Bipolar Disorder Prediction using Machine Learning Techniques

Aarya Bhatia[1], Jashwanth Rushi[2], Ayush Patel[3]

[1,2]Department of Computer Science & Technology, Manav Rachna University, Faridabad, Haryana, India
[3] Manav Rachna International Institute of Research and Studies, Faridabad, Haryana
{[1]maddoganddemigod123@gmail.com,
[2]jashwanthrushi004@gmail.com,
[3]aayushpl06@gmail.com}

Abstract. Bipolar disorders are mental disorders characterized by extreme mood swings, also known as manic depression. Individuals with bipolar disorder oscillate between periods of intense elation and energy (mania) and profound sadness. However, the machine learning models could assist psychiatrists or psychologists identify children or teens who are in jeopardy of developing bipolar disorder. Additionally, the researchers verbally express that the ML models could be integrated into the medical system to avail psychiatrists catch early warning signs of bipolar disorder. For the diagnoses, we have described Multi-Output (Random Forest) Regressor utilizing dataset consisting of mood-ness scale and physical assessments. The model being utilized for regression model will be taking 4 outputs as Y and rest of them are X. Among the regressor model, we will also use other regressors for soothsaying the test results and optically discern how well it performs compared to our First model.

Keywords: machine Learning, Regression, Bipolar disorder, Clinical assessments, Multioutput Regressor, Gradient Boosting, Support Vector Machine, K-Nearest Neigh-bours, R-squared

1 Introduction

Due to booming use of social media, peer stress, accelerated pressure stage, opposition, expectancies, and so forth. have led to unstable teenage minds. Bipolar disorders (BD) are one of the most debilitating diseases, with an approximate lifetime incidence of four-five% inside the preferred population. Artificial intelligence (AI) and machine learning (ML) have in recent years identified rapid optical zooming in data analysis. The initial optimization of computer technologies usually approves applications to work effectively in the real world. Machine learning, a branch of artificial astuteness, deals with training machines according to a given dataset and utilizing that learning for future decisions. A trained system can mimic the way humans learn and analyze, improving predictive precision over time. Machine learning is a promising new method for diagnosing and treating bipolar disorder. Machine learning algorithms can be used to analyze brain and clinical data to identify patterns associated with bipolar disorder. This can be used to create diagnostic tools and personalized treatment plans. More research is needed to develop and validate machine learning tools for bipolar disorder, but the results are promising. This research [1] has attracted considerable attention from the psychiatric and artificial intelligence communities by automating the diagnosis of psychiatric disorders such as bipolar disorders (BDs) by machine learning technology. These methods are based mainly on a variety of biomarkers extracted from data from electroencephalograms (EEG) and magnetic resonance imaging (MRI)/functional MRI (fMRI).The introduction of multiple output regressors (MOR) in the field of bipolar disorder creates great opportunities to improve the understanding and treatment of this complex disease. By analyzing comprehensive databases containing patient information, MOR has been shown to be able to reveal complex patterns and is associated with the onset and progression of mental illness (bipolar). By leveraging MOR based machine learning algorithms, we can increase early diagnosis, develop personalized treatment plans, and predict changes in mood. This new approach could revolutionize the way we treat bipolar disorder by combining technology with mental health, improving the quality of life for people affected by the disease. The potential of MOR provides a paradigm shift in our understanding of bipolar disorder disease by leveraging the potential to collect extensive patient data. By delving into this rich data, MOR is able to identify complex and relevant patterns and provide new clues to disease onset and progression.

In the following section, the related work in the field have been thoroughly examined and discussed. In Section 2, there are literature reviews of other studies contributing to the understanding of the diagnosis and treatment of bipolar disorder. This review examines recent research on the genetic and environmental risk factors, neural mechanisms and clinical interventions for bipolar disorder in adults and adolescents. In Section 3, we provide a detailed description of the dataset used in our study. Our research objectives and the solutions that we have proposed to achieve them are presented in Section 4. Finally, in Section 5, we provide a comprehensive list of all the referred papers in this research.

