Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhanced Ensemble Machine Learning Technique to detect Bipolar Disorder

Lingeswari Sivagnanam [1*], N. Karthikeyani Visalakshi [2]

[1] *Research Scholar, Computer Science, Government Arts and Science College, Kangeyam, India. Email: tejumitu1279@gmail.com*

[2] *Assistant Professor, Computer Science, Government Arts and Science College, Kangeyam, India. Email: karthichitru@yahoo.co.in*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

The study introduced a unique Enhanced Ensemble Machine Learning (EEML) method for the detection of bipolar disorder, a serious medical diagnoses problem that acquires an immediate and timely diagnosis for an effective and earlier treatment. The EEML approach combines several machine learning models with the data based on patient's attitude on different perspectives such as real, signal based, textual and behavioural data and more. In order to train individual classifiers for identifying mental diseases, pertinent characteristics are taken from each data source using a thorough feature selection technique. Relevance Vector Machine, Adaboost, Multi-Layer Perceptron and Recurrent Neural Network are among the classifiers; Random forest optimizes the classifiers. By capturing a greater variety of traits associated with mental diseases, the ensemble of classifiers improves overall performance. Three bipolar disorder datasets of various are used in the study to assess the EEML method on Dataset1 (a multimodal dataset), Dataset2 (signals on sensor-based dataset), and Dataset3 (real-time dataset). The EEML model excels with its higher accuracy of about 96.28% with Dataset1, 95.71% with Dataset2, and 98.5% with Dataset3. This study enhances the field of mental health diagnostics by taking advantage of ensemble machine learning's strengths to enhance detection accuracy and reliability.

Keywords: Bipolar disorder, EEML, RVM, RNN, classifiers.

INTRODUCTION

Bipolar disorder (BD) is considered as a mental illness (MD) that is characterized by repeated spells of overly depression and elevated mood, as well as fluctuations in their activity [1], [3]. Bipolar disorder is one of the major mental illness where people suffered of this throughout the world. Stress, depression and loneliness plays an important role in serious development of bipolar disorder. This research, detects the patients who are affected precipitately of bipolar disorder. People who have BD exhibits high impulsivity and activity as well as poor focus [2]. However, these characteristics are episodic, which means that they tend to arise and alter at unpredictable intervals. This makes BD a difficult mental diseases world to diagnose [4].

The primary aim of this study is to create an Enhanced Ensemble Machine Learning classifier (EEML) that accurately and cost effectively detects bipolar disorders by utilizing various types of machine learning models, of which have been shown to produce more accurate findings over time. The main objective of the study is to determine a person's mental state affected of bipolar disorder by developing system models, to help the needy at the earliest. In order to build these systems, using different types of machine learning modals and data sets acquired from various sources. Bipolar disorder is considered as a lifelong illness but it is long-term, ongoing treatment can help people to control symptoms and enable to live a healthy life. Patients with early-stage bipolar disorder have substantially better odds of

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

recovery and long-term well-being. Hence by these kind of research cannot stop it from happening, but can be useful for early detection. So that the needy people get the right psychiatric care at the earliest.

The significant contributions of this paper are as follows:

- Have proposed an Enhanced Ensemble Machine Learning (EEML) model by combining Relevance Vector Machine, Adaboost, Multi-Layer Perceptron and Recurrent Neural Network for earlier and effective diagnosis of bipolar disorder.
- To improve accuracy in detection of bipolar disorder by means of attribute selection measure with its relevant features.
- Here, three different datasets from different sources have been applied in the detection using the EEML modal. The modal applies to different datasets which involves Dataset1 as signal based sensor devised dataset, Dataset2 as multimodal dataset and Dataset3 as real-time dataset.
- The individual base classifiers and the EEML model performance is evaluated with its performance metrics such as Accuracy, Recall, Precision, F1 score and Cohen kappa score.

