VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2025, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 13, Number 3, July-September 2025 pp: 129-139

# Evaluation of Machine Learning–Based Methods to Detect Bipolar Disorder in Individuals With Mental Health Conditions

**Syed Ibad Hasnain** [1*], **Hafsa Israr** [2], **Muhammad Faris** [1], **Rabika Kamal** [1], **Hafiza Syeda Yusra Tirmizi** [1]

[1]Faculty of Engineering Science and Technology, Hamdard University, Karachi 74600, Pakistan ;
[2]Department of Biomedical Engineering, Sir Syed University of Engineering, Karachi, Pakistan

***Keywords***: *Mental Health, Random Forest, Support Vector Machine, Bipolar Disorder, Psychiatry.*

***Journal Info***:
*Submitted:*
*June 25, 2025*
*Accepted:*
*August 30, 2025*
*Published:*
*September 09, 2025*

## Abstract

Bipolar disorder (BD) is still one of the most incapacitating of neuroaffective disorders in psychiatry. The strong mood swings from states of euphoria to depression often destabilize interpersonal relationships and can undo occupational stability. Early and reliable diagnosis facilitates prompt pharmacological intervention and mental-health education that may protect not only the patient and their immediate social circle but also the entire social structure from general distress. In this research study the performance of machine learning algorithms such as random forest (RF), support vector machine (SVM) and gradient boosting (GB) has been investigated for classification and prognostication of BD and its subtypes. The machine learning models were validated using a clinical dataset, which included 120 participants: 28 of BD I, 31 of BD II, 31 of Major Depressive Disorder and 30 healthy controls. Model performance was evaluated with stratified cross-validated train-test-split and a set of metrics, including accuracy, precision, recall, F1-score, and Receiver Operating Characteristic - Area under the Curve (ROC vs. AUC). In other words, the RF model had the highest accuracy (88%), precision (90%), and recall (88%). The discriminative performance of RF and SVM models was comparable with an ROC-AUC of 97%. These results emphasize the potential of machine learning (ML), specifically ensemble techniques like Random Forest (RF), as an effective supplement to traditional early clinical diagnosis in bipolar disorders and related psychiatric illnesses.

***Correspondence author email address:** ibad.hasnain@hamdard.edu.pk
DOI: 10.21015/vtse.v13i3.2173

## 1 Introduction

Bipolar disorder (BD) is clinically defined by a pattern of alternating euthymia, mania, depression and sometimes mixed states; the profound mood fluctuations have brought BD to a level of public-health concern [1, 2] . Epidemiological surveys have suggested that ap-

proximately one-percent of the world's population is affected by the disorder with prevalence rates remaining constant across different cultural settings [3].

Formally, bipolar I is characterized by fully formed manic episodes while bipolar II is characterized by recurrent depressive episodes with hypomania of a lesser intensity [4]. Subtypes of patients show marked deficits in interpersonal relationships, occupational performance and fundamental executive functioning, and the risk of suicidality is elevated by an approximate twenty to thirty-fold compared to non-affected populations [5].

In clinical practise, BD is only exceptionally found in isolation; co-morbid expressions such as emotional dysregulation, depression and anxiety make diagnostic clarification and pharmacological therapy even more complicated [6]. Economic analyses suggest that direct medical costs are high, but that indirect costs (expressed as lost wages and reduction in production output) typically minimizes these direct costs [7]. With high prevalence, shortened life expectancy and substantial economic cost, research however remain on track looking for more effective algorithms of treatment, new pharmacotherapeutics, or even improvement in psychotherapeutic interventions [8, 9].

It continues to be evidenced that the etiology of bipolar disorder (BD) is multifactorial, meaning several potential pathways must be taken into account, with no one likely to be responsible for its presentation. Strong heritability is shown by genetic studies, but family history is only a first approximation. Recently, genome-wide association scans have added the loci CACNA1C and TRANK1 to the risk map, but these contribute only a small fraction of the global heritability [10, 11]. Neuroimaging studies by the ENIGMA BD consortium contribute a further layer of knowledge, reporting on atrophic changes and cross-associational features in the cortex and subcortical regions of the brain and in white matter tracts [12]. Biomarker studies have focused on the serum levels of insulin-like growth factor 2, a neurotrophic peptide which has been shown to be reduced correlatively with the severity of manic symptoms, and may serve as a clinical biomarker [13]. Finally, the role of environmental context needs to be taken into account; childhood abuse, chronic stress, recreational drug use often lead to latent vulnerability open [14].

Clinicians have previously diagnosed bipolar disorder according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) and cross-referencing the patient's report with a checklist of symptoms [8, 15]. Although theoretically sound, this balancing act is often time consuming, impressionist, and financially costly to low-resource clinics [5, 16]. Over the last decade, machine-learning methods have started to be used as an auxiliary diagnostic tool in psychiatry [17]. Machine-learning algorithms are used to analyze large carefully labelled datasets to detect hidden patterns and make predictions for new patients. The typical workflow begins with a complete training set that train the model, and then proceeds through a further, unseen hold-out set that objectively assesses the generalizability of the learned decision rules to real-world situations [18, 19].

