples. Therefore, a classification algorithm is required to process the large amount of data. The Long Short-Term Memory (LSTM) and bidirectional LSTM are technique models that are able to process large time-series data[8]. These models utilize memory cells that possess self-connections and retain the temporal state of networks using a three-gate mechanism, including the input, output and forget gate. Different from one-way LSTM, Bi-LSTM adds a layer of reverse LSTM. The reverse LSTM reverses the data and the hidden layer synthesizes the forward and reverse information so that cells in the network can simultaneously obtain context information[15].

TABLE III. LSTM & BI-LSTM MODEL DETAILS

Model	Layer (type)	Output shape
LSTM	LSTM	(None, 1, 32)
	Dropout Layer (0.2)	(None, 1, 32)
	LSTM-1	(None, 16)
	Dropout-1 (0.2)	(None, 16)
		(None, 2)

	Dense	
Bi-LSTM	LSTM Dropout Layer (0.2) LSTM-1 Dropout-1 (0.2) Dense	(None, 1, 32) (None, 1, 32) (None, 16) (None, 16) (None, 2)

III. RESULTS AND DISCUSSION

This section contains an explanation of the experimental scenarios that have been carried out. The study involves the evaluation of three datasets using LSTM and Bi-LSTM models to determine the impact of the amount of features and the type of classification model on the obtained results.

A. Results

TABLE IV. ACCURACY COMPARISON

Classification Model	Number of Selected Features	Accuracy (%)
LSTM	25	**89.28**
	50	85.04
	75	81.87
Bi-LSTM	25	**92.83**
	50	86.59
	75	83.74

Table 4 reveals that the LSTM and Bi-LSTM classification models achieve their maximum performance when using a dataset containing 25 features. However, with further comparison, it is evident that the dataset with 25 characteristics and a bidirectional LSTM classification model achieves a superior accuracy of 92.83%. Concurrently, the Long Short-Term Memory (LSTM) model applied to a dataset containing 25 features has an accuracy rate of 89.28%.

Subsequently, an accuracy of 85.04% was achieved with 50 features, while a number of features of 25 resulted in an accuracy of 81.87%. The same trend was observed in the Bi-LSTM model. Once the dataset with 25 features achieved the highest accuracy, the subsequent accuracy values were 86.59% for a dataset with 50 features and 83.74% for a dataset with 75 features. These results suggest that the number of characteristics affects the accuracy of the categorization model. Machine learning algorithms will achieve superior performance when provided with data that has pertinent features and optimal amounts.

When comparing Figures 3 and 4, which depict the accuracy of the LSTM and Bi-LSTM models, it can be observed that the disparity in accuracy between the two models is rather small. Nevertheless, the conspicuous disparity is in the quantity of epochs. For the LSTM model, reaching an accuracy of 89.28% necessitates 50 epochs. By using Bi-LSTM, an accuracy of 92.83% may be achieved with just 30 epochs. Applying a Bi-LSTM for classification enables the sequential processing of input data in both forward and backward directions, allowing for the accumulation of relevant information while discarding irrelevant data. This approach can yield good results with reduced training time.

