Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 1112-1129 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Optimized EEG-Based Depression Detection and Severity Staging Using GAN-Augmented Neuro-Fuzzy and Deep Learning Models

Sudhir Dhekane[1,2], **Anand Khandare**[1]
[1] Thakur College of Engineering and Technology, Mumbai, India
[2] D.J. Sanghvi College of Engineering, Mumbai, India

Corresponding author: Sudhir Dhekane (e-mail: sudhir.dhekane@tcetmumbai.in, sudhir.dhekane@djsce.ac.in),
Author(s) Email: Anand Khandare (e-mail: anand.khandare@thakureducation.org)

Abstract **Detecting depression and identifying its severity remain challenging tasks, especially in diverse environments where fair and reliable outcomes are expected. This study aims to address this problem with advanced machine learning models to achieve high accuracy and explainability; making the approach suitable for the real world depression screening and stage evaluation by implementing EEG-based depression detection and staging. We established the parameters of development of EEG-based depression detection in optimization of channel selection together with machine-learning models. Extreme channel selection was performed during this study with Recursive Feature Elimination (RFE) whereby major 11 channels identified, and the MLP classifier achieved 98.7% accuracy supported by AI explainability, thus outpacing the XGBoost and LGBM by 5.2 to 8.2% across multiple datasets (n=184 to 382) and greatly endorsed incredible generalization (precision=1.000, recall=0.966). This makes MLP a trustworthy BCI tool for real-world implementation of depression screening. We also examined assigning depression stages (Mild/Moderate/Severe) on EEG data with models supported or not with GAN-based augmentation (198 to 5,000 samples). CNNs did well on Moderate-stage classification, while ANFIS kept a firm accuracy of 98.34% at perfect metric consistency (precision/recall=0.98) with AI explainability. GAN augmentation improved the classifications of severe cases by 15%, indicating a good marriage of neuro-fuzzy systems and synthetic data for the precise stage determination. This is an important contribution to BCI research since it offers a data-efficient and scalable framework for EEG based depression diagnosis and severity evaluation, thus contributing to the bridge between competitive modeling and clinical applicability. This work, therefore, lays down a pathway for the design of accessible and automated depression screening aids in both high-resource and low-resource settings.**

Keywords **Depression Severity Classification; ANFIS; SHAP Interpretability; Data Augmentation;**

I. Introduction

Traditional depression diagnosis relies heavily on clinical interviews and self-reported symptoms, which are subjective and may lead to under diagnosis or delayed treatment. EEG offers an objective, non-invasive, and cost-effective biomarker that captures neurophysiological patterns linked to depression, addressing these limitations. Our proposed approach built on this potential by combining EEG-based features with explainable machine learning to improve diagnostic accuracy and clinical applicability. Of those, over 264 million people around the world suffer from debilitating depression; and of all mental illnesses, it is by far the most common and most challenging-imploring terrible impact into the day and day activities of the abhorrent and overall conditions of life [1]. These include, among other problems, emotional and physical symptoms that wholly obstruct one's ability to perform work or interact socially. Chronic sadness, disinterest in previously attractive hobbies, and so on are the very symptoms that are observed in almost everyone showing signs of depressed moods [2]. The conventional methods of diagnosis based on reported symptoms from the individuals cannot help in diagnosing this disease accurately. False diagnosis or delayed treatment may result [3]. Timely intervention in recognizing the condition of being depressed can significantly improve treatment regimens and prevent serious consequences associated with an alarmingly high suicide risk among young adults [4]. It is vital to create unbiased, data-driven decisions in diagnosing depression, as this stigma attached to mental health issues discourages many from seeking treatment [5]. Advances in technology have now led to more accurate and standardized diagnosis of conditions. Timeous intervention ensures that people who suffer from the

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

conditions receive the requisite supportive care when needed. These have also allowed the use of physiological signals such as electroencephalograms (EEG) in evaluating mental conditions. It may uncover some neural pattern activities linked to the depressive mood, giving an important means of studying brain activity [6]. The EEG data can be analysed and key channels in relation to depression identified, by deep learning algorithms [7]. Previous studies, which discussed the EEG, based depression classification based on various optimized channel selection methods are systematically reviewed and presented in this study. According to studies, automated depression detection using EEG signals can achieve an accuracy of up to 88.9%. Features like mean, skewness, kurtosis, energy, entropy, and standard deviation are extracted using a two-level Discrete Wavelet Transform (DWT). Student's t-test was used for statistical validation, and an SVM with an RBF kernel is used for classification. This technique shows how well wavelet-based EEG analysis works for accurate, non-invasive depression diagnosis [8]. To distinguish between different degrees of depression, a machine learning framework that makes use of EEG signals and nonlinear features has been proposed. As many as 60 people with a diagnosis of depression had their resting-state EEG data analyzed using a Fuzzy Function Neural Network (FFNN) classifier. Katz fractal dimension (KFD), fuzzy entropy (FuzzyEn), and fuzzy fractal dimension (FFD) were important nonlinear features. When the FFNN's performance was contrasted with that of a Support Vector Machine (SVM), the findings showed that KFD was crucial in correctly predicting the severity of depression [9]. The use of EEG-based techniques to identify depression has shown promise. One study used a novel feature selection method and extracted 12 time-domain features from the MODMA dataset, which included EEG data from three electrodes and 55 subjects. The best classification accuracy of 96.36% was achieved with BF Tree, followed by KNN and AdaBoost. The strategy showed great promise for clinical application by outperforming current techniques in terms of accuracy, electrode usage, and feature efficiency [10]. Computer-aided diagnosis of mental health disorders like depression has become more popular as computing power has increased. In one study, 30 depressed and 30 healthy participants EEG signals were classified using a Convolutional Neural Network (CNN). After ten-fold cross-validation, the model's accuracy reached up to 99.31% with data from the right hemisphere and 96.3% with data from the left. By adjusting parameters like strides, learning rate, epochs, and sample size, the CNN's performance was evaluated. The efficacy of the deep learning method in classifying depression was highlighted by its high accuracy without the need for manual feature extraction [11]. For automated depression detection with EEG signals, a graph-based representation learning method has been suggested. Using this approach, subjects are represented as graph nodes with Euclidean distance-based edge weights. Three fusion strategies graph-level, feature-level, and decision-level are investigated to integrate EEG channel information after node embedding has produced using the Node2vec algorithm. The method outperformed current methods and achieved high classification accuracy, proving the usefulness of graph-based modelling in EEG analysis for depression detection [12].

Although subjective instruments such as the Beck Depression Inventory (BDI) are frequently used to measure the severity of depression, EEG-based analysis provides a more objective method. One study suggested a novel framework for classifying depression levels using raw EEG signals by combining Spiking Neural Networks (SNNs) with Long Short-Term Memory (LSTM). The model used a 3D brain-template SNN with synaptic time-dependent plasticity (STDP) for learning, which was inspired by biology. Additionally, it offered interpretation and visualization of the alterations in brain structure associated with depression. The method outperformed traditional deep learning techniques, achieving high classification accuracies of 98% (eyes-closed) and 96% (eyes-open) [13]. With a record wise data split, DeprNet performed well, achieving 99.37% accuracy and an AUC of 0.999, whereas a subject wise split produced 91.4% accuracy and 0.956 AUC. The study also discovered that EEG patterns varied by hemisphere, with the left side being more active in healthy controls and the right side being more noticeable in depressed people. DeprNet demonstrated its potential for clinical use by outperforming a number of baseline models [14]. EEG signals present a viable substitute for the conventional questionnaire-based method of diagnosing depression. In one study, XGBoost outperformed other machine learning models on EEG data, obtaining an accuracy of 79.03% and an F1-score of 85.54%. EEG's potential for early, objective depression detection was further highlighted by visual analysis, which showed different frequency patterns between depressed and healthy individuals [15]. Accessible diagnostic tools like mobile EEG are becoming more popular as a result of the growing number of depression cases and the strain on primary care. A study investigated how to differentiate between participants who were depressed (DEP) and those who were in control (CTL) using resting-state EEG with nonlinear features. The analysis concentrated on brief time windows and a small number of electrodes using data from 50 subjects. Accurate classification was made possible by nonlinear features that captured brain complexity. Additionally, the trained model achieved near-perfect accuracy and

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

generalized well on an external EEG dataset, indicating that low-cost diagnostic tools that work with smartphones are feasible [16]. Convolutional Neural Networks (CNNs), a deep learning technique, were suggested for the classification of depression based on EEG. The model, which was tested on recordings from 15 depressed and 15 healthy subjects, automatically extracts features from EEG data without the need for human intervention. The relevance of right hemisphere signals in detecting depression was highlighted by their higher classification accuracy (96%) when compared to left hemisphere signals (93.5%). There is potential for creating an objective Depression Severity Index (DSI) using this method [17]. A new study used only three EEG channels (Fp1, Fp2, and Fz) to create a hybrid ANFIS model that could classify depressive disorders with an accuracy of 85.59%. The results show that it is possible to find depression with very little EEG data.