Supervised machine-learning methods are now routine in clinical research, where domain experts annotate historical patient data and behavioral profiles to train algorithms that predict future presentations [20, 21]. Within the Machine Learning toolkit, the Random Forest algorithm has become a strength by accumulating the outputs of numerous decision trees, it dampens noise, resists over-fitting and remains stable when faced with incomplete records. Its ensemble structure easily engrosses variability arising from error prone screening questionnaires, densely documented intake notes and uneven demographic stratification, frequently achieving cross-validation performance that outstrips more conventional classifiers [22, 24]. Longitudinal studies increasingly reports Random Forest derived odds ratios alongside standard evaluation metrics, indicating growing confidence in the approach among busy clinical teams, reflecting the reliability of the model [25, 26].

Few studies have observed that the way a data set is split can radically bias the performance of a Random Forest [27, 28]. Cross-validation has emerged as a prevalent technique as it folds the training and testing portions so that every instance gets to be

evaluated in both parts [28]. By defying the algorithm with multiple, slightly different views of the data, this method limits overfitting and creates a robust scale of how much error the model is likely to encounter. The same cross-validation folds reveal whether the model's predictions remain reliable on data it has never seen. [29, 30]. Because of these advantages, the proposed study leans on cross-validation when tuning and checking the Random Forest and other machine learning models which includes Support Vector Machine and Gradient Boost.

## 2 Methodology

This section outlines the research framework employed in this study. Our approach was designed to ensure a systematic and reproducible analysis of the data. The entire process, depicted in Figure1.

### 2.1 Data Source

The dataset obtained for this study was from the Harvard Dataverse repository[31]. The data comprises of 120 patients and 17 essential symptoms psychiatrist use in diagnosing the mood disorders. These patients are divided into four diagnostic categories Bipolar Type I, Biploar Type II, Major Depressive Disorder and a control group labeled "Normal". Each patient is described by 17 behavioral attributes such as sadness, euphoric, exhausted, sleep disorder, mood swings, suicidal thoughts, anorexia, authority respect, try explanation, aggressive response, ignore & move on, nervous breakdown, admit mistakes, overthinking, sexual activity, concentration and optimism. The target variable is the expert's clinical diagnosis for each patient.

### 2.2 Data Pre-processing

To handle mixed data types, a comprehensive preprocessing pipeline was applied. Firstly, feature separation is performed segregating features into numeric and categorical subsets than using standard scalar transformer numeric features were transformed to have zero mean and unit variance which ensures all numeric features contribute equally.Secondly, categorical features using the One Hot Encoder were encoded into one-hot binary dummy variables. Lastly, a column transformer is used to combine these steps into a single unified pipeline. This preprocessing approach certifies that all transformations are learned from the training folds only and consistently applied to validation or test folds, ensuring no data leakage.

### 2.3 Machine Learning Models

Three different machine learning models were evaluated for this multi-classification task. The models were implemented with certain set of hyper parameters, chosen based on preliminary tuning.

#### 2.3.1 Random Forest (RF)

Random Forest (RF) algorithm is an ensemble learning technique that works through the creation of a large number of decision trees in the training. The main idea of the model is to create several decision trees, each of which is based on another bootstrap sample ( subset ) of the original data. The end result of the classification tasks is that the ultimate prediction is a majority vote of the predictions of all of the trees in the forest. This team methodology dramatically increases predictive strength and predictive power through minimizing variance and virtually eliminating overfitting that is often characteristic of individual decision trees. Figure2: provides a conceptual description of this ensemble structure.

The particular structure that the Random Forest model used in this research adopted was a group of 120 distinct decision trees. A maximum depth of four levels was placed in each tree to further control complexity of the model and overfitting. In order to achieve the highest level of reproducibility of the results a fixed random seed of 0 was explicitly defined and used all the way through the model building process.

These hyperparameters did not just happen to provide good results, but were a result of careful and trial-and error process of tuning. The preemptive grid search was first carried out to investigate a wide hyperparameter space. This was then followed by the cycles of narrower manual tuning. This was done by means of a lot of experimentation with numbers of estimators (trees) and with maximum tree depths. Each candidate configuration was strictly tested by

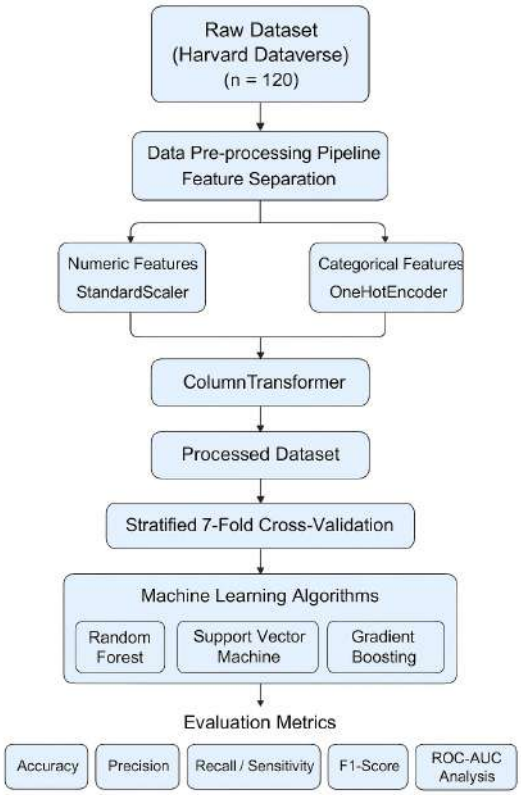


**Figure 1.** Methodology Workflow

stratified k-fold cross validation, a method that maintains the class distribution within a fold in order to give credible performance estimates.