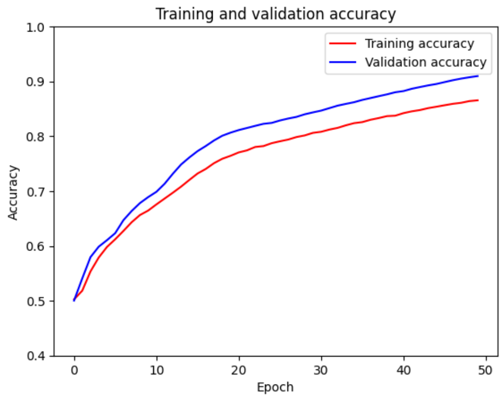

Fig. 3. The Accuracy Plot from LSTM with 25 Selected Features

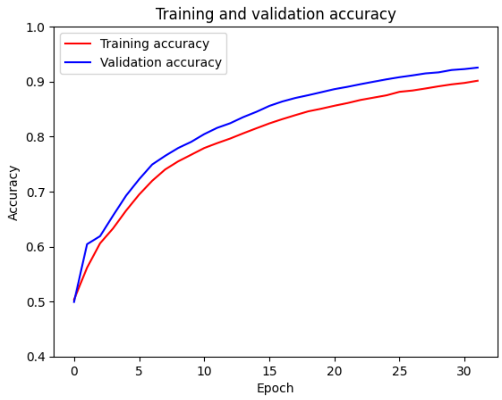

Fig. 4. The Accuracy Plot from Bi-LSTM with 25 Selected Features

However, evaluations entirely reliant on accuracy can be considered as biased. It is essential to assess additional aspects, including the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This evaluation will determine the accuracy of the predictions by assessing the proportion of testing data that was correctly predicted and the proportion that was not. This will enhance the assessment of the model.

According to the results presented in figures 5 and 6, the Bi-LSTM model has a higher number of true positives (TPs) compared to the LSTM model. Specifically, the Bi-LSTM model has 20,596 TPs, whereas the LSTM model has 20,000 TPs. According to this, the Bi-LSTM classification model has a relatively low amount of false negatives (1658) and false positives (1512). The recall, precision, and F1-

Score values in Table 5 are derived from the TP, TN, FP, and FN counts acquired from pictures 5 and 6.

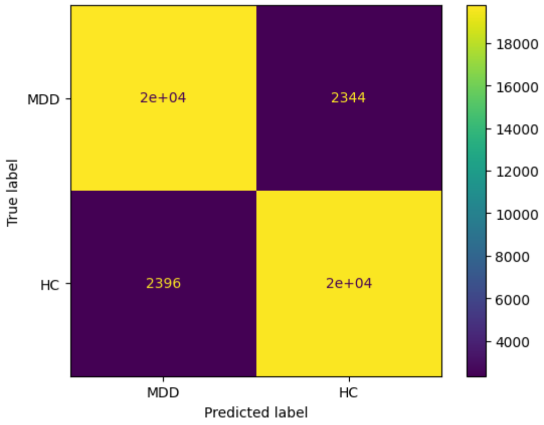

Fig. 5. Confusion Matrix from LSTM with 25 Selected Features

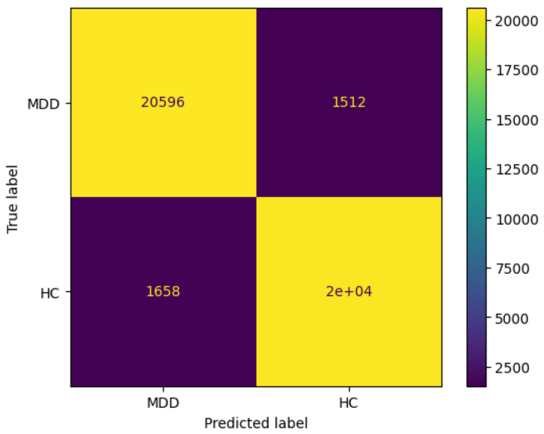

Fig. 6. Confusion Matrix from Bi-LSTM with 25 Selected Features

TABLE V. RECALL, PRECISION, & F1-SCORE

	LSTM (%)	**Bi-LSTM (%)**
Recall	89.3	92.5
Precision	89.5	93.2
F1-Score	89.4	92.8

Based on the recall, precision, and F1-Score findings of the LSTM and Bi-LSTM models, it is apparent that these three parameters have similar values or a small difference compared to the accuracy value. The trained LSTM and Bi-LSTM models can accurately differentiate between the EEG signals of patients with Major Depressive Disorder (MDD) and those of Healthy Control (HC).