LITERATURE REVIEW

The proposed method of W Jiji., 2022 [5] supervised learning framework focused on detecting bipolar disorder (BD), shows that this method has the potential to identify BDs with an average accuracy of 77.77%. R. Achalia, et al., 2020 [6] applies SVM, a supervised machine learning method, with accuracy of 87.60%, a sensitivity of 82.3%, and a specificity of 92.7%. LS ROTENBERG., 2021[7] studies about the ML algorithms (Support Vector Machines, Random Forest, Naive Bayes, and Multilayer Perceptron) showed reasonable performance on the prediction task, with F values ranging from 61 to 80%. The random forest algorithm yielded higher average performance. (68% relapsed vs. 74% non-relapsed). Ensemble approaches help reduce variance and bias mistakes associated with individual ML models. For example, bagging reduces variance without raising bias, whereas boosting decreases bias [8], [9], [10].

A study by M. B. Fonseca et al., 2018[11] found 92% accuracy when comparing BD patients with healthy individuals and 93% accuracy when comparing SZ with healthy individuals. These 18 findings distinguish patients with bipolar disorder from those with schizophrenia with her 92% accuracy. The Peerbasha's [12], a sequential deep learning model was used to predict a patient with bipolar disorder and was found to be 99% accurate.

An ensemble classifier model is proposed by author L. Sivagnanam and N.K. Visalakshi [13] with attaining accuracy of 98%. This high accuracy rate shows that an effectiveness of ensemble method on improving the BD detection. In order to detect two various mental illness such as anorexia and depression, Rao et al. [14] proposed a model which is extremely effective. This model make a promising nature for medical diagnosis in a wide range based on its robustness in real time cases. Fitriyani et al. [15] used the patient's risk variables data to create an ensemble model for the early diagnosis of type2 diabetes as well as hypertension.

DATASET DESCRIPTION

In data collection, patient records are collected from various sources such as smart devices, medical consultations and Electronic Health Records (EHRs). Medical history, mood swings, activity range, demographic data, sleep pattern, medication compliance, and speech mode are needed for this system to assess the patient.

Dataset-1

Dataset 1 contains the Wearable Stress and Affect Detection (WESAD) dataset. These datasets are available publicly for detecting stress through wrist-worn devices. This dataset includes physiological signals recorded under various emotional states such as amusement, stress and neutral. For mental health research, it provides preprocessed device-monitored data to identify patients' normal and abnormal situations on basis of their activity called actography.

Dataset-2

Dataset 2 contains psychotherapy interview transcriptions processed through the Google Cloud platform. The study uses two different datasets including Extended Distress Analysis Interview Corpus (E-DAIC) and Bipolar Disorder Corpus (BDC) to analyses the depression and bipolar disorders that enables researchers to examine their systems on standard datasets, which helps emotion detection and related areas move forward. The dataset consists of synchronized audio and video recordings and recorded sessions with a range of emotional states identified to represent emotional conditions or actions.

Dataset-3

The perusal records of data collected from the concerned department with maximum number of samples from the recent data. The data is retrived from the patients' data sheet. It contains both patients' history as well as clinical details. The retrieved data have been stored in excel sheet in .CSV format. The features such as age, gender, sleepiness, aggressive, depression etc.. arebeen collected from the clinical sheet.

DATA PREPROCESSING

Data preprocessing includes the following steps: data cleaning, data assembly, and data encoding. The lost data is operated using the imputation approach. By establishing a limit for the arithmetic values, the attributes' consistency is preserved. To translate the classified data into arithmetic values, this system uses either label encoding or one-hot encoding. Additionally, to maximize data quality, noise reduction techniques like Principal Component Analysis (PCA) or Independent Component Analysis (ICA) are used.

Data Splitting and Feature Extraction

The preprocessed data is divided into training and testing datasets in an 80-20 or 70-30 ratio to guarantee that the model can be tested with unknown data. K-fold cross-validation is one kind of cross-validation approach that is used to minimize overfitting and maximize the resilience of the model. Through feature extraction, the model performs at its best and becomes less complex. To extract characteristics from heart rate deviation, activity range, and sleep pattern, time-series analysis is utilized. It is possible to derive statistical parameters like mean, standard deviation, skewness, and kurtosis from the mood and activity levels. Social behavior, daily routine, and mood rating patterns are examined in relation to behavioral feature extraction. In addition to subject simulation and emotional analysis, the natural language processing (NLP) technique extracts features from the medical history and consultation notes.