The last selected configuration, consisting of 120 estimators at a maximum depth of 4, was chosen with a careful tradeoff between three important factors namely; high predictive accuracy, high computational efficiency and high resilience to model overfitting. This was a realization that provided the best trade-off and provided strong performance without redundancy.

### 2.3.2 Support Vector Machine (SVM)

Support Vector Model Support Vector machine (SVM) The Support Vector Model (SVM) framework revolves around finding the best hyperplane that maximally separates classes over a high dimensional feature space. This paper adopts the methodology suggested in this paper using RBF kernel SVM. The configuration of the model used is regularization parameter C= 2.33 and a kernel coefficient, gamma, that is specified as scale automatically calculates gamma= 1/n features X.var ).

These particular hyperparameter values were determined by an exhaustive optimization process. A preliminary grid search was carried out through a large array of possible C and gamma values to find potentially promising regions in the hyperparameter space. This initial exploration was accompanied by several successive rounds of manual these optimization steps to optimize the choice of parameters. Through this dual-phase method, the model stability was improved and uniform performance metrics among all folds of cross-validation were observed, thus allowing the final model to be more generalizable.

The full set of optimization steps, starting with grid search to the eventual choice of parameters is graphically represented in Figure3, which shows the correlation between the values of the hyperparameters with model performance in validation folds. The chosen pa-

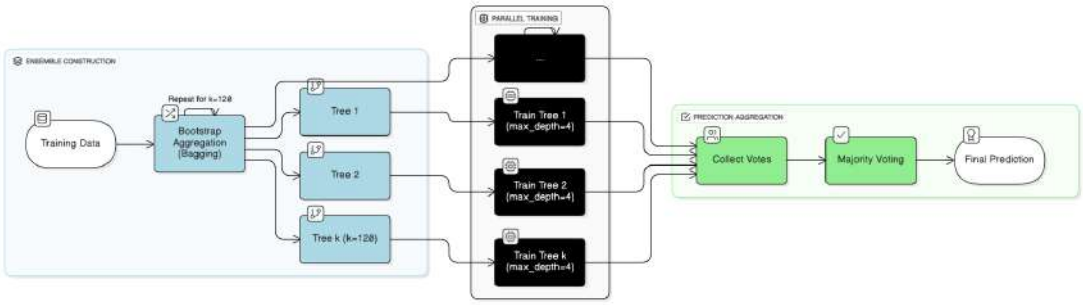


**Figure 2.** Random Forest Classifier (RF) Workflow

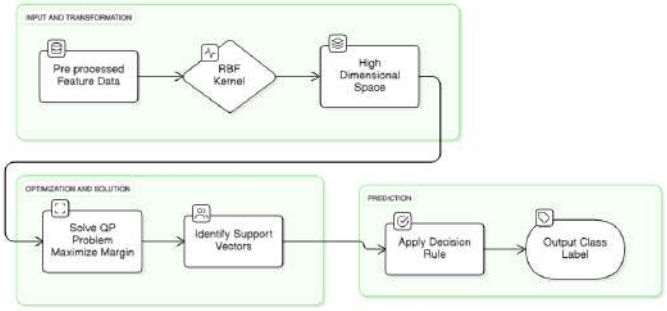


**Figure 3.** Support Vector Machine (SVM) Workflow

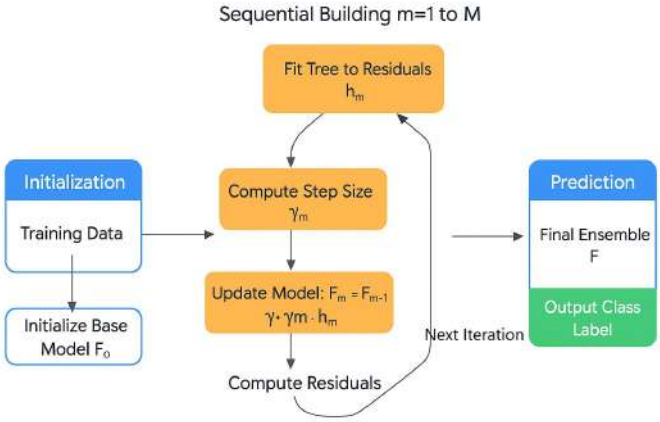


**Figure 4.** Gradient boosting (GB) Classifier Workflow

rameters (C=2.33, gamma='scale') are the set that ensured the best tradeoff between the complexity of the models and the generalization ability, maximizing margin separation without becoming overfit.