2 Related Work

This paper [2] aims to summarize information on the use of ML techniques to predict schizophrenia and BPD, and to assist in early and timely diagnosis of disease. In this review, 1243 documents were extracted through database searches, of which 15 were included according to full-text assessment. The main algorithms used were support vector machines (SVMs), random forests (RFs) and gradient boosting (GBs). The input and output characteristics are very diverse and are maintained to allow future research. RF algorithms are much more accurate and sensitive than SVM and GB algorithms.

The GB has a much higher specificity than SVM and RF. We did not find any significant differences in the specificity between RF and SVM. ML can predict results accurately and help to make clinical decisions regarding schizophrenia and BPD and for that RF is often better suited to supervised learning tasks than other algorithms. This review [3] looked into the relation between child hood trauma and bipolar disorder in adults. It has been found that any abuse of children can cause long-term neurobiological and permanent morphological changes in victim's brains. Overstimulation of gray matter leads to decreased hippocampus volume, activated Amygdala and impairment of prefrontal, frontal, and posterior, and these changes lead to personality changes in individuals with BPD. Child abuse is a risk factor for BPD development and is a public concern. Understanding the impact of early childhood negative stress factors on adulthood requires a serious emphasis on early diagnosis and intervention. This study [4] aimed to investigate the effects of borderline personality disorder on the outcome of inpatient treatment for anorexia nervosa. There was no age or gender difference between patients with anorexia and patients with BPD. Patients with AN+BPD differ from AN patient in that they show a higher general and specific mental illness of food disorders and only partially improve under special inpatient treatment. In particular, aspects of the regulation of emotions and core AN symptom, such as body dissatisfaction and perfectionism, need to be addressed even more explicitly in patients with common diseases. The purpose of this research [6] is to use machine learning (ML) to distinguish between BP and BPD and to distinguish between concurrent bipolar disorders (BP)/borderline personality disorders (BPD). Participants were assigned DSM diagnoses and compared to self-reporting measures to examine personality, emotional control strategies and child-oriented experiences. We identified more distinctive clinical variables in BP/BPD associated with BP, which differ with greater precision than BP and less than BPD alone. The aim of the study [7] is to distinguish between patients with bipolar disorders and healthy control patients with greater precision by using a broader cognitive assessment and new machine learning algorithms. For neurocognitive evaluation, six neurocognitive tests from the CANTAB test battery were used, and the polyhedral conic function algorithm was used to classify participants. Patients with bipolar disorders differ from healthy controls with a 78% accuracy. Our study demonstrates an algorithm for predicting high accuracy using the CANTAB cognitive brain battery that separates bipolar disorders from healthy controls.

3 Dataset Description

The Bipolar Disorder Dataset [8] contains relevant information for individuals diagnosed with bipolar disorder. Data are collected for research and clinical purposes and may include a variety of behaviour and measures relevant to the study of bipolar disorder. This dataset is taken from Kaggle under no author, no DOI citation, no coverage and license is unknown. This table represents a data set that is multivariate and has clinical features in nature, with 77,500 instances. The area to which the data pertains is not specified. The attribute characteristics are real, and there are 31 attributes included in the data set. The data donated is not specified and the associated tasks for the data set are classification. There are no missing values in the data set.

4 Research Objectives

1. **To Evaluate the Predictive Performance of Machine Learning Models:** Assessing the effectiveness of machine learning models, specifically the Random Forest Regressor and Gradient Boosting Regressor, in predicting bipolar disorder clinical assessments.
2. **To Compare the Accuracy of Model Predictions:** Comparing the accuracy of predictions made by the Random Forest Regressor and Gradient Boosting Regressor in diagnosing bipolar disorder clinical outcomes.
3. **To Contribute to Mental Health Care:** Contribute findings that can potentially improve the accuracy and efficiency of diagnosing bipolar disorder, thus enhancing patient care and mental health outcomes.

5 Methodology

The proposed methodology for predicting the bipolar disorder is shown in the figure1.