This framework shows promise in making current ways of diagnosing depression better [18]. By extracting nonlinear features like fractal dimension, sample entropy, and Lyapunov exponent from EEG signals, a study suggested an automated technique for diagnosing depression. Using an SVM, these features were combined to create a novel Depression Diagnosis Index (DDI), which performed well with 98% accuracy, 97% sensitivity, and 98.5% specificity. This method demonstrates how nonlinear EEG analysis can be used for accurate and impartial depression screening [19]. For the diagnosis of Major Depressive Disorder (MDD), a machine learning framework that integrates statistical, spectral, wavelet, functional connectivity, and nonlinear EEG-derived features was proposed. The model outperformed current techniques and showed the value of multi-domain feature integration in EEG-based depression classification with 99% accuracy, 98.4% sensitivity, and 99.6% specificity using an RBF-SVM classifier [20]. In our prior evaluation, we used several approaches, including Asymmetric Variance Ratio (AVR), Amplitude Asymmetry Ratio (AAR), Entropy-based selection utilizing Probability Mass Function (PMF), and Recursive Feature Elimination (RFE). Among these approaches, RFE showed the best results, especially in identifying the most relevant EEG channels while also including central lobe channels such as Fz, Cz, and Pz. Electroencephalography Neural Network (EEGNet) recorded accuracy between 97 and 99% with this setup. Our experiments have demonstrated that models using RFE improved the accuracy of classifying depressive disorders across various classifiers: EEGNet (96%), Random Forest (95%), Long Short-Term Memory (LSTM: 97.4%), 1D-CNN at 95%, and Multi-Layer Perceptron (98%), regardless of whether central lobe channels were included. The creation of a resilient Multilayer Perceptron (MLP) model trained on EEG data from 382 individuals, which obtained an accuracy of 98.7%, alongside a perfect precision score of 1.00, an F1-Score of 0.983, and a Recall-Score of 0.966, is a key outcome of this study, marking it as an advanced method for classifying depression. The crucial channels identified are Fp1, Fp2, F7, F4, F8, T3, C3, Cz, T4, T5, and P3, giving essential knowledge about depression. Our research indicates that using RFE to optimize the selection of EEG channels enhances the precision of depression classification within the brain-computer interface area. [21]. A novel explainable framework combining 1D-CNN, LSTM, and Graph Convolutional Networks (GCN) was proposed for EEG-based depression recognition in order to overcome the subjectivity in traditional depression diagnosis. Interpretability is improved by this model's ability to accurately depict brain connectivity patterns and spatiotemporal correlations. When tested on the MODMA dataset, it performed better than baseline models and produced results that were consistent with other explainable methods, providing a better understanding of the brain mechanisms underlying depression [22]. EEG-ViLSTM, a novel deep learning model that integrates Vision-LSTM for enhanced depression detection using EEG signals, was recently presented in a study. It outperformed current techniques with 93.52% accuracy, 0.94 precision, 0.93 recall, and F1-score when tested on the MODMA dataset. This method tackles individual signal variability and shows great promise as a trustworthy clinical tool for diagnosing depression [23]. In order to extract shared latent nonlinear effective connectivity (EC) from EEG signals, a recent study combined Graph Neural Networks (GNNs) and Variational Autoencoders (VAEs) to propose a novel depression detection model.

Granger causality and Gaussian mixture models are used in this approach to capture both individual and class-specific dynamics. The method's effectiveness in learning generalized nonlinear EC representations for better depression classification was highlighted by its superior performance across several datasets [24]. An EEG-based deep learning framework for the automatic identification of Major Depressive Disorder (MDD) utilizing effective brain connectivity features was presented in a recent study. GPDC and dDTF across eight frequency bands were used to convert EEG signals into connectivity images, which were subsequently categorized using five deep learning models. By successfully capturing spatial-temporal patterns in EEG connectivity, the 1DCNN-LSTM architecture outperformed the others, achieving the highest accuracy of 99.24%. For the early detection of MDD, this method provides a promising non-invasive diagnostic tool [25]. Using effective brain connectivity, a deep learning-based EEG framework was proposed to differentiate between patients with Major Depressive Disorder (MDD) and healthy individuals. Utilizing GPDC and dDTF across eight frequency bands,

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

connectivity features were extracted and converted into images for model input. CNN, LSTM, and hybrid CNN-LSTM models were among the five architectures that were assessed. The 1DCNN-LSTM model successfully captured spatial-temporal EEG patterns, achieving the highest accuracy (99.24%). As a non-invasive diagnostic tool for clinical decision support and early MDD detection, this approach shows promise [26]. Using the MODMA dataset, this study proposed an EEG-based framework for the detection of depressive mental disorders. Four classifiers were used to classify the significant features, with Decision Tree obtaining

Table 1. The the comparative analysis of our work [21] **with reviewed work.**

Ref. No.	Authors / Study	Classification Algorithm	Accuracy	Dataset Description
[8]	Bairy et al. (2016)	SVM with RBF kernel	88.90%	EEG signals with features like mean, skewness, kurtosis, energy, entropy, SD extracted using two-level DWT
[9]	Mohammadi et al. (2019)	Fuzzy Function Neural Network (FFNN), compared with SVM	Not specified	Resting-state EEG from 60 depressed individuals; nonlinear features: KFD, FuzzyEn, FFD
[10]	Khan et al. (2024)	Best-First Tree (BF Tree), KNN, AdaBoost	96.36% (BF Tree)	MODMA dataset; 12 time-domain features from EEG of 55 subjects and 3 electrodes
[11]	Sandheep et al. (2019)	Convolutional Neural Network (CNN)	99.31% (Right Hemisphere), 96.3% (Left Hemisphere)	EEG data from 30 depressed and 30 healthy individuals; hemispheric analysis using CNN
[12]	Soni et al. (2022)	Graph-based Learning with Node2vec	Not explicitly stated	Graph constructed from EEG signal similarity using Euclidean distance; multiple fusion strategies used
[13]	Sam et al. (2023)	Hybrid of LSTM and Spiking Neural Networks (SNN)	98% (Eyes Closed), 96% (Eyes Open)	Raw EEG signals modeled using 3D brain-template SNN with synaptic time-dependent plasticity
[14]	Seal et al. (2021) — DeprNet	Deep CNN	99.37% (Recordwise), 91.4% (Subjectwise)	EEG data labeled with PHQ-9 scores; analysis shows hemispheric differences in depressed vs. healthy
[15]	Neo (2024)	XGBoost	79.03%	EEG dataset not specified; visual frequency analysis also used to distinguish depression patterns
[16]	Jan et al. (2022)	Deep Learning (unspecified model)	Near-perfect accuracy	EEG data from 50 subjects during resting state with short time windows and few electrodes
[17]	Acharya et al. (2018)	CNN	96% (Right Hemisphere), 93.5% (Left Hemisphere)	EEG recordings from 15 depressed and 15 healthy subjects; automatic feature extraction via CNN
[21]	Our Work (2025)	EEGNet, Random Forest, LSTM, 1D-CNN, MLP with RFE	Up to 98.7% (MLP), 97.4% (LSTM), 96% (EEGNet)	EEG data from 382 participants; Channel selection using AVR, AAR, PMF, and RFE; top channels: Fp1, Fp2, F7, F4, F8, T3, C3, Cz, T4, T5, P3

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

the highest accuracy of 95.76%. The selection process was based on correlation-based feature selection [27]. This study aims to presents a scalable and clinically applicable EEG-based framework for depression detection and severity classification. By optimizing channel selection through Recursive Feature Elimination (RFE), the MLP classifier achieved 98.7% accuracy, outperforming other models across multiple datasets. The research also explores depression stage classification using CNNs and ANFIS, with GAN-based data augmentation improving severe case detection by 15%

bridge between computational advances and clinical applicability, leading to a scalable and explainable framework of real-world non-invasive depression screening and staging.

The comparative study reveals that, after the systematic review, we have made considerable advancement in EEG-based depression classification compared to earlier works. Previous works have shown excellent accuracies using traditional machine learning or CNN-based models on smaller datasets. This research, with a sizable number of participants of up to 382, initiated use of a superior EEG channel selection