### 2.3.3 Gradient Boosting Classifier (GB)

Gradient Boosting approach builds the predictive model in a serial, cumulative manner of decision trees. The next tree in the series is fitted, in a particular way to the negative gradient, or the deviance, of the multiclass log loss function based on the ensemble of the preceding trees. This stage-by-stage method permits the model to reduce the loss function gradually, paying special attention to the errors in earlier iterations.

To conduct this study, the gradient boosting model is applied with the following settings: 100 estimator trees, a learning rate of 0.2, a tree depth of 4 and a subsample ratio of 1.0 (the full dataset is used during each tree). The architectural formulation of this model construction and the process of iterative fitting is graphically illustrated in Figure 4.

Gradient Boosting methodology builds on the predictive model in a series of sequential additive decisions in the form of decision trees. The successive trees that follow are explicitly trained on the negative gradient- or the deviance- of the multiclass log loss, to the ensemble of the prior trees. This step-by-step process allows this model to gradually reduce the loss function by targeting the errors that occurred in previous iterations.

To implement the gradient boosting algorithm in this study, the following parameters have been used: 100 estimator trees, learning rate of 0.2, maximum depth of the tree is 4 and the subsample ratio is 1.0 (uses the entire dataset to build the tree). Figure 4 displays the architectural form of this model configuration and the iterative fitting process visually.

These hyperparameters were selected as the

result of a strict optimization process. A preliminary automated grid search was used to find promising parameter sets, and the search was refined by manual tuning to stabilize the results. Many different configurations were tested using several criteria, such as the accuracy of the validation, the trade-off between the precision and the recall measures, and resistance to overfitting. The last parameter set was selected because it exhibited the most desirable and stable trade-off among model complexity, predictive performance and generalization capability over all the validation folds.

Each classifier was slotted into a single Scikit-learn pipeline that also handled the preliminary scaling and encoding. Cross-validation thus kept the fitting of those transforms strictly to the training folds, ensuring the untouched test fold remained a fair yardstick.

### 2.4 Performance Evaluation

A strong comparative analysis was made utilizing numerous performance matrix. The simplest of these is overall classification accuracy which measures the fraction of correct predictions on a test set held out of the computation. Stratified seven-fold cross-validations were performed to obtain a robust indication of model stability by splitting the data into evenly-sized mini-batches that maintained class balance, then accumulating the results over all folds. Error patterns were analyzed using a confusion matrix that included true positives, false positives, true negatives, and false negatives. A simplified classification report from scikit-learn summarized precision, recall, F1 Score and support, reflecting the number of correctly labelled, as well as the confidence in each label. Finally, receiver operating characteristic (ROC) curves offered an all-encompassing diagnostic view by plotting the true-positive versus false-positive rates trade-off; the area under the curve (AUC) for each class was measured and then macro-averaged to define a single measure that encapsulates the discriminant performance of the models. The Figure 1 .illustrates the sequential workflow of the methodology of the proposed study.

## 3 Results and Discussion

This study compared the performance of Random Forest (RF), Support Vector Machine (SVM) and Gradient Boost (GB) using Harvard Mental Health Dataset of 120 patients which includes 28 BD Type I, BD Type II, Major Depressive Disorder and Normal. For evaluation the matrices used were accuracy, precision, recall, F1 score, confusion matrix and ROC-AUC curves.

**Table 1.** Classification report of machine learning models

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Random Forest | 0.8833 | 0.8982 | 0.8804 | 0.8785 |
| Support Vector Machine | 0.8329 | 0.8560 | 0.8339 | 0.8227 |
| Gradient Boost | 0.8156 | 0.8436 | 0.8143 | 0.8052 |

The Table 1. above shows the comparative performance of the machine learning models used in this study. Random Forest outperformed the other methods, achieving the highest accuracy (0.8833) and simultaneously leading in precision (0.8982), recall (0.8804) and F1-score (0.8785), confirming its ability to limit both false positives and false negatives. Support Vector Machine ranked second with its accuracy of 0.8329 and balanced precision–recall pair (0.8560 / 0.8339) yielded an F1-score of 0.8227. Gradient Boost, despite marginally elevated precision (0.8436), underwent reduced recall (0.8143), leading to the lowest accuracy (0.8156) and F1-score (0.8052); thus, its extra model complexity brought no compensating benefit. Collectively, these metrics establish Random Forest as the most reliable classifier for this dataset, with SVM providing a moderate alternative and Gradient Boost offering the least favorable trade-off.