B. Discussions

Several tests were conducted in this study with the objective of enhancing the precision of the depression detection system. Prior research in this field has utilized classification methods to distinguish EEG signals of patients with Major Depressive Disorder (MDD) from those without the disorder (HC). Cai et. al [5] study, which also focused on analyzing the prefrontal cortex, achieved an accuracy rate of approximately 79%. Approaches for improving accuracy in this research include comparing the amount of selected features and evaluating the performance of LSTM and Bi-LSTM classification models.

The dataset with 25 features yields the highest accuracy value when comparing the number of features in both LSTM and Bi-LSTM models. The research applies the minimum redundancy maximum relevance (mRMR) algorithm for feature selection. The algorithm yields feature rankings based on their highest level of importance. The relevance of a feature increases proportionally with its rating. In contrast, features that receive low rankings are deemed to be less relevant. Based on the obtained results, it can be inferred that the dataset with the selected features has a higher accuracy value compared to the datasets with 50 and 75 selected features. This suggests that the dataset with the selected features has a more relevant feature set.

Another finding obtained is that the accuracy value of Bi-LSTM is higher when compared to LSTM. Not only is the accuracy higher but the epochs required for the training process are fewer compared to LSTM. This is caused by the advantages of Bi-LSTM where BLSTM allows information flow in both directions, adding a new LSTM layer that inverts the sequence, and the outputs of both layers are combined, for example, with average, sum, multiplication, or concatenation. The possibility of two flow directions enables a better learning process[16].

Numerous research had demonstrated that in some classification cases, Bi-LSTM exhibits superior performance in comparison to LSTM. Some examples of research areas include text classification[17], power load forecasting[15], and financial data forecasting[18]. The comparison of Bi-LSTM and LSTM has also been shown from several studies that use EEG signals as classification data such as emotion classification[8][19] and response classification to music and sound[20]. In the domain of EEG signal classification, similar to text classification and other types of classifications, Bi-LSTM demonstrates superior performance compared to LSTM.

IV. CONCLUSIONS

This study aimed to classify EEG patterns in individuals diagnosed with major depressive disorder (MDD) and healthy controls (HC). Classification is conducted through multiple situations, specifically by considering the quantity of selected features and the classification model used. The study assessed a total of 25, 50, and 75 selected attributes. The highest accuracy value was achieved by utilizing 25 selected characteristics in both the LSTM and Bi-LSTM models, out of the three features that were chosen. Regarding the investigation comparing the LSTM and Bi-LSTM classification models, the Bi-LSTM model achieved the highest accuracy value of 92.83%. In addition to offering superior precision, it utilizes a lower epoch value compared to LSTM. These results suggest that the combination of mRMR feature selection with 25 features and the Bidirectional LSTM classification model can be employed to categorize the EEG signals of patients with Major Depressive Disorder (MDD) and healthy control (HC) individuals. In order to enhance future study, it is desirable to incorporate supplementary attributes into the EEG signal and investigate alternative techniques for feature selection.

Acknowledgment

The author would like to acknowledge the research funding by Universitas Multimedia Nusantara under the grant of numbers 0037-RDLPPM-UMN/P-INT/II/2023.

References

[1] J. Heymann and A. Sprague, "Meeting the UN Sustainable Development Goals for mental health: why greater prioritization and adequately tracking progress are critical," *World Psychiatry*, vol. 22, no. 2, pp. 325–326, May 2023, doi: https://doi.org/10.1002/wps.21090.

[2] M. É. Czeisler, "Mental Health, Substance Use, and Suicidal Ideation during the COVID-19 Pandemic," *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, no. 32, Aug. 2020, doi: https://doi.org/10.15585/mmwr.mm6932a1.

[3] H. Akbari, M. T. Sadiq, M. Payan, S. S. Esmaili, H. Baghri, and H. Bagheri, "Depression Detection Based on Geometrical Features Extracted from SODP Shape of EEG Signals and Binary PSO," *Traitement du Signal*, vol. 38, no. 1, pp. 13–26, Feb. 2021, doi: https://doi.org/10.18280/ts.380102.