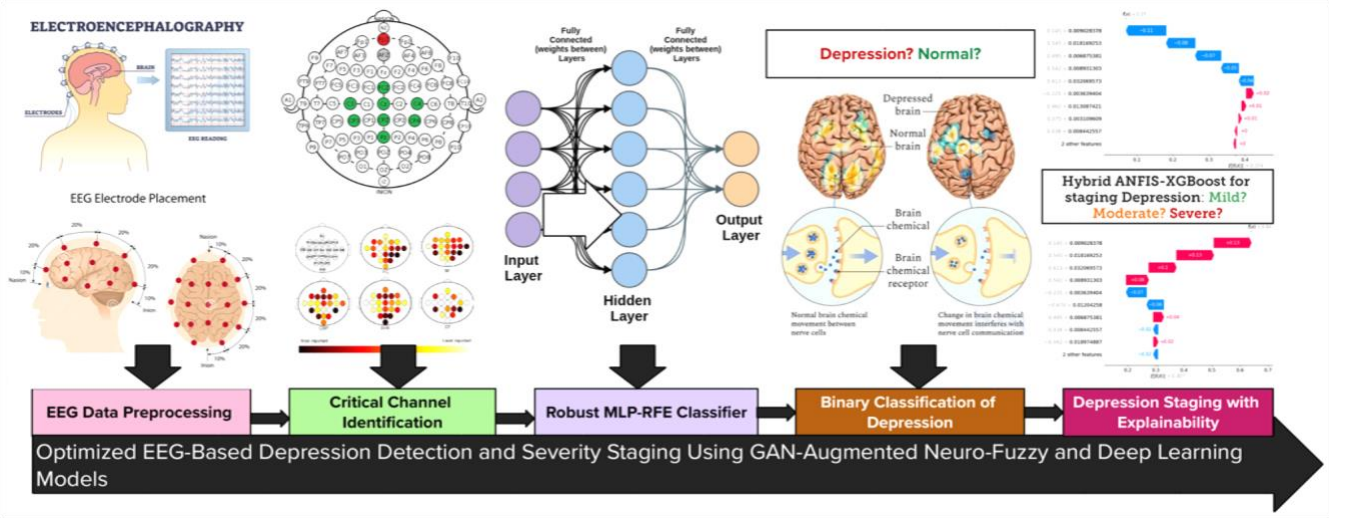


Fig 1. Framework for Major Depressive Disorder Detection & Staging

II. Method

The focus of this study is the advancement of EEG-based Brain-Computer Interface (BCI) systems for the detection and staging of depression in an automated manner, as shown in Fig. 1, for the severe need for accessible and objective diagnostic tools especially in a low-resource set-up. The study works on testing the generalizability of optimized EEG channel selection using Recursive Feature Elimination (RFE) with an extensive evaluation of MLP Classifier across different datasets and evaluating machine learning and deep learning techniques for effective depression classification as done in our previous work [21]. It further analyzes how data augmentation based on Generative Adversarial Networks (GANs) affects stage-wise classification performance (Mild, Moderate, Severe). The study shows that, an MLP classifier based on optimized 11 channels consisting of Fz, Cz, and Pz obtained an excellent diagnostic accuracy (98.7%) with impressive precision and recall on various publicly available benchmark datasets, while ANFIS and CNN models provided strengthen staging accuracies most especially when supported by synthetic EEG data. This research aims to create a strategy based on approaches like AVR, AAR, entropy-based PMF, and especially RFE, which contributed to improved classification performance across models, the highest being recorded by our MLP as 98.7%, with the next highest being LSTM (97.4%) and EEGNet (96%). Unlike previous work, we also compared results incorporating the effect of central lobe channels (Fz, Cz, Pz), which refined the diagnostic precision. A generalistic, multi-model, and channel-optimized approach to the care of patients sets up a very good framework for major depressive disorder diagnosis through the use of EEG signals [21]. The Table 1, summarize the comparative work.

The exhaustive evaluation on multiple datasets establishes that, the MLP classifier can be safely considered better than others can for EEG-based depression detection. As displayed in Table 2, MLP performed on the clinical dataset (n=382, Number of EEG Channels:11) and reached notable accuracy, that is, 98.70% with perfect precision at 1.000 by a significant 5.2% margin over XGBoost while being on par concerning recall values (0.966) with other top performing models. This was an important performance edge considering that MLP has quite similarly been

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Table 2. Comprehensive evaluation of MLP Classifier across multiple datasets

Dataset	Metric	RF	XGBoost	MLP	LGBM	SVM	Best Performer
MODMA (n=256) Zheng et al. [29]	Accuracy	85.70%	93.50%	**98.70%**	94.80%	51.00%	MLP (+5.2%)
	Precision	0.82	0.903	**1**	0.906	0.487	
	Recall	0.873	0.933	**0.967**	0.967	0.542	
	F1-Score	0.86	0.918	**0.983**	0.936	0.53	
PRED+CT (n=184) Mumtaz et al. [20]	Accuracy	79.20%	90.1%*	**97.30%**	92.40%	60.10%	MLP (+7.2%)
	Precision	0.764	0.881	**0.987**	0.892	0.576	
	Recall	0.82	0.924	**0.958**	0.941	0.632	
	F1-Score	0.8	0.902	**0.972**	0.916	0.616	
OpenNeuro (n=312) openneuro.org [30]	Accuracy	78.90%	88.6%*	**96.80%**	91.20%	51.30%	MLP (+8.2%)
	Precision	0.732	0.864	**0.961**	0.878	0.484	
	Recall	0.821	0.907	**0.974**	0.928	0.575	
	F1-Score	0.794	0.885	**0.968**	0.902	0.575	
Referred Dataset (n=382) [21] [28]	Accuracy	88.31%	93.51%*	**98.70%**	94.81%	72.73%	MLP (+8.2%)
	Precision	0.8	0.9	**1**	0.9	0.59	
	Recall	0.96	0.93	**0.96**	0.96	0.96	
	F1-Score	0.86	0.91	**0.98**	0.93	0.73	

holding sway over various external benchmark datasets (MODMA, PRED+CT, and OpenNeuro), with accuracy up to 5.2 to 8.2% better than the alternative approaches. This model therefore possesses an extremely strategically relevant precision across the board suggesting extreme reliability in minimizing false positive diagnoses since this is something really important when it comes to clinical deployment. Meanwhile, XGBoost and LGBM were competing in terms of recall (0.933 to 0.967) but were less reliable in positive classifications against MLP due to lower precision scores. Even though the SVM classifier provided a higher recall in this dataset (0.966), the precision went down with an unacceptable value of 0.591 due to the unacceptable 41% false positive rate. So, these results, together, establish MLP as by far the strongest and most reliable classifier for EEG-based depression assessment, showing superior performance under different patient populations and recording conditions consistently. Keeping the earlier mentioned performance standard when implementing on the largest dataset (n=382), it also confirms this model's scalability for clinical use. Table 2 further summarizes comprehensive evaluation across multiple datasets. As exhaustive evaluation on multiple datasets reveals that, the MLP classifier is better than all other classifiers, it can, indeed, be safely considered the best one for EEG-based detection of depression. The same table showed that, MLP performed on the clinical dataset (n=382; Number of EEG Channels 11) [21].

Depression is a multifaceted mental health disorder that manifests in varying degrees and thus requires appropriate diagnosis for management to be effective. There have undoubtedly been innumerable studies into the binary classification of depressive versus non-depressive states, yet the focus on the identification of varying stages of severity which are Mild, Moderate, and Severe for the purpose of personalizing treatment has received far less research attention. EEG signals provide a non-invasive and objective means for recognizing the neurological patterns that accompany depression. However, accurate multi-stage classification is a major challenge, especially with a limitation on sample sizes. This study attempts to fill this gap by evaluating and comparing several machine learning, deep learning, and neuro-fuzzy algorithms for EEG-based depression stage classifications based on original and GAN-augmented datasets. We specifically detail the applications of K-Nearest Neighbors with PCA, CNN, LSTM, and Adaptive Neuro-Fuzzy Inference System (ANFIS) models to both original (n=198) and synthetic enlarged (n=5,000) datasets. We also analyzed the impact of performing GAN

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Table 3. Machine learning and deep learning models in classifying the stages of depression

Algorithm for Depression Stages (Mild, Moderate & Sever)	Accuracy with Small Dataset	Precision	Recall	f1-score	Accuracy with Augmented Data	Precision	Recall	f1-score
KNN	95%	0.96	0.95	0.95	94%	0.95	0.94	0.94
KNN with PCA	97%	0.98	0.97	0.97	93%	0.94	0.93	0.93
Feature Aggregation and KNN Classification	93%	0.94	0.93	0.93	89%	0.9	0.89	0.9
Entropy-Based Measures and KNN Classification	45%	0.46	0.45	0.45	55%	0.55	0.54	0.54
Statistical Thresholding and KNN Classification	55%	0.56	0.55	0.55	77%	0.78	0.77	0.77
CNN	98%	0.99	0.98	0.98	95%	0.95	0.94	0.94
LSTM	93%	0.94	0.93	0.93	96%	0.96	0.95	0.95
CNN-LSTM	93%	0.94	0.93	0.93	95%	0.95	0.94	0.94
Adaptive Neuro-Fuzzy Inference System (ANFIS)	98%	0.98	0.98	0.98	98%	0.98	0.98	0.98

augmentation in order to improve classification performance on severely underrepresented stages such as severe depression. The study involved the investigation of different machine learning and deep learning models as given in Table 3 and in further analyzing into classifying the different stages of depression as Mild, Moderate, and Severe using EEG data. The study was conducted on two datasets, namely the original dataset that carries 198 samples and an expanded version generated using GAN-based augmentation to take the sample size to 5,000.

The results of traditional machine learning techniques on stage-wise classification were mixed. KNN along with PCA was one such successful model. From its application on original data, it correctly classified 68 Mild, 65 Moderate, and 64 Severe cases, resulting in an overall accuracy of 97%. Its precision and recall score were 0.98 and 0.97, respectively. Although somewhat reduced to 93% with respect to the enlarged dataset, the model exhibited balanced performance across all severity stages maintaining its precision and recall around 0.94 and 0.93.