Training data

Selecting and training a collection of various machine learning techniques is the initial step in the training module. Adaboost, Relevant Vector Machine (RVM), RNN and Multilayer Perceptron (MLP) models are all used in this system. Each model captures the different data patterns to maximize the overall stability of the ensemble. Using models such as grid search or random search, hyperparameter tuning is used to attain the best model performance.

Testing data

The trained models are examined in the testing sets to assess the accuracy, F-measure, recall, and precision. This process ensures that the model can properly handle new and unknown data. Highlight the areas that need improvement, and provide specific details about the model's effectiveness based on performance metrics.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

PROPOSED METHODOLOGY

Bipolar disorder is an important mental illness that causes daily emotional tension and mood fluctuations. The early detection and effective treatment of this condition are critical. An ensemble heterogeneous classifier is presented in this study to accurately identify bipolar disorder. To obtain the best possible prediction accuracy and durability, this system combines a number of machine-learning approaches.

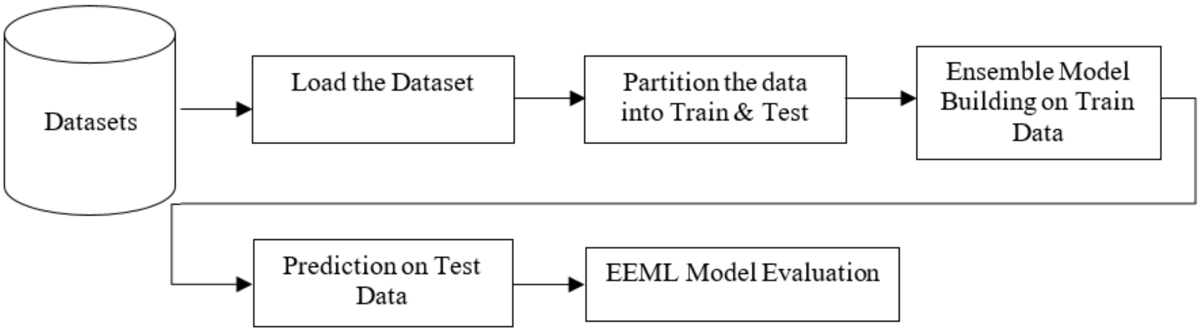

Figure 1. EEML architecture

Figure 1 shows the proposed EEML machine learning framework as architecture representation of the contribution applied in this work.

EEML MODEL

EEML stands for Enhanced Ensemble Machine Learning Classifier. This model has the ability to combine the individual models into an ensemble classification model to improve the overall performance in order to get better detection and prediction of the expected output. As the multiple models are combined together to strengthens the EEML. Hence reproducing the detections with better accuracy because of its robustness [11]. Thus optimizes the performance on the testing set. The below is the building model of EEML,

EEML Algorithm

Procedure

Step 1: Splitting the data by dividing the dataset into training and testing set.

Step 2: Each and every individual model to be trained using training data.

- Fitting the Adaboost model on the training data.
- Fitting the RVM (Relevance Vector Machine) model on the training data.
- Fitting the MLP (Multilayer Perceptron) model on the training data.
- Fitting the RNN (Recurrent Neural Network) model on the training data.

Step 3: Generating detections on the datasets:

- Using the trained models to generate detections on the testing data.

Step 4: Combining these detections using random forest:

Step 5: Evaluating the performance of the ensemble by applying performance metrics.

- Measuring the accuracy, precision, recall, or F1-score and Kappa Score of the ensemble on the testing data.

Step 6: Final result of the detection on Enhanced Ensemble classifier.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

RESULTS AND DISCUSSION

This Model shows the analysis of the individual base classifiers of Adaboost Classifier, Multilayer Perceptron Classifier, RNN and Relevance Vector Machine learning Classifier. The combination or the integration these algorithms is its EEML model to find the results which deals with the performance metrics such as Accuracy, Precision, Recall, F1 Score and Cohen Kappa Score.