Figure. 5 shows the confusion matrix of the Random Forest (RF) model. The Random Forest classifier achieved an overall accuracy of 88.3% (106/120), demonstrating particularly strong performance on Depression (90.3% sensitivity, 28/31 correctly identified, with three cases mislabeled as Normal) and Normal subjects (93.3% sensitivity, 28/30, with two misclassified as Depression). Bipolar Type II was detected with 87.1% sensitivity (27/31, all four errors routed to Type I), while Bipolar Type I showed the lowest sensitivity at 82.1% (23/28), with two instances mislabeled as

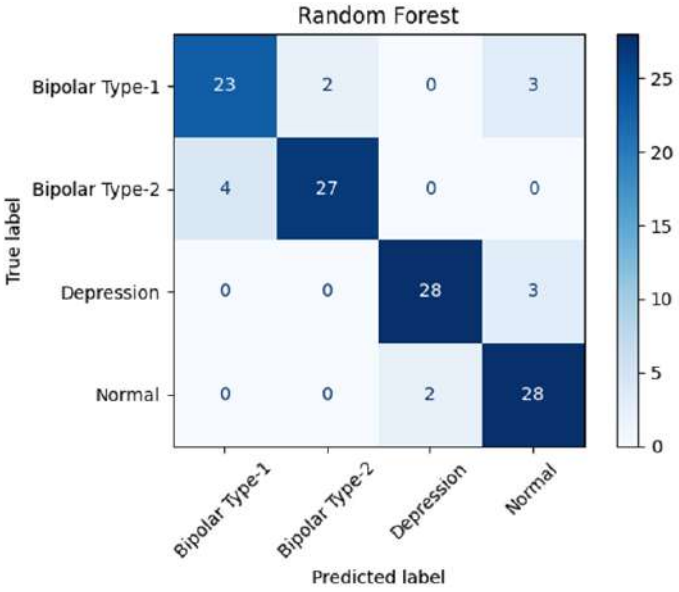


**Figure 5.** Confusion Matrix of Random Forest (RF) model

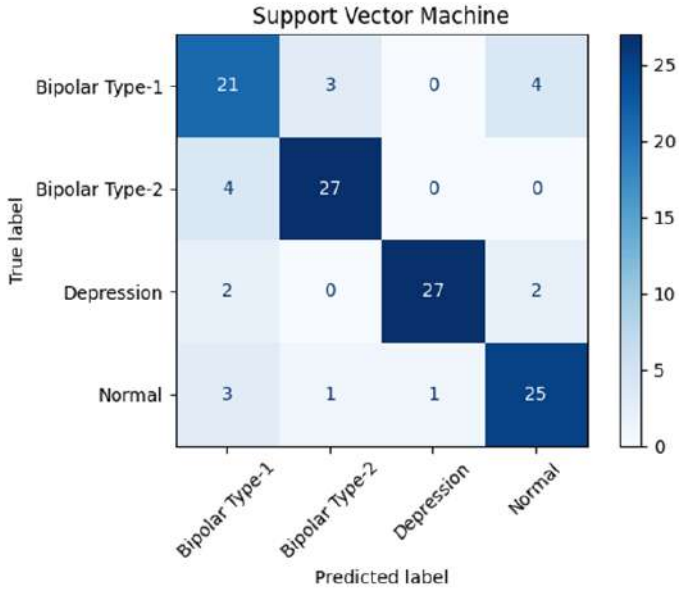


**Figure 6.** Confusion Matrix of Support Vector Machine (SVM) model

Type II and three as Normal. These results underscore the model's robust identification of Depression and healthy controls, alongside the remaining challenge of distinguishing between bipolar subtypes.

Figure 6. shows the confusion matrix of the Support Vector Machine (SVM) model. The Support Vector Machine classifier achieved an overall accuracy of 83.3% (100/120). It detected Bipolar Type I with only 75.0% sensitivity (21/28), mislabeling three cases as Type II and four as Normal. Bipolar Type II and Depression were identified more reliably—87.1% sensitivity in each case (27/31), the former's errors all routed into Type I and the latter split evenly between Type I (2) and Normal (2). Normal subjects were correctly classified 83.3% of the time (25/30), with the remaining five cases scattered across all three pathological labels. These results highlight the SVM's strength in identifying Bipolar II and depressive presentations, while confirming that distinguishing Bipolar I from both its subtype and euthymic controls remains its primary limitation.

Figure7. shows the confusion matrix of the Gradient Boost Classifier (GB) model. The Gradient Boost model attained an overall accuracy of 81.7% (98/120), with class-wise sensitivities varying notably. Bipolar Type I proved the most challenging to detect (67.9%, 19/28 correct), yielding six misclassifications as Type II and three as Normal. Both Bipolar Type II and Depression achieved 87.1% sensitivity (27/31 each), the former's errors all routed to Type I and the latter's entirely to Normal. Normal subjects were correctly identified 83.3% of the time (25/30), with the remaining five cases dispersed across Type I (one), Type II (two), and Depression (two). These results underscore that, while Gradient Boost reliably recognizes Bipolar II and depressive presentations, it struggles most with Bipolar I and exhibits cross-talk between pathological and healthy classes.