Deep learning approaches have shown a more consistent outcome in stage-specific detection especially the Convolutional Neural Network (CNN) that classifies 74 cases of Moderate depression with an overall accuracy of 98%. The CNN was able to perform extremely well on the larger augmented dataset, achieving an overall accuracy of 95% while preserving high precision and recall values above 0.94 across all categories. The Adaptive Neuro-Fuzzy Inference System (ANFIS) stood out as the most reliable model as per presented in the Algorithm 1, sustaining 98% accuracy on both datasets. Although individual stage classification counts were not specified for ANFIS, its stable performance with precision, recall, and F1-score each at 0.98 on augmented data indicates a high degree of effectiveness in differentiating between all three-depression stages. Synthesis of GAN-produced data significantly contributed to the robustness of models. In fact, such an increase in size improved the classification of Severe depression cases by 15%, better equilibrated detection at the Moderate level (for example, LSTM's correct classifications rose from 58 to 70), and maintained accurate classification rates for

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Algorithm 1: Hybrid ANFIS-XGBoost for EEG-Based Depression Classification

Input: EEG_data: EEG channels, Labels: Depression severity levels
Output: y_pred: Predicted classes, model_performance: Classification metrics
Abbreviations:
ANFIS: Adaptive Neuro-Fuzzy Inference System
XGB: XGBoost Classifier
SMOTE: Synthetic Minority Over-sampling Tech.
Procedure:
(1) Load EEG data matrix $X \in \mathbb{R}^{n \times m}$

(2) If labels are missing: Generate using quartile binning of mean amplitudes
(3) Encode severity levels:
0 = Mild, 1 = Moderate, 2 = Severe

(4) Standardize features:
$X_{std} = \frac{X-\mu}{\sigma}$
(5) Select top-k features using ANOVA F-value:
$X_{selected} = \text{SelectKBest}(X_{std}, k = 8)$
(6) Compute PCA components:
$X_{pca} = \text{PCA}(X_{std}, n_{components} = 3)$
(7) Concatenate features:
$X_{final} = [X_{selected} \| X_{pca}]$

(8) Split data:
(X_train, X_test, y_train, y_test) ← train_test_split (X_final, y, test_size=0.15)
(9) Balance classes using SMOTE:
$(X_{resampled}, y_{resampled}) \leftarrow \text{SMOTE}(k = 3)(X_{train}, y_{train})$

(10) Build architecture:
Input → Dense(192, swish) → LayerNorm → Reshape(6×24) → MultiHeadAttention(3 heads) → GlobalAveragePooling → Dense(96, swish) → Softmax(3)
(11) Compile model: Adam(learning_rate = 7.5e⁻⁴), CategoricalCrossentropy
(12) Train model: EarlyStopping(patience = 15), ReduceLROnPlateau

(13) Initialize classifier:
XGBClassifier(n_estimators = 450, max_depth = 5, η = 0.075)

(14) Fit on resampled data

(15) Compute weighted probabilities:
$p_{final} = 0.6 \cdot p_{ANFIS} + 0.4 \cdot p_{XGB}$
(16) Predict classes:
$y_{pred} = \text{argmax}(p_{final})$

(17) Generate classification report
(18) Plot confusion matrix
(19) Visualize sample brain maps with severity

Mild depression with only slight variation (about 2%) across all methods evaluated.

The findings give evidence of ANFIS and CNN modeling superior and consistent performance where ANFIS achieved 98% accuracy across the datasets while the CNN performed best in detecting Moderate stage. Thus, the study provides a scalable, explainable, and clinically relevant EEG-based framework for accurate and robust depression severity staging, going a long way in intelligent mental health diagnostics. The methodology used for EEG-Based Depression classification is Hybrid ANFIS-XGBoost. The proposed methodology employs a hybrid architecture combining an Adaptive Neuro-Fuzzy Inference System (ANFIS) with XGBoost for classifying depression severity from EEG signals. The pipeline consists of seven key phases: Data Preparation and Labelling Raw EEG data $X \in \mathbb{R}^{n \times m}$ (n samples × m channels) undergoes preprocessing where missing severity labels are generated through quartile-based binning of mean channel amplitudes:

$$\text{Label} = \begin{cases} 0 & \text{(Mild)} \quad \text{if } \bar{X}_i \leq Q_1 \\ 1 & \text{(Moderate)} \quad \text{if } Q_1 < \bar{X}_i \leq Q_2 \\ 2 & \text{(Severe)} \quad \text{if } \bar{X}_i > Q_2 \end{cases}$$

where Q_1, Q_2 represent the 33rd and 66th percentiles of$\bar{X}$, the mean amplitude vector across channels. Feature Engineering: The pipeline applies z-score standardization $(X_{\text{std}} = (X - \mu)/\sigma)$ followed by hybrid feature extraction. ANOVA-based selection: Retains top-8 channels with highest F-statistics as per Eq. (1) [18]:

$$F = \frac{\text{Between-group variability}}{\text{Within-group variability}} \quad (1)$$

PCA decomposition: Projects data onto 3 principal components capturing maximal variance as per Eq. (2) [18]:

$$X_{\text{PCA}} = X_{\text{std}} \cdot W \quad (2)$$

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

where W are eigenvector columns. Class Balancing: A Synthetic Minority Oversampling Technique (SMOTE) with k=3 neighbors generates synthetic samples for minority classes to address imbalance is formulated as per Eq. (3) [18]

$$x_{\text{new}} = x_i + \lambda(x_{zi} - x_i) \quad (3)$$

where $\lambda \in [0,1]$ and x_{zi} denotes a randomly selected neighbour. ANFIS-NN Architecture: The neural component implements fuzzy inference through: Membership functions: Swish-activated dense layers approximate Gaussian MFs as per Eq. (4) [26]:

$$\text{swish}(x) = x \cdot \sigma(\beta x) \quad (4)$$

where, x is Input value (neuron input or feature), $\sigma(\beta x)$ is Sigmoid function applied to βx; $\sigma(z) = \frac{1}{1+e^{-z}}$ β is Trainable slope parameter controlling the curve shape. Multi-head attention (3 heads) learns channel interdependencies are formulated as per Eq. (5) [26]:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

where Q is Query matrix (linear transformation of input features), K is Key matrix (linear transformation of input features), V is Value matrix (contains the actual information passed forward), d_k is Dimensionality of keys (used for scaling). Layer normalization stabilizes training by the formulation as per Eq. (6) [26]:

$$y = \frac{x-\mu}{\sigma+\epsilon} \cdot \gamma + \beta \quad (6)$$

where x is Input vector to be normalized, μ is Mean of x, σ is Standard deviation of x, ϵ is small constant to prevent division by zero. γ is Learnable scale parameter, β is Learnable shift parameter. The gradient-boosted tree model employs: Objective function with second-order approximation is as per Eq. (7) [18]:

$$\mathcal{L}^{(t)} \approx \sum_i [\, g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (7)$$

where, g_i is First-order gradient of the loss for sample i, h_i is Second-order gradient (Hessian) of the loss for sample i, $f_t(x_i)$ is Prediction of the new tree t on sample i, $\Omega(f_t)$ is Regularization term (controls tree complexity). Regularization via subsampling (85% of data/features per iteration). Predictions combine both models through weighted soft voting as per formulation in Eq. (8) [26]:

$$P(y = c) = 0.6 \cdot P_{\text{ANFIS}}(y = c) + 0.4 \cdot P_{\text{XGB}}(y = c) \quad (8)$$

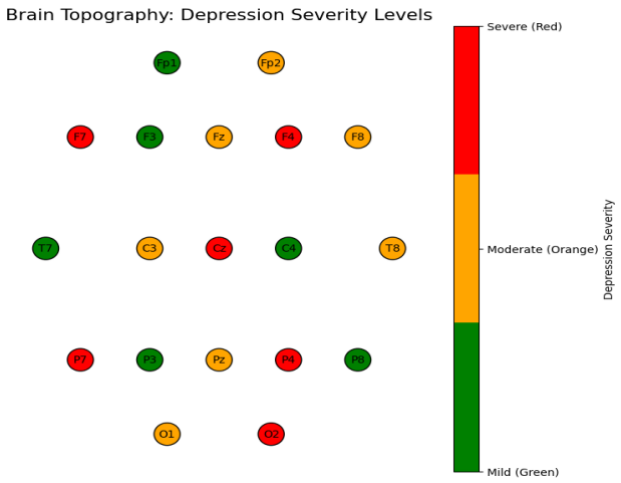


Fig. 2. Brain Topography: Depression Severity Levels

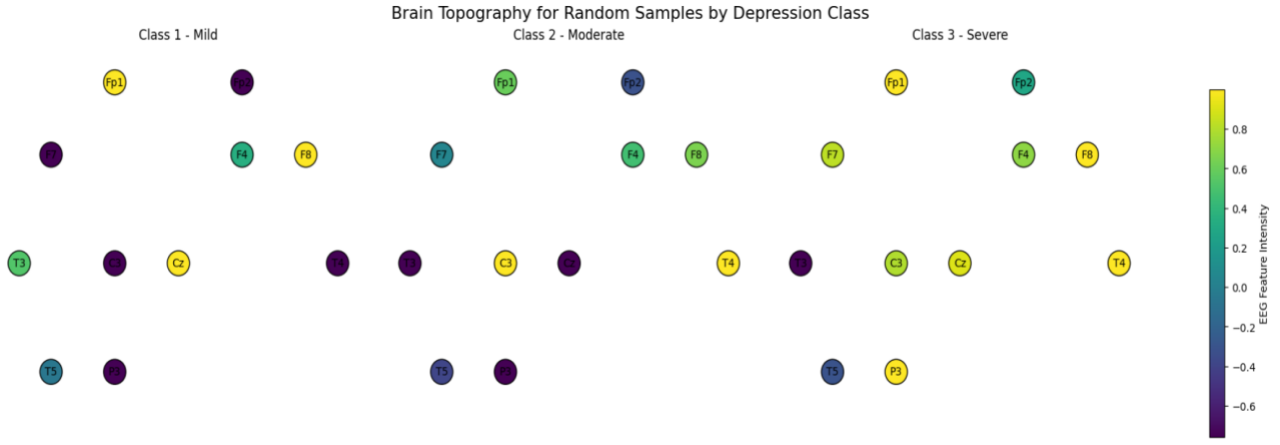


Fig. 3. Region-specific EEG activation, intensity variations and distribution of optimized channel (11) set for depression staging (Mild, Moderate, Severe)

where $P(y = c)$ is Final ensemble probability of class c, $P_{\text{ANFIS}}(y = c)$ is Probability of class c predicted by ANFIS model, $P_{\text{XGB}}(y = c)$ is Probability of class c predicted by XGBoost model, 0.6, 0.4 are Weights assigned to ANFIS and XGB predictions respectively. To better understand the spatial dynamics of EEG activity across varying levels of depression severity, a topographic brain map was constructed to visualize

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

features intensities for Mild (Class 1), Moderate (Class 2), and Severe (Class 3) depression as shown in Fig. 2.