[4] W. Zhang, H. Zhang, J. Huang, H. Zhou, J. Yang, and N. Zhong, "Single-Channel EEG-Based Machine Learning Method for Prescreening Major Depressive Disorder," *International Journal of Information Technology and Decision Making*, vol. 18, no. 05, pp. 1579–1603, Sep. 2019, doi: https://doi.org/10.1142/s0219622019500342.

[5] H. Cai *et al.*, "A Pervasive Approach to EEG-Based Depression Detection," *Complexity*, vol. 2018, pp. 1–13, 2018, doi: https://doi.org/10.1155/2018/5238028.

[6] D. A. Pizzagalli and A. C. Roberts, "Prefrontal cortex and depression," *Neuropsychopharmacology*, vol. 47, no. 1, Aug. 2021, doi: https://doi.org/10.1038/s41386-021-01101-7.

[7] H. Cai *et al.*, "A multi-modal open dataset for mental-disorder analysis," *Scientific Data*, vol. 9, no. 1, p. 178, Apr. 2022, doi: https://doi.org/10.1038/s41597-022-01211-x.

[8] M. Pratiwi, Adhi Dharma Wibawa, and Mauridhi Hery Purnomo, "EEG-based Happy and Sad Emotions Classification using LSTM and Bidirectional LSTM," Jul. 2021, doi: https://doi.org/10.1109/icera53111.2021.9538698.

[9] I. Grout, "Introduction to Digital Signal Processing," in *Digital Systems Design with FPGAs and CPLDs*, Elsevier Ltd, 2008.

[10] M. Shouran and E. Elgamli, "Design and Implementation of Butterworth Filter," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 9, 2020.

[11] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, and J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," *Biomedical Signal Processing and Control*, vol. 57, p. 101702, Mar. 2020, doi: https://doi.org/10.1016/j.bspc.2019.101702.

[12] H. Nee. Oon, A. Saidatul, and Z. Ibrahim, "Analysis on Non-Linear Features of Electroencephalogram (EEG) Signal for Neuromarketing Application," *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, Aug. 2018, doi: https://doi.org/10.1109/icassda.2018.8477618.

[13] Z. Zhao, R. Anand, and M. Wang, "Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform," *IEEE Xplore*, Oct. 01, 2019. https://ieeexplore.ieee.org/abstract/document/8964172.

[14] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC Bioinformatics*, pp. 1–14, 2017, doi: 10.1186/s12859-016-1423-9.

[15] J. Du, Y. Cheng, Q. Zhou, J. Zhang, X. Zhang, and G. Li, "Power Load Forecasting Using BiLSTM-Attention," *IOP Conference Series: Earth and Environmental Science*, vol. 440, no. 3, p. 032115, Feb. 2020, doi: https://doi.org/10.1088/1755-1315/440/3/032115.

[16] Silva and A. Alvarenga, "Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction," *Energy Reports*, vol. 10, pp. 3315–3334, Nov. 2023, doi: https://doi.org/10.1016/j.egyr.2023.09.175.

[17] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling," *ACLWeb*, Dec. 01, 2016..

[18] S. Siami-Namini, N. Tavakoli, and Akbar Siami Namin, "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," *arXiv (Cornell University)*, Nov. 2019.

[19] J. Yang, X. Huang, H. Wu, and X. Yang, "ScienceDirect ScienceDirect EEG-based classification based on Bidirectional International emotion Knowledge in the Long Short-Term Memory Network EEG-based classification based Bidirectional Jinru emotion Wu on Short-Term Memory Network," *Procedia Comput. Sci.*, vol. 174, no. 2019, pp. 491–504, 2020, doi: 10.1016/j.procs.2020.06.117.

[20] I. Ariza, A. M. Barbancho, L. J. Tardón, and I. Barbancho, "Energy-based features and bi-LSTM neural network for EEG-based music and voice classification," *Neural Computing and Applications*, Oct. 2023, doi: https://doi.org/10.1007/s00521-023-09061-3.