Figure 8. presents the class-specific ROC curves for the Random Forest classifier, revealing very high discrimination capacity across all diagnostic categories. Depression achieves the highest AUC (0.995), closely followed by Normal (0.984) and Bipolar Type II (0.975), indicating near-perfect discrimination from the other groups. Even Bipolar Type I, which proved most challenging in the confusion analysis, attains an AUC of 0.926, reflecting strong true-positive rates at low false-positive thresholds. The steep rises at the left of each curve underscore the model's ability to

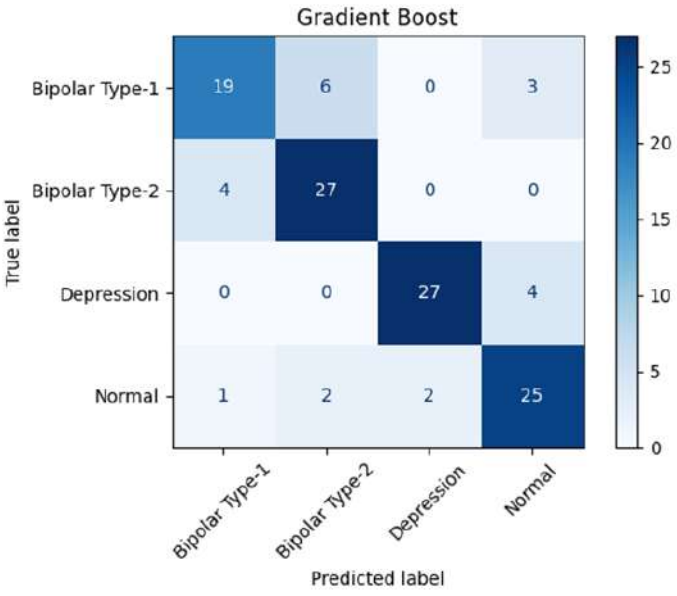


**Figure 7.** Confusion Matrix of Gradient Boost Classifier (GB) model.

capture genuine cases with minimal misclassification, while the small differences in AUC quantify the relative difficulty of distinguishing subtypes. Collectively, these ROC-AUC metrics confirm that the Random Forest approach provides exceptional overall performance in differentiating between mood disorder subtypes and healthy controls.

Figure 9. illustrates the per-class ROC curves for the Support Vector Machine, demonstrating excellent discrimination across all diagnostic groups. Depression again reaches near-perfect discrimination (AUC = 0.997), followed by Bipolar Type II (0.984) and Normal controls (0.973), indicating that these categories are almost fully distinguishable from the others. Bipolar Type I, while still strong (AUC = 0.944), exhibits the shallowest initial slope, reflecting greater overlap with non-Type I cases at low false-positive rates. The sharp ascents of the Depression and Bipolar II curves near the y-axis confirm the model's capacity to capture true positives with minimal spurious classifications, whereas the more gradual rise for Bipolar I quantifies its relative diagnostic challenge. Overall, the SVM's ROC-AUC profile underscores its robust performance in detecting mood disorder subtypes, with the only modest decrement observed for the most clinically subtle category.

Figure 10. displays the ROC curves for the Gradient Boost classifier, revealing strong but variable discrimination across diagnostic categories. Depression attains the highest AUC (0.969), followed closely by Bipolar Type II (0.959) and Normal controls (0.950), indicating that these conditions are largely separable from other classes. Bipolar Type I yields the lowest AUC (0.898), with its curve rising more gradually at low false-positive rates, reflecting residual overlap with non-Type I presentations. The sharp initial ascent of the Depression and Type II curves underscores the model's capacity to capture true cases with minimal misclassification, whereas the more modest slope for Bipolar I quantifies its relative diagnostic subtlety. Overall, the Gradient Boost detector maintains excellent overall performance, with its only notable decrement occurring for the most clinically nuanced category.

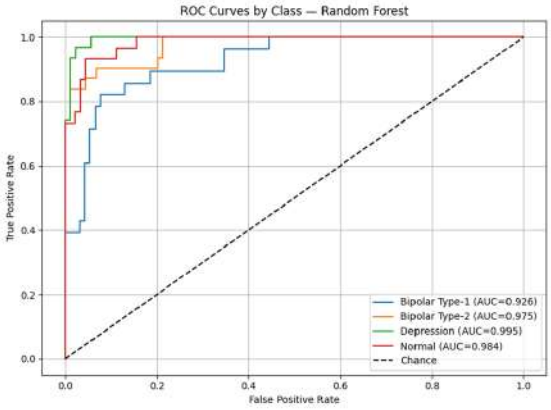


**Figure 8.** ROC curves by class for the Random Forest classifier.

To comprehensively evaluate machine learning approaches for multi-class mental health classification, we compared previous structured experimental design with our proposed method. As Figure 11. illustrates that the proposed study is a contribution over prior work, not only at the methodology level, but also at the empirical one since the search space is extended from the use of a Random Forest model to a combination of 3 classifiers including Random Forest,