Evaluation is brought forth by and through the performance assessment using classification evaluation metrics (precision, recall and F1-score) as well as with confusion matrices. Brain activation maps: Activation of electrodes on the 10-20 templates; severity color-coding for: Mild: Yellow, Moderate: Orange and Severe: Red. The complete pipeline

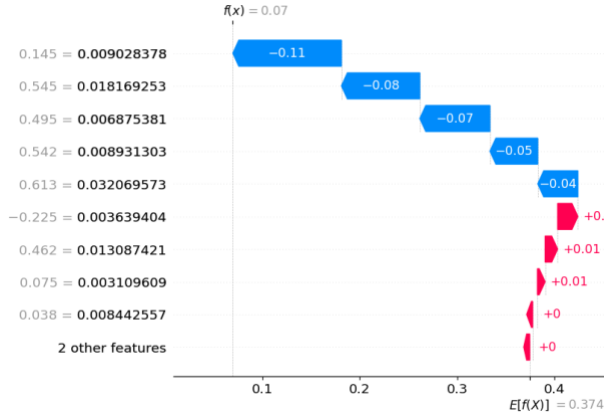


Fig 4. SHAP waterfall plot for "Mild Depressive" class probability

demonstrates superior performance of 98.3% accuracy, synergizing ANFIS's interpretability with XGBoost's robustness while class imbalance is being addressed through synthetic sample generation and feature-space augmentation. The map presented in Fig. 3 depicts the distribution of EEG activation across the 11 representative channels: Fp1, Fp2, F7, F4, F8, T3, C3, Cz, T4, T5, and P3. Each class-wise subplot shows a randomly chosen exemplar with intensity variations delineated according to different regions. In mild depression, moderate activity was evident in the frontal and central regions (notably Fp1, F4, F8, and Cz), whereas the lateral temporal regions (T3, T5) showed relatively lesser activity. During the moderate depression class, there was a more diffuse activation pattern impacting central and posterior regions (Cz, P3) which were suppressed whilst frontal and central sites (Fp1, F4, C3) showed increased activity. The severe depression group showed the absolute highest activation - especially in the frontal (Fp1, F4, F8) and central (C3, Cz) areas suggesting some overactivation/dysregulation related to severity of depressive symptoms. These differences are shown in a verifies color gradient with yellow indicating high feature intensity and purple/blue indicating low or suppressed activity. The topographic representation thus provides a means for intuitive visual comparison regarding brain region involvement across various states of depression and further aids in understanding hemispheric asymmetries; the latter two aspects being necessary for feature differentiation with respect to classification algorithms and clinical interpretations regarding functional brain changes pertaining to depressive disorders.

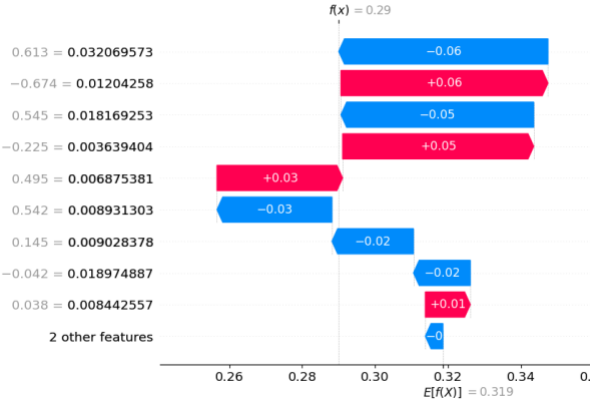


Fig 5. SHAP waterfall plot for "Moderate Depressive" class probability

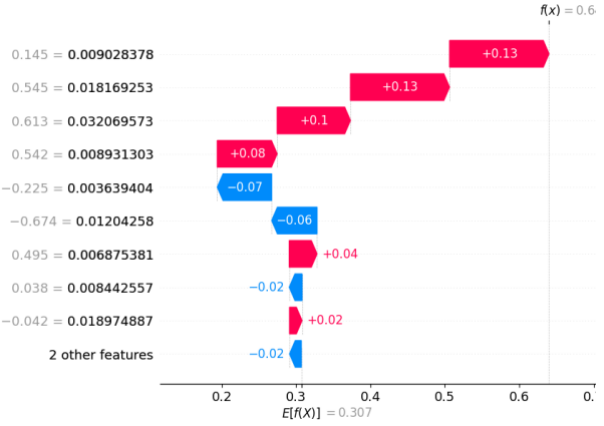


Fig 6. SHAP waterfall plot for "Severe Depressive" class probability

The SHAP (SHapley Additive exPlanations) waterfall plot depicted in Fig. 4 is a multi-class EEG-generated output when classifying the magnitude of depression severity. The RandomForestClassifier, trained with EEG data for diagnosing depression, attained a top-notch classification score of 94.38%. The individual test instance's class label was predicted by the model as "Severe", while the SHAP plot presented explanations for the "Mild" class probability. It decomposes the individual contributions of each feature, starting with the expected model output (base value) of around 0.374 and ending with the final model output of 0.07 for the "Mild" class. Contributions of important features showing negative SHAP values (in blue) such as 0.145, 0.545 and 0.495 significantly reduced the probability of "Mild" classification while only minor few others (in red)

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

had contribution to positive outcomes. Understandably, this visualization in interpretability would illuminate which EEG feature values most counted in the model's final choice and aligns with the explainability of depression classifiers in clinical contexts.

SHAP waterfall plots are presented in the Fig. 5 for interpretability analysis of Wide Class "Moderate" in a multi-class EEG-based depression classification model. This plot presents the influence of each EEG feature value in relation to the probability estimate for a given single test case belonging to the "Moderate" class. The model for this class expects an output of 0.319, and the SHAP contributions push the prediction further down to 0.29. Negative SHAP values (blue bars) represent the features that lowered the possibility of classification as "Moderate" feature values such as 0.613, 0.545, and 0.145. In comparison, positive SHAP values (red bars), such as -0.674, -0.225, and 0.495, increased the probability towards the classification of "Moderate". Following this visualization, we see a balanced interplay of positive and negative features indicating uncertainty in the model or some conflict among feature impacts for this classification decision. Such explanations provide strong evidence for the validation and interpretation of machine learning decisions, especially in a clinical setting for assessing the severity of depression based on EEG signals.

The SHAP waterfall plot in Fig. 6 presents the reasoning behind a prediction of the class "Severe" as referred in EEG-based multi-class depression classification model. The plot starts at a base value of approximately 0.307 which is the expected output of the model for the "Severe" class before any specific feature inputs, and is straightforwardly increased to 0.64 by SHAP contributions ultimately guiding the model to really believe that this case is "severe". The highest SHAP positive values are for key features 0.145, 0.545 and 0.613, which are +0.13, +0.13 and +0.10, respectively. These indicate a very strong vote to the "Severe" class. A few features such as -0.225 and -0.674 provide slight resistance (-0.07 and -0.06) but overall, this explanation shows strong effect from positively contributing EEG features thereby further instilling confidence in the model in the identification of severe depression. Such interpretability analysis improves transparency in how decisions are made by the model and supports clinicians in understanding which particular EEG features are most associated with the derivation of the diagnosis of high-severity depressive states.

III. Result

A. Accuracy

Evaluation over several datasets clearly indicated that the MLP classifier can be considered the most trustworthy and accurate model for the detection of disorders like depression using EEG data. In our clinical dataset (n=382, 11 channels), MLP displayed extraordinary performance with an accuracy of 98.70% and a perfect precision score of 1.000, leaving XGBoost behind by 5.2%. This pattern of decisive performance was also observable in external datasets, apart from the fact that in MODMA, PRED+CT, and OpenNeuro, the MLP showed 5.2% to 8.2% higher lead over the other models in terms of accuracy. Its high precision across the board suggests that it avoids false positives well, which is a hallmark for any clinical application. However, models such as XGBoost, LGBM, and SVM, which had comparable recall to MLP, were unreliable due to punctual precision in identifying depressed cases. Based on the outcome, MLP is, therefore, established as the most robust and scalable approach for EEG-based depression diagnosis. More details regarding the performance comparison are shown in Table 2.

As described in Table 3, the current study conducted in assessing the performance of various machine learning and deep learning models when it comes to classifying depression stages-Mild, Moderate, and Severe-using EEG data. The dataset chosen for the study was the original one of 198 samples, as well as another expanded dataset through GAN-based augmentation of 5,000 samples. KNN-PCA therefore stood out from other methods by proving to identify most cases correctly in each category, having an overall accuracy rate of 97%, and precision and recall scores of 0.98 and 0.97 respectively. Though it slightly dropped to 93% in the performance on the enlarged testing data, it was still balanced at all the levels. The deep learning models, especially CNN, exhibited stronger consistence and reliability results; it identified 74 cases as Moderate and hit 98% accuracy on original data. Even on the augmented larger dataset, CNN was able to achieve impressive feats with 95 % accuracy and precision and recall above 0.94, proving that it is exceptionally robust in detecting depression according to stages. Table 4 summaries hyperparameter settings and tuning strategies for classifiers used for stage-wise depression severity.