Accuracy

Accuracy refers to the ratio of correctly classified instances out of the sum of instances in the dataset. It is calculated as follows

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$ Equation 1

Precision

Precision refers to the ratio of true positive predictions out of sum of positive predictions. Precision is calculated as follows

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$ Equation 2

Recall

Recall also known as sensitivity, refers to the proportion of true positive predictions out of sum of actual positive instances in the dataset. Recall is calculated as follows

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$ Equation 3

F1 Score

The F1 score is considered as the harmonic mean of precision and recall. This provides a balance between precision and recall that can be applied when the class distribution is imbalanced. The F1 score is calculated as follows

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$ Equation 4

Cohen Kappa Score

Cohen's kappa score can also be referred to as Cohen's kappa or kappa coefficient, it's a statistic score to measure the level of agreement between two different raters to classify items into mutually exclusive categories, very common in machine learning techniques. This ranges from -1 to 1, where 1 indicates perfect agreement between the raters and 0 indicates agreement equivalent to chance whereas Negative values indicate agreement worse than chance. The formula for Cohen's kappa is as follows

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$ Equation 5

Where Po is the observed agreement between the raters.

Pe is the expected agreement between the raters if the agreement occurred by chance.

The above figures depicts the confusion matrix of the classifiers applied in this paper, thus shows the clear view of the work.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The tables and graphs are generated depending on the performance values attained to assure the effectiveness of the proposed EEML. The following are the tables involved in the implementation of three different datasets used and results are analysed with respect to the algorithms.

Table 1: Performance metrics of individual classifiers on Dataset1

Classification Model	Accuracy	Precision	Recall	F1_score	Cohen Kappa Score
Adaboost Classifier (ABC)	96.84	96.96	96.84	96.78	95.58
Multilayer Perceptron Classifier (MLP)	80.02	78.88	80.02	79.51	71.34
Relevance Vector Machine Learning Classifier (RVM)	93.04	92.99	93.00	92.89	90.95
Recurrent Neural Network (RNN)	95.63	95.01	94.27	94.99	95.12
Enhanced Ensemble Machine Learning Model (EEML)	96.28	99.37	99.29	99.27	98.99

The table1 using Dataset1 shows the results of the individual classifiers with their performance metrics. The accuracy of Adaboost Classifier (ABC) shows 96.84%, Multi-Layer Perceptron (MLP) is about 80.02%, Relevant Vector Machine Learning Classifier (RVM) has about 93.04%, and Recurrent Neural Network produces 96.53%. Finally Enhanced Ensemble Machine Learning Model (EEML) shows the result with the highest accuracy of about 96.28% compared with the individual classifiers.

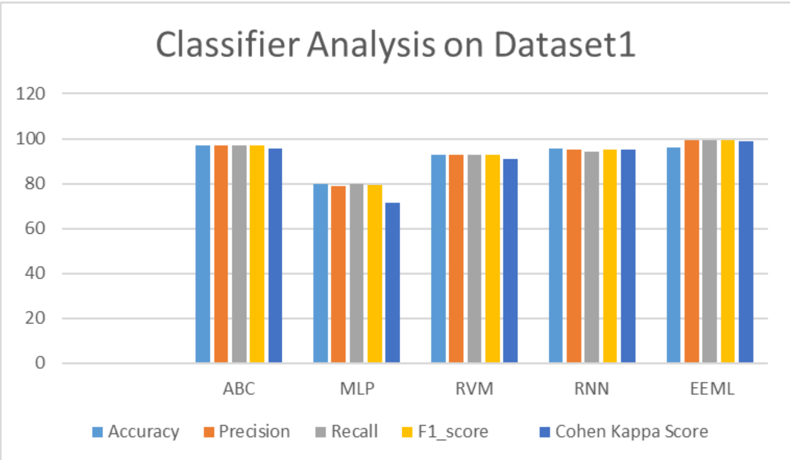

Figure. 2. Analysis Graph on Dataset1

The above figure 6.2. Illustrates the visual representation of results of the individual classifiers with its Accuracy, Precision, Recall, F1 score and Cohen Kappa score on Dataset1.