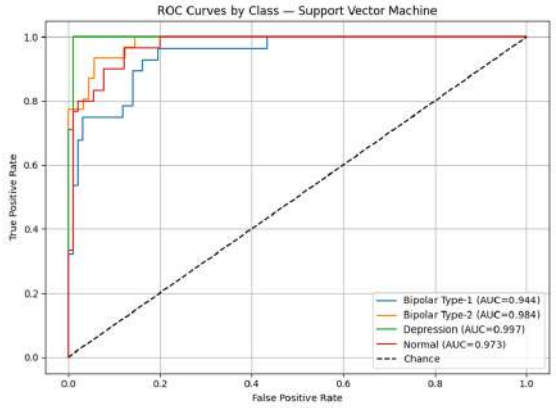


**Figure 9.** ROC curves by class for the Support Vector Machine classifier.

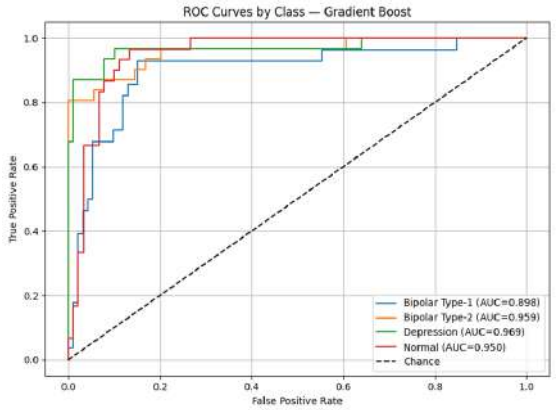


**Figure 10.** . ROC curves by class for the Gradient Boost classifier.

SVM, and Gradient Boosting. In addition, it uses a more compact stratified seven-fold cross-validation and, unlike the prior approach, replaces the ordinal or min-max preprocessing with Standard Scaler on numeric features and One Hot encoding on categorical features [30]. Overall, these refinements improve Precision from 0.87 to 0.89; Recall from 0.87 to 0.88; and Accuracy 0.87 to 0.88 with F1 score of 0.87, while the new concept of macro-averaged F1 and ROC equally pivotal in affecting Discriminative Performance in favor of class-balanced measurements [30]. These results indicate that richer model diversity and more rigorous data handling yield modest yet meaningful gains in predictive performance and produce metrics that better capture class-level behavior, thereby enhancing the robustness and generalizability of the findings.

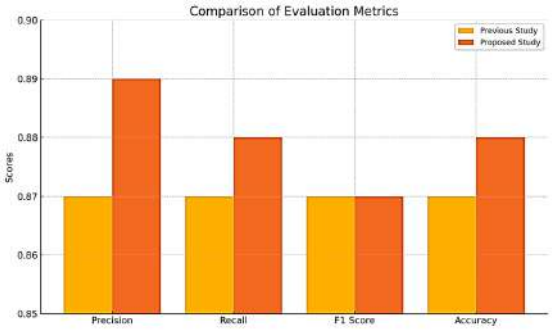


**Figure 11.** Comparisons of previous[30] and proposed study.

## 4 Conclusion

This research analyzed the predictive effectiveness of three popular machine learning techniques Random Forest, Support Vector Machine (SVM), and Gradient Boosting concerning recognizing various mental health disorders from a structured dataset. Leveraging a stratified 7-fold cross-validation framework and a robust preprocessing pipeline combining standard scaling and one-hot encoding, we evaluated each model across multiple performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Out of all the models we evaluated, Random Forest emerged as the most robust and balanced in performance. Even though SVM slightly underperformed in ROC AUC, its lower recall and F1 scores suggested inconsistency across several metrics. Although Gradient Boosting scored above baseline expectations, it underperformed in all core metrics compared to other models. The results suggest that not only is Random Forest robust, but also performs reliably well on complex multi-class mental health datasets which makes it suitable for real-world scenarios pertaining to mental health evaluations if preprocessing and balanced validation techniques are used.

## 5 Limitation and Future Direction

While the present study does offer promising evidence for the value of machine learning in classifying bipolar disorder and comorbid conditions, there are some

limitations to note. Firstly, the sample size of the dataset is low, which might limit the generalizability of the results. Secondly, the dataset is restricted for analyses across different clinical or demographic populations. Furthermore, limited behavioral features were available, which together are relatively impoverished relative to multimodal inputs such as neuroimaging, genetic, or digital behavioural markers. Such constraints provide a step forward in future research based upon larger and more diverse datasets comprising more feature representations in order to help in generalizing and strengthening the models.

## Author Contributions

**Syed Ibad Hasnain**: Conceptualization, Methodology, Software **Rabika Kamal**: Data curation, Writing- Original draft preparation. **Hafsa Israr**: Visualization, Investigation. **Muhammad Faris**: Software, Validation. **Hafiza Syeda Yusra Tirmizi**: Writing- Reviewing and Editing

## Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors.

## References


[1] I. Grande, M. Berk, B. Birmaher, and E. Vieta, "Bipolar disorder," *The Lancet*, vol. 387, no. 10027, pp. 1561–1572, 2016.

[2] D. J. Miklowitz and S. L. Johnson, "Bipolar disorder," *Psychopathology: History, Diagnosis, and Empirical Foundations*, 3rd ed., 2017, pp. 317–363.