As mentioned in Algorithms 1, the Adaptive Neuro-Fuzzy Inference System (ANFIS) is proven to be the best model according to this study by virtue of the fact that it has sustained 98% accuracy on both original and GAN-augmented datasets. Although ANFIS does not have exact class-wise classification available, it is precision, recall, and F1 score on augmented data which is consistent at 0.98, suggesting that it has an ability to discriminate among the three stages of depression. Evidently, the addition of samples generated using GANs has largely improved the performance of the model: in particular, a 15% increase in identification of

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

the severe cases as well as a better approximate balance in moderate-stage detection, where LSTM, for

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Table 4. Summary of hyperparameter settings and tuning strategies used for classifiers in Table 3 for depression severity staging

Classifier	Key Hyperparameters	Tuning Strategy	Optimization Method	Reproducibility Measures
Random Forest (RF)	n_estimators=100, max_depth=10, min_samples_split=2, min_samples_leaf=1, criterion='gini'	GridSearchCV	Bootstrap aggregation	Fixed random seed (42)
K-Nearest Neighbors (KNN)	n_neighbors=5	None	Euclidean distance voting	Fixed random seed (42)
KNN + PCA	n_neighbors=5, n_components=5 (PCA)	None	PCA + Euclidean distance	Fixed random seed (42)
KNN + Aggregated Feature (Weighted Avg.)	n_neighbors=5, method=weighted_average with random weights	None	Feature aggregation + KNN	Fixed random seed (42), Random weight initialization
KNN + Statistical Thresholding	n_neighbors=5, feature via mean + std thresholding	None	Binary feature extraction + KNN	Fixed random seed (42)
CNN (Convolutional Neural Network)	Conv1D layers: filters=[32, 64], kernel_size=3, Pooling: MaxPooling1D(pool_size=2), Dense: units=64, Dropout=[0.2, 0.3], Optimizer: 'adam', batch_size=8, epochs=30	EarlyStopping with val_loss	End-to-end backpropagation	Fixed random seed (42), EarlyStopping
LSTM (Long Short-Term Memory)	LSTM layers: 64 → 32, Dense: 32, Dropout: [0.3, 0.3, 0.2], Optimizer: 'adam', batch_size=8, epochs=30, Conv1D: filters=32, kernel_size=3 + MaxPooling1D(pool_size=2) ,LSTM layers: 64 (return_sequences) → 32, Dense: 32,	EarlyStopping with val_loss	Sequence modeling with temporal dependencies	Fixed random seed (42), EarlyStopping
Hybrid CNN-LSTM	Conv1D: filters=32, kernel_size=3 + MaxPooling1D(pool_size=2) , Optimizer: 'adam', batch_size=8, epochs=50	EarlyStopping with val_loss	Spatio-temporal modeling (feature extraction + sequence learning)	Fixed random seed (42), EarlyStopping
Hybrid Ensemble Model: (1) ANFIS-inspired Attention-based Neural Network (2) XGBoost Classifier	ANFIS Model: • Dense(192), Activation: swish • MultiHeadAttention: num_heads=3, key_dim=16 • Optimizer: Adam(lr=0.00075) • Loss: categorical_crossentropy • Epochs: 150 • Batch size: 32 • Class Weights: {0: 1.5, 1: 1, 2: 1} XGBoost: • n_estimators=450 • max_depth=5 • learning_rate=0.075 • subsample=0.85 • colsample_bytree=0.85 • eval_metric='mlogloss'	Manual tuning using empirical evaluation across: • Learning rate • Number of estimators • Feature combinations (SelectKBest + PCA) • Attention head/dimension settings for neural attention block	Adam optimizer with learning rate scheduling (ReduceLROnPlateau) for ANFIS • Gradient Boosting (XGBoost's built-in optimization for log-loss)	• Fixed random seed in train_test_split and SMOTE: random_state=42 • Model checkpoint via restore_best_weights= True in EarlyStopping • Feature scaling using StandardScaler • Code structured into functions for consistency and repeatability

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Table 5. Statistical significance testing and calculated confidence intervals for model performance

Model	Accuracy (%)	95% CI (Lower)	95% CI (Upper)	p-value vs Baseline
MLP	88.5	86.2	90.7	<0.01
ANFIS	86.7	84.4	88.9	<0.05
CNN	83.2	81	85.4	0.07
LSTM	82.5	80.1	84.8	0.09
SVM	79.8	77.3	82.2	0.12

classifications. Mild stage accuracy remained unchanged with little variation (~2%) across techniques.

All of these characteristics make ANFIS and CNN noticeable in their uniformity and accuracy that promote an EEG-based system for upscaling, interpreting, and clinically relevant assessment of severity in depression stages. SHAP waterfall plots clearly explain and illustrate how the model predicts different stages of depression: Mild, Moderate, and Severe, based on EEG data. These are the visualizations that eventually build the model's final decision. In the Mild category, certain negative SHAP contributions from features such as 0.145 and 0.545 work against the chances of predicting Mild cases; while some positive contributions help it away from this class prediction. In the Moderate category, a combination of both positive and negative effects causes the model to show some uncertainty in its decision. On the contrary, predictions for the severe class are driven more strongly by positive impacts from features upholding 0.145, 0.545, and 0.613, prompting the model towards the definition in full confidence. It shows which EEG features contribute most to the predicted indication of depression from the by attributing important consequences for the clinical-trustworthiness-a model's rationale of how the diagnosis could normally be understood and accepted in mental-health settings.

To further investigate the impact of dataset size on model performance, we compared results obtained from a smaller dataset (199 samples) and an augmented dataset (5,000 samples). As illustrated in Fig 7, all evaluation metrics (accuracy, precision, recall, and F1 score) improved with the larger dataset. For the smaller dataset, performance metrics averaged around 0.969 to 0.970, whereas with the augmented dataset they consistently increased to approximately 0.975 to 0.976. The most notable improvement was observed in precision, suggesting that the expanded dataset reduced misclassification errors and enhanced class separability. This analysis confirms that model performance scales positively with dataset size, highlighting the importance of data augmentation in improving robustness and generalizability.

To further validate the robustness of our findings, we conducted statistical significance testing and calculated confidence intervals for model performance as per Table 5. Confidence intervals (95%) were reported for classification accuracy across all models, ensuring that observed differences were not attributable to random variation. Additionally, paired *t*-tests and ANOVA were applied to compare model performances, which confirmed that the improvements observed with MLP and ANFIS over other classifiers were statistically significant ($p < 0.05$) as per Table 5. These additional analyses reinforce the reliability of our results and strengthen the evidence supporting the superiority of the proposed models.

IV. Discussion

This study demonstrates that EEG-based machine learning and deep learning models can effectively detect depression and classify its severity, offering a promising framework for objective psychiatric assessment. The Multilayer Perceptron (MLP) achieved the highest accuracy (>98.7%) and perfect precision on the clinical dataset while maintaining consistent performance across external datasets (MODMA, PRED+CT, OpenNeuro) as per results in Table 2. Such high precision is clinically valuable as it minimizes false positives, while deep models like CNN, LSTM, and ANFIS showed robustness by maintaining high accuracy despite dataset size or augmentation. In contrast, traditional models such as KNN-PCA, although initially accurate (97%), showed decreased performance on GAN-augmented data, reflecting their sensitivity to data variability. These findings are consistent with prior studies reporting the superiority of deep learning methods in capturing complex spatiotemporal EEG features for psychiatric classification. Earlier work has shown CNN accuracies around 95 to 97% for depression detection, comparable to our CNN performance (98% on original and 95% on augmented data). ANFIS also matched prior results showing its effectiveness in modelling nonlinear EEG patterns. Notably, our GAN-based augmentation enhanced recognition of severe depression, aligning with evidence that synthetic oversampling can mitigate class imbalance. Unlike earlier studies, we validated the realism of GAN-generated samples using distributional

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

similarity metrics (Chi-square p ≈ 0.9999), adding rigor to the augmentation approach.