Table 2: Performance metrics of individual classifiers on Dataset2

Model name	Accuracy	Precision	Recall	F1_score	Cohen Kappa Score
Adaboost Classifier (ABC)	91.1.	91.09	91.47	90.86	91.36
Multilayer Perceptron Classifier (MLP)	95	95.26	95.14	95.02	94.60

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Relevance Vector Machine Learning Classifier (RVC)	88.92	90.62	90.11	89.56	88.06
Recurrent Neural Network (RNN)	93.84	93.21	94.62	93.02	93.78
Enhanced Ensemble Machine Learning Model (EEML)	95.71	96.03	95.81	95.85	95.37

The table 6.2. using Dataet2 shows the results of the individual classifiers with their performance metrics. The accuracy of Adaboost Classifier (ABC) shows 91.1%, Multi-Layer Perceptron (MLP) is about 95%, and Relevance Vector Machine Learning Classifier (RVC) has about 88.92%, Recurrent Neural Network (RNN) results in 93.84%. Finally Enhanced Ensemble Machine Learning Model (HEML) shows the result with the highest accuracy of about 95.71% compared with the individual classifiers.

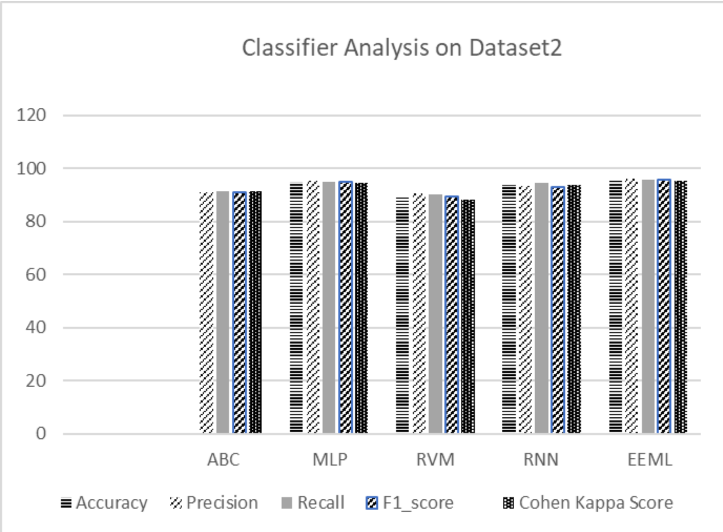

Figure 3. Analysis Graph on Dataset2

The above figure 4 illustrates the visual representation of results of the individual classifiers with its Accuracy, Precision, Recall, F1 score and Cohen Kappa score on Dataset2.

The table 3 shows the results of the individual classifiers with their performance metrics. As it is a perfectly balanced dataset, the classes are equally represented, hence accuracy and recall

might be the same as long as the provided models correctly rectify all instances of the positive class (true positives) meanwhile also identifying all instances of the negative class (true negatives). The accuracy of Adaboost Classifier (ABC) shows 95%, Multi-Layer Perceptron (MLP) is about 95%, Relevance Vector Machine Learning Classifier (RVC) has about 96.5%. Finally Enhanced Ensemble Machine Learning Model (EEML) shows the result with the highest accuracy of about 98.5% compared with the individual classifiers.