[3] R. S. McIntyre *et al.*, "Bipolar disorders," *The Lancet*, vol. 396, no. 10265, pp. 1841–1856, 2020.

[4] V. Oliva *et al.*, "Bipolar disorders: an update on critical aspects," *The Lancet Regional Health–Europe*, vol. 48, 2025.

[5] T. M. Gencheva *et al.*, "Diagnostic value of structural, functional and effective connectivity in bipolar disorder," *Acta Psychiatrica Scandinavica*, vol. 151, no. 3, pp. 192–209, 2025.

[6] M. Leboyer and D. J. Kupfer, "Bipolar disorder: new perspectives in health care and prevention," *J. Clin. Psychiatry*, vol. 71, no. 12, p. 1971, 2010.

[7] H. B. Türközer *et al.*, "Familial High-Risk for Psychosis and Bipolar Disorder Amongst Youth," *Early Intervention in Psychiatry*, vol. 19, no. 2, p. e70019, 2025.

[8] D. A. Clayton and G. M. Burlingame, "Evidence-based group therapy for mood disorders," *Group Dynamics*, 2024.

[9] P. J. Harrison *et al.*, "Innovative approaches to bipolar disorder and its treatment," *Ann. N. Y. Acad. Sci.*, vol. 1366, no. 1, pp. 76–89, 2016.

[10] A. Liberona *et al.*, "BDNF as a predictor of treatment response," *Int. J. Mol. Sci.*, vol. 25, no. 20, p. 11204, 2024.

[11] M. N. Modesti *et al.*, "Lipidomics and genomics in mental health," *Lipids in Health and Disease*, vol. 24, no. 1, p. 89, 2025.

[12] L. M. Dietze *et al.*, "White matter microstructure in obesity and bipolar disorders," *Molecular Psychiatry*, 2024.

[13] S.-Y. Ye *et al.*, "Lower serum IGF-2 in BD," *Frontiers in Psychiatry*, vol. 15, p. 1354999, 2024.

[14] L. D. Hawke, S. V. Parikh, and E. E. Michalak, "Stigma and bipolar disorder," *J. Affect. Disord.*, vol. 150, no. 2, pp. 181–191, 2013.

[15] J. R. Geddes and D. J. Miklowitz, "Treatment of bipolar disorder," *The Lancet*, vol. 381, no. 9878, pp. 1672–1682, 2013.

[16] T. de A. Cardoso *et al.*, "Digital tools for detection and treatment of BD," *JMIR*, vol. 11, p. e58631, 2024.

[17] L. de S. Rotenberg *et al.*, "Predict depressive relapses in BD," *J. Affect. Disord.*, vol. 295, pp. 681–687, 2021.

[18] J. Mateo-Sotos *et al.*, "Identify BD patients via ML," *CSSP*, vol. 41, no. 4, pp. 2244–2265, 2022.

[19] I. Perez Arribas *et al.*, "Distinguishing BD and BPD," *Translational Psychiatry*, vol. 8, no. 1, p. 274, 2018.

[20] A. K. Jihoon *et al.*, "Chronotherapeutic intervention," *BMJ Mental Health*, vol. 28, no. 1, p. e301338, 2025.

[21] M. Suárez *et al.*, "Random Forest for accurate BD classification," *Life*, vol. 15, no. 3, p. 394, 2025.

[22] F. Colombo *et al.*, "ML for BD prediction," *Neurosci. Biobehav. Rev.*, vol. 135, p. 104552, 2022.

[23] C.-H. Cho *et al.*, "Mood prediction by ML using digital phenotypes," *J. Med. Internet Res.*, vol. 21, no. 4, p. e11029, 2019.

[24] S. Kim, H.-K. Lee, and K. Lee, "Screening mood symptoms using MMPI-2-RF," *J. Pers. Med.*, vol. 11, no. 8, p. 812, 2021.

[25] N. Topolski, S. H. Jeong, and B. Mwangi, "ML applications in mood disorders," in *Mood Disorders*, 2021, p. 206.

[26] Q. Yang *et al.*, "Classification for psychiatric disorders using ML," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 5054–5064, 2022.

[27] E. Turska, S. Jurga, and J. Piskorski, "Mood disorder detection in adolescents," *Entropy*, vol. 23, no. 9, p. 1210, 2021.

[28] H. Abou-Warda *et al.*, "Random forest model for mental disorders," in *AISI 2016 2*, Springer, 2017, pp. 670–680.

[29] J. M. Rodolpho *et al.*, "Biomarkers and Mental Disorders: RF analysis," *Biomolecules*, vol. 15, no. 6, p. 793, 2025.

[30] R. Choirunnisa, M. Anshori, and W. Kusuma, "Improving RF evaluation in mental health," *RIGGS J. AI & Digital Business*, vol. 4, pp. 3526–3534, 2025.

[31] H. Karbalaeipour *et al.*, "A Collection of 120 Psychology Patients with 17 Essential Symptoms to Diagnose Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, and Normal Individuals," Harvard Dataverse, doi:10.7910/DVN/0FNET5.