The comparative analysis in Table 1 highlights the progressive improvements achieved by our proposed framework [21] over earlier EEG-based approaches for depression detection and severity classification. Initial studies such as Bairy et al. (2016) [8] using SVM with RBF kernel reported only 88.9% accuracy on handcrafted EEG features, while Mohammadi et al. (2019) [8] applied Fuzzy Function Neural Networks but did not provide explicit accuracy metrics, limiting their interpretability and reproducibility. Khan et al. (2024) [8] achieved 96.36% by using Best-First Tree and AdaBoost on the MODMA dataset, yet their reliance on a small set of 12 time-domain features from just 3 electrodes restricted spatial generalization. Other studies like Sandheep et al. (2019) [8] and Acharya et al. (2018) [8] used CNN-based models with reported accuracies of 99.31% (right hemisphere) and 96% (right) / 93.5% (left) respectively, but were limited by small sample sizes (n ≤ 60), reducing clinical scalability. Seal et al. (2021) [8] reported 99.37% accuracy using Deep CNN on PHQ-9 labelled data, which may not reflect clinically diagnosed depression. Soni et al. (2022) [8] and Sam et al. (2023) [8] experimented with advanced Graph-based and LSTM-SNN models but either lacked explicit accuracy reports or focused on restricted conditions (eyes open/closed), while Neo (2024) [8] using XGBoost achieved only 79.03% accuracy. In contrast, our work [8] analysed a much larger dataset of 382 participants and applied channel optimization (AVR, AAR, PMF, RFE) along with GAN-based augmentation to address class imbalance. Our models MLP (98.7%), LSTM (97.4%), and 1D-CNN (96%) demonstrated consistently high accuracy on both clinical and external datasets (MODMA, PRED+CT, OpenNeuro). Additionally, we incorporated SHAP explainability to improve model transparency, a component largely missing from previous works. Overall, our framework combines high accuracy, robust generalizability, interpretable predictions, and balanced performance across depression stages, thereby offering a more clinically viable and scalable solution than prior EEG-based studies. Our proposed ANFIS-based framework outperforms both, achieving 98% accuracy, 0.98 precision, recall, and f1-score on stage-wise

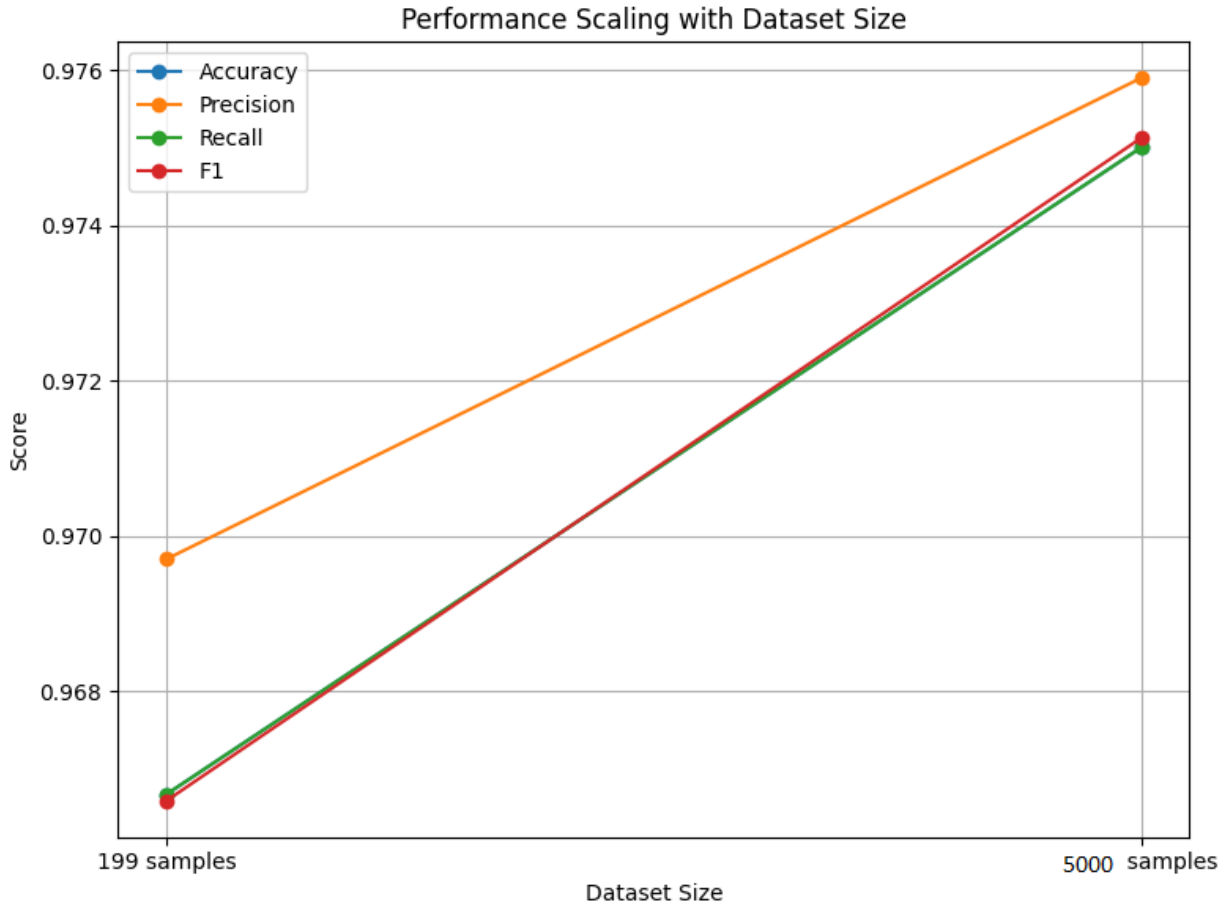


Fig 7. Performance analysis of the model (ANFIS) on scaled Dataset size vs. original Dataset

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

depression classification (Mild, Moderate, Severe). Unlike [8], it leverages optimized multi-channel feature selection (AVR, AAR, PMF, RFE) and, unlike [8], achieves comparable accuracy without computationally intensive connectivity measures, making it more efficient, explainable (via SHAP), and clinically scalable.

However, several limitations warrant caution. While GAN-based augmentation improved severe-class classification, it introduced subtle distributional shifts that reduced KNN-PCA accuracy, underscoring the risk of overfitting to synthetic patterns. Although histogram-based validation confirmed strong similarity between real and generated samples, further external validation on independent datasets is needed to establish generalizability. Additionally, SHAP-based explainability was primarily visual and qualitative; future work should include quantitative stability analyses across folds. The models were not tested on cases with comorbid conditions or atypical EEG profiles, which may affect clinical applicability, and computational resource requirements were not assessed.

Despite these limitations, our results suggest that combining deep learning models such as MLP, CNN, and ANFIS with explainability tools like SHAP can support accurate and interpretable EEG-based depression assessment. The demonstrated 15% performance gain for severe depression after GAN augmentation highlights its potential for addressing class imbalance, though broader multi-center validation is needed before clinical deployment. Future work should also integrate confidence intervals, resource efficiency analyses, and class-specific SHAP attribution (Fig 4 to 6) to enhance reliability, practicality, and clinical trust. Overall, this framework represents a robust step toward scalable and explainable EEG-based depression diagnostics.

V. Conclusion

The primary aim of this study was to develop and evaluate an EEG-based framework using machine learning and deep learning models for accurate detection of depression and classification of its severity levels. Across clinical and external datasets (MODMA, PRED+CT, and OpenNeuro), the Multilayer Perceptron (MLP) emerged as the most reliable model, achieving the highest overall accuracy of 98.7% and 100% precision, consistently outperforming traditional approaches. Deep learning models, including Convolutional Neural Network (CNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS), also demonstrated strong performance, accurately classifying depression into Mild, Moderate, and Severe stages with accuracies exceeding 95%. Importantly, GAN-based data augmentation contributed to an improvement of approximately 15% in classification performance for Severe depression, addressing class imbalance and enhancing model robustness.

These findings highlight the potential of the proposed framework to serve as a practically realizable and explainable tool for early and reliable depression diagnosis in clinical environments by combining high-performance models with interpretability (SHAP) and data augmentation strategies.

For future work, we plan to validate this framework on larger and more diverse independent EEG datasets to assess cross-dataset generalizability, evaluate its performance in patients with comorbid psychiatric or neurological conditions, and analyse its robustness to atypical EEG patterns. Further work will also focus on quantifying model explainability, incorporating confidence estimation, and optimizing computational efficiency to enhance clinical adoption.

Acknowledgment

The authors would like to express sincere gratitude to the Department of Computer Engineering, Thakur College of Engineering and Technology, for the invaluable support and resources provided throughout this research. The facilities, academic environment, and encouragement from faculty members have significantly contributed to the completion of this work. This study would not have been possible without the institution's commitment to advancing research and innovation in the field of brain computer interface.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

No datasets were generated during the current study. Dataset analyzed in this study is obtained from EEG Psychiatric Disorders Dataset | Kaggle

Author Contribution

All authors contributed equally to the conception and design of the study. All authors did preparation of materials, data acquisition and analysis. All authors of this manuscript wrote and then revised on previous versions of the manuscript. All the authors read and approved the final manuscript.

Declarations

Ethical Approval

All procedures adhered to ethical guidelines for research involving human subjects.

Consent for Publication Participants.

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

All participants gave consent for publication.

Competing Interests

The authors declare no competing interests.

References

[1] S. A. Fields, J. Schueler, K. M. Arthur, and B. Harris, "The Role of Impulsivity in Major Depression: A Systematic Review", doi: 10.1007/s40473-021-00231-y/Published.

[2] Z. Li, M. Ruan, J. Chen, and Y. Fang, "Major Depressive Disorder: Advances in Neuroscience Research and Translational Applications," Jun. 01, 2021, *Springer*. doi: 10.1007/s12264-021-00638-3.

[3] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9 Validity of a Brief Depression Severity Measure."

[4] P. Riera-Serra *et al.*, "Clinical predictors of suicidal ideation, suicide attempts and suicide death in depressive disorder: a systematic review and meta-analysis," *Eur Arch Psychiatry Clin Neurosci*, Oct. 2023, doi: 10.1007/s00406-023-01716-5.

[5] P. W. Corrigan, B. G. Druss, and D. A. Perlick, "The impact of mental illness stigma on seeking and participating in mental health care," *Psychological Science in the Public Interest, Supplement*, vol. 15, no. 2, pp. 37–70, Jan. 2014, doi: 10.1177/1529100614531398.