Table 3: Performance metrics of individual classifiers on Dataset3

Model name	**Accuracy**	**Precision**	**Recall**	**F1_score**	**Cohen Kappa Score**
Adaboost Classifier (ABC)	95	95.14	95	95.01	89.90
Multilayer Perceptron Classifier (MLP)	95	95.51	95	95.01	89.96

Relevance Vector Machine Learning Classifier (RVC)	96.5	96.36	96.45	96.49	93.85
Recurrent Neural Network (RNN)	97.1.	96.84	97.03	97	96.52
Enhanced Ensemble Machine Learning Model (EEML)	98.5	100	97.77	98.87	97.46

The below figure 4 illustrates the visual representation of results of the individual classifiers with its Accuracy, Precision, Recall, F1 score and Cohen Kappa score on Dataset3

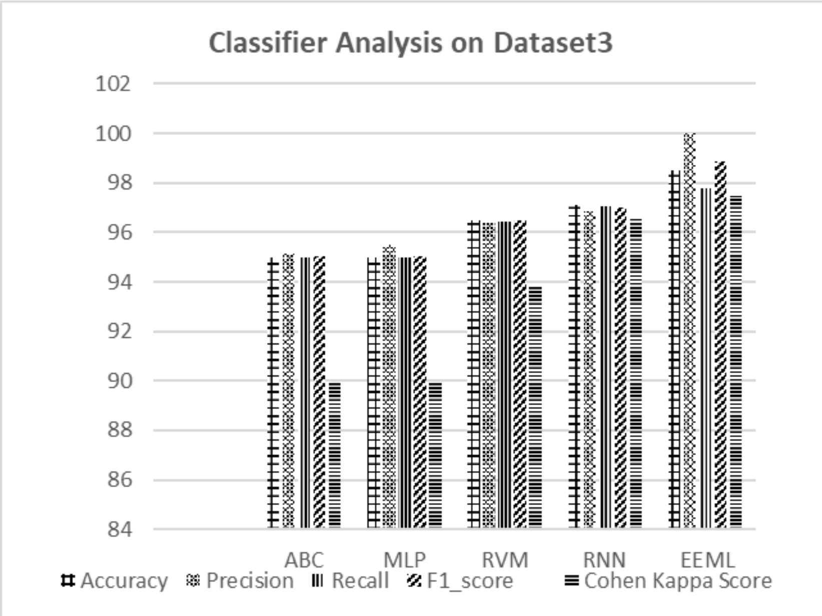

Figure 4. Analysis Graph on Dataset3

Table 4: Different Data Sets Compared with the Proposed Classifier

Datasets	**Proposed Accuracy**	**Classifier**
Dataset 1	EEML	96.28%
Dataset 2	EEML	95.71%
Dataset 3	EEML	98.5%

Table 4 describes about an accuracy of proposed EEML classifier is obtained for three various data sets is represented. Hence, the outstanding performance of classifier is demonstrated from this outcome. Also it can handle the multiple datasets effectively and offer persistent accuracy in all circumstances.

CONCLUSION

The Ensemble model classifier has been adopted by combining Adaboost Classifier, Multilayer Perceptron Classifier, Relevance Vector Machine learning Classifier, RNN and the model has been applied on the data set with complete features. The performance of ensemble model is measured by its evaluation parameters. When the results are compared, the Enhanced ensemble model provides a best accuracy of 96.28% in analysis with dataset1, 95.71% in analysis with dataset2 and 98.5% in analysis with dataset3 than the individual models. Thus the novelty of this research EEML, is one of the better classifier in the real world scenarios, can be a powerful building model in the detection mental health issues.

This study can provide findings regarding ML as a feasible and accurate detector for patients suffering from bipolar disorder but at limited range of samples and that strongly encourages, future studies is to compare the findings with larger patient cohorts, as well as include a larger sample of variables to ML findings. Furthermore, this kind of machine learning classifier can be implemented in other sorts of mental and physical illnesses for better and accurate results a short time.

Acknowledgement: I thank Dr. Karthikeyani Visalakshi for her contributions to this work.

Funding Statement: No financing / There is no fund received for this article.

Data Availability: Data sharing is not applicable to this article”.

Conflict of interest: No conflict of interest.

REFERENCES

[1] Müller-Oerlinghausen, B.; Berghöfer, A.; Bauer, M. Bipolar Disorder. Lancet 2002, 359, 241–247.
[2] Hajek, T.; Cullis, J.; Novak, T.; Kopecek, M.; Blagdon, R.; Propper, L.; Stopkova, P.; Duffy, A.; Hoschl, C.; Uher, R.; et al. Brain Structural Signature of Familial Predisposition for Bipolar Disorder: Replicable Evidence for Involvement of the Right Inferior Frontal Gyrus. Biol. Psychiatry 2013, 73, 144–152.
[3] Dwyer, D.B.; Falkai, P.; Koutsouleris, N. Machine Learning Approaches for Clinical Psychology and Psychiatry. Ann. Rev. Clin. Psychol. 2018, 14, 91–118.
[4] Arnone, D.; Cavanagh, J.; Gerber, D.; Lawrie, S.M.; Ebmeier, K.P.; McIntosh, A.M. Magnetic Resonance Imaging Studies in Bipolar Disorder and Schizophrenia: Meta-Analysis. Br. J. Psychiatry 2009, 195, 194–201.
[5] Ganasigamony, Wiselin Jiji, and Muthuraj Antony Arul Selvaraj. "Computer assisted diagnosis of bipolar disorder using invariant features." Concurrency and Computation: Practice and Experience (2022): e6984
[6] Achalia R, Sinha A, Jacob A, Achalia G, Kaginalkar V, Venkatasubramanian G, Rao NP. A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder. Asian J Psychiatr. 2020 Apr; 50:101984. https://doi: 10.1016/j.ajp.2020.
101984. Epub 2020 Feb 26. PMID: 32143176.
[7] Rotenberg LS, Borges-Júnior RG, Lafer B, Salvini R, Dias RDS. Exploring machine learning to predict depressive relapses of bipolar disorder patients. J Affect Disord. 2021 Dec 1; 295:681-687. doi: 10.1016/ j.jad.2021.08.127. Epub 2021 Sep 2. PMID: 34509784.
[8] S. Doroudi, ``The bias-variance tradeoff: How data science can inform educational debates,'' AERA Open, vol. 6, no. 4, Oct. 2020, Art. no. 2332858420977208, doi: 10.1177/2332858420977208.
[9] L. Breiman, ``Arcing classi_ers,'' Ann. Stat., vol. 26, no. 3, pp. 801_824, 1998.
[10] S. Alelyani, ``Stable bagging feature selection on medical data,'' J. Big Data, vol. 8, no. 1, p. 11, Jan. 2021, doi: 10.1186/s40537-020-00385-8.
[11] Fonseca, M., Andrades, R., Bach, S., Wiener, C. and Oses, J. (2018) Bipolar and Schizophrenia Disorders Diagnosis Using Artificial Neural Network. Neuroscience and Medicine, 9, 209-220. Doi: 10.4236/nm.2018.
94021.
[12] A Predictive Model to identify possible affected Bipolar disorder students using Naive Baye's, Random Forest and RVM machine learning techniques of data mining and Building a Sequential Deep Learning Model using Keras. S. Peerbasha1* and M. Mohamed Surputheen2*(Research Scholar1*, Associate Professor2*) Vol. 21 No. 5 pp. 267-274.
[13] Sivagnanam, L., & Visalakshi, N. K. (2023). Detection of bipolar disorder by means of ensemble machine learning classifier. Data and Metadata, 2, 134-134. https://doi.org/10.56294/dm2023134.

[14] Rao, G., Peng, C., Zhang, L., Wang, X., Feng, Z. (2020). A Knowledge Enhanced Ensemble Learning Model for Mental Disorder Detection on Social Media. In: Li, G., Shen, H., Yuan, Y., Wang, X., Liu, H., Zhao, X. (eds) Knowledge Science, Engineering and Management. KSEM 2020. Lecture Notes in Computer Science (), vol. 12275. Springer, Cham. https://doi.org/10.1007/978-3-030-55393-7_17

[15] N. L. Fitriyani, M. Syafrudin, G. Al_an, and J. Rhee, ``Development of disease prediction model based on ensemble learning approach for diabetes and hypertension,'' IEEE Access, vol. 7, pp. 144777_144789, 32019, https://doi: 10.1109/ACCESS.2019.2945129.