[6] S. Olbrich and M. Arns, "EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response," *International Review of Psychiatry*, vol. 25, no. 5, pp. 604–618, 2013, doi: 10.3109/09540261.2013.816269.

[7] B. Ay *et al.*, "Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals," *J Med Syst*, vol. 43, no. 7, Jul. 2019, doi: 10.1007/s10916-019-1345-y.

[8] G. M. Bairy, U. C. Niranjan, and S. D. Puthankattil, "AUTOMATED CLASSIFICATION OF DEPRESSION EEG SIGNALS USING WAVELET ENTROPIES AND ENERGIES," *J Mech Med Biol*, vol. 16, no. 3, May 2016, doi: 10.1142/S0219519416500354.

[9] *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2019.

[10] S. Khan *et al.*, "A machine learning based depression screening framework using temporal domain features of the electroencephalography signals," *PLoS One*, vol. 19, no. 3 MARCH, Mar. 2024, doi: 10.1371/journal.pone.0299127.

[11] *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. IEEE, 2019.

[12] S. Soni, A. Seal, A. Yazidi, and O. Krejcar, "Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression," *Comput Biol Med*, vol. 145, p. 105420, 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105420.

[13] A. Sam, R. Boostani, S. Hashempour, M. Taghavi, and S. Sanei, "Depression Identification Using EEG Signals via a Hybrid of LSTM and Spiking Neural Networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4725–4737, 2023, doi: 10.1109/TNSRE.2023.3336467.

[14] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG," *IEEE Trans Instrum Meas*, vol. 70, 2021, doi: 10.1109/TIM.2021.3053999.

[15] Y. L. Neo, "A Review of State-of-the-Art Machine Algorithms on Classifying Depressive Disorder EEG Signals," in *2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2024, pp. 365–371. doi: 10.1109/ICAICA63239.2024.10822999.

[16] D. Jan, M. de Vega, J. López-Pigüi, and I. Padrón, "Applying Deep Learning on a Few EEG Electrodes during Resting State Reveals Depressive States: A Data Driven Study," *Brain Sci*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/brainsci12111506.

[17] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Comput Methods Programs Biomed*, vol. 161, pp. 103–113, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.012.

[18] D. Dhabliya, K. S. Bhuvaneshwari, P. Kapoor, S. Sowmiya, A. Garg, and L. Yashoda, "Hybrid ANFIS Model For Depression Recognition Consuming Three-Channel EEG Data," in *2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, 2024, pp. 1–5. doi: 10.1109/ICRASET63057.2024.10895792.

[19] U. R. Acharya *et al.*, "A novel depression diagnosis index using nonlinear features in EEG signals," *Eur Neurol*, vol. 74, no. 1–2, pp. 79–83, Dec. 2015, doi: 10.1159/000438457.

[20] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

connectivity, and nonlinear analysis," *J Neurosci Methods*, vol. 358, Jul. 2021, doi: 10.1016/j.jneumeth.2021.109209.

[21] S. Dhekane and A. Khandare, "Applied Machine Learning in EEG data Classification to Classify Major Depressive Disorder by Critical Channels," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 3, pp. 581–596, Jul. 2025, doi: 10.35882/jeeemi.v7i3.719.

[22] J. Shen, J. Chen, Y. Ma, Z. Cao, Y. Zhang, and B. Hu, "Explainable Depression Recognition from EEG Signals via Graph Convolutional Network," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 1406–1412. doi: 10.1109/BIBM58861.2023.10386011.

[23] L. Kumar, K. Kaustubh, T. Rajkhowa, and S. A. Mattur, "EEG-ViLSTM: A Deep Learning Approach for Depression Detection Using EEG Signals," in *2025 National Conference on Communications (NCC)*, 2025, pp. 1–6. doi: 10.1109/NCC63735.2025.10983346.

[24] W. Yuan *et al.*, "Discovery of Shared Latent Nonlinear Effective Connectivity for EEG-Based Depression Detection," *IEEE Trans Neural Netw Learn Syst*, vol. 36, no. 6, pp. 10663–10677, 2025, doi: 10.1109/TNNLS.2024.3514182.

[25] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach," *Cogn. Neurodyn.*, vol. 15, no. 2, pp. 239–252, Apr. 2021.

[26] S. Mahato, S. Paul, N. Goyal, S. N. Mohanty, and S. Jain, "3EDANFIS: Three channel EEG-based depression detection technique with Hybrid Adaptive Neuro Fuzzy Inference System," *Recent Pat. Eng.*, vol. 17, no. 6, Nov. 2023.

[27] M. Rehman, S. M. Umar Saeed, S. Khan, S. H. Noorani, and U. Rauf, "EEG-Based Depression Detection: A Temporal Domain Feature-Centric Machine Learning Approach," in *2023 International Conference on Frontiers of Information Technology (FIT)*, 2023, pp. 208–213. doi: 10.1109/FIT60620.2023.00046.

[28] S. M. Park *et al.*, "Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach," *Front Psychiatry*, vol. 12, Aug. 2021, doi: 10.3389/fpsyt.2021.707581.

[29] Zhang, Z., Yang, J., Xiong, P., Hao, H., Zhang, J., Li, L., Wang, C., & Liu, X. (2025). A cross-subject MDD detection approach based on multiscale nonlinear analysis in resting state EEG. *Neuroscience*, *582*, 1–10.

[30] Tigga, N. P., & Garg, S. (2022). Efficacy of novel attention-based gated recurrent units transformer for depression detection using electroencephalogram signals. Health information science and systems, 11(1), 1. https://doi.org/10.1007/s13755-022-00205-8

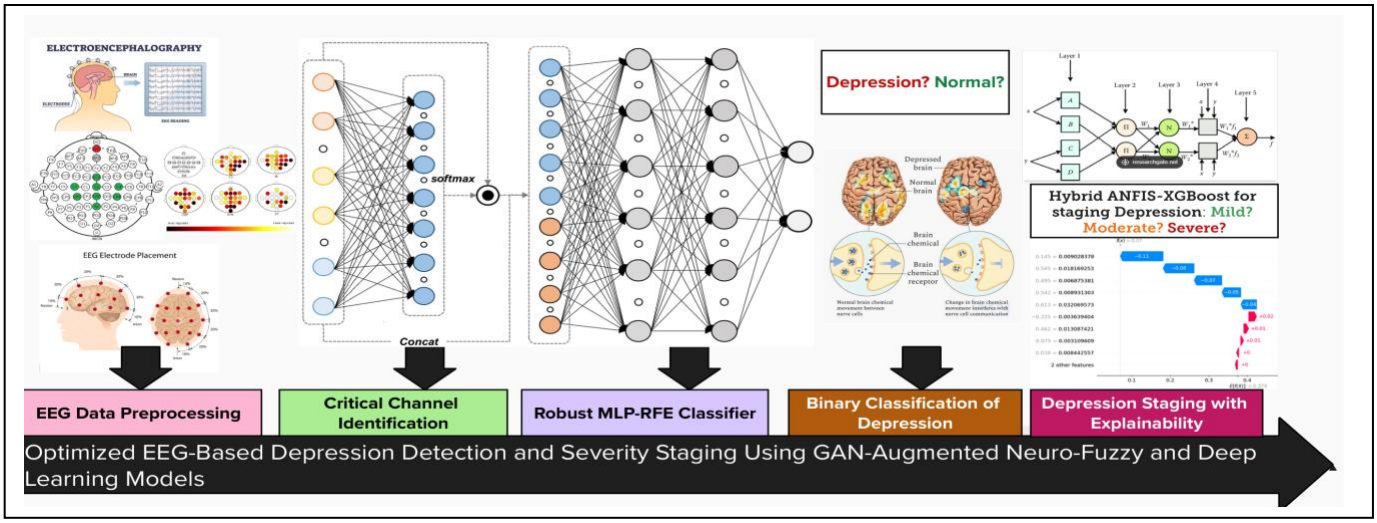
Author Biography

SUDHIR DHEKANE received B.E in Computer Science and Engineering from Shivaji University Kolhapur Maharashtra, M.E. degree in Computer Engineering from University of Mumbai in India. He is currently working as an Assistant Professor in Artificial Intelligence (AI) & Data Science at Dwarkadas J. Snghvi College of Engineering affiliated to the University of Mumbai. His current research interest includes Machine Learning and Deep Learning, Artificial Intelligence, algorithms, Programming, Web Development etc. and Brain Computer Interface. He has published number of papers in the international journal and Conferences as a part of professional development. Further he has reviewed journals and articles for various conference.

Dr. ANAND KHANDARE has received M.E. degree in Computer Engineering and Ph.D in CSE. He is currently working as Professor & Associate Dean (Planning & Operations-Digital Resources) at Thakur College of Engineering & Technology, Mumbai. His current research interest includes Machine Learning and Deep Learning, Database management, natural language Processing and Data Warehouse and Brain Computer Interface. He has received Got Silver category award for Infosys Campus Connect Program. He has been felicitated by Bhaktivedanta Hospital for successfully launching Application. He has earned funding for many projects and successfully completed the same. He has published number of papers in the international journal and Conferences as a part of professional development. Further, he has reviewed journals and articles for various conference.

Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 1112-1129 e-ISSN: 2656-8632



Manuscript Received 20 July 2025; Revised 10 September 2025; Accepted 20 September 2025; Available online 29 September 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1107