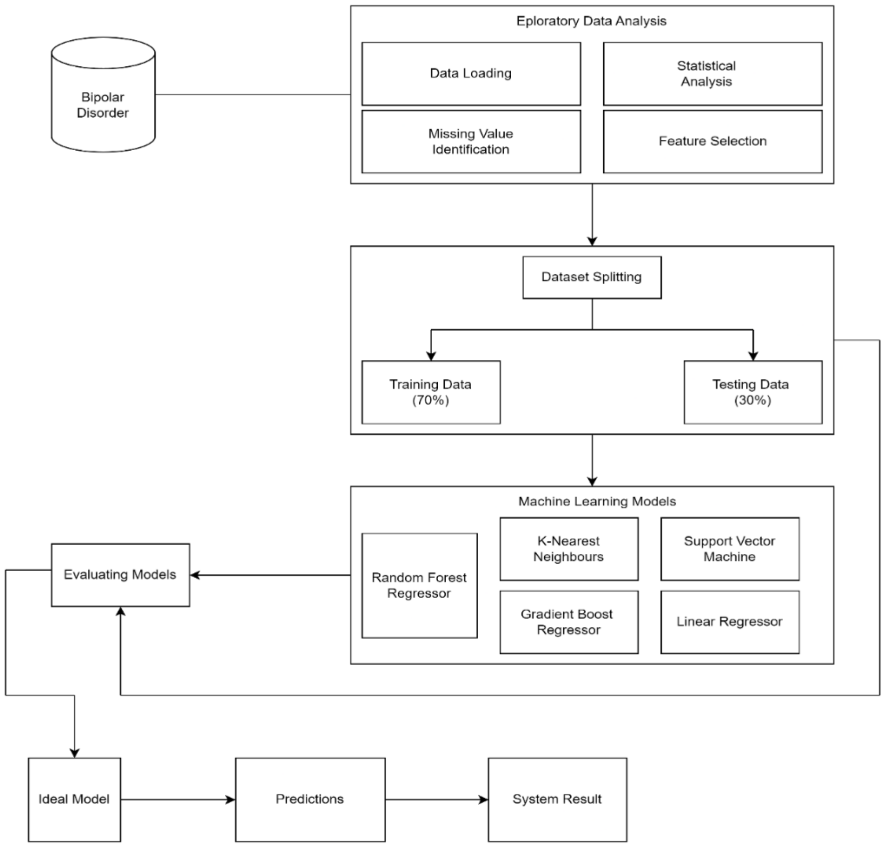

Fig. 1. Block Diagram of Model System

We have split the dataset into training and testing set. Then, we used a 70-30 split, with 70% of the data used for the training set and 30% for testing set. Some of the columns were not used because they did not have any use such as ('s.no', 'Sub1', 'Sub2', 'Sub3', 'Sub4', 'Ts', 'name'). So, after dropping them, X and y set were used for train-test-split method. There are 4 outputs in the dataset, so y1, y2, y3, y4 were created to train and 'Sub1', 'Sub2', 'Sub3', 'Sub4' were used as the outputs. The utilization of multioutput regression techniques, consisting of the Multi-Output Regressor, inside the context of regression modeling has received prominence due to its ability to deal with critical demanding situations in records analysis. In this research paper [5], the authors emphasize the importance of multioutput regression, which surpasses the constraints of conventional single-output methods. one of the key benefits lies in its capability to model complex relationships in each unimodal and multi-modal records, making it relevant to a huge range of regression troubles. moreover, multioutput regression handles multi-output regression responsibilities, even when co-dependencies between reaction variables exist, that's an important function in actual-world statistics analysis. As bipolar disease clinical tests may additionally involve a couple of associated metrics, we used the Multi-Output-Regressor wrapper to address the multioutput nature of the problem. It allows us to expect a couple of clinical metrics simultaneously. Then, we wrapped both of them models in multi-output-regressor so that we could avoid 1-D array problem as the models don't use 2-D array. We Trained both the Random Forest Regressor and the Gradient Boosting Regressor on the training dataset. The models learned to predict the combined set of clinical metrics. After fitting both the multioutput regression model with the training set comprising both the (X) and the corresponding variable (y), we proceeded to evaluate its predictive performance. To evaluate the model's accuracy, we calculated scores indicative of its ability to correctly predict the outcomes. Moreover, we computed the R-squared rating as a measure of the model's goodness-of-fit, offering insights into how well the model captures the variance in the facts. This technique allowed us to thoroughly compare the version's overall performance and its ability to explain the variability in the observed outcomes. We have also used Linear regression, Support vector regressor and KNN regressor to compare their scores with the above models. After finding about their scores and comparing them with each other, We have concluded that our Random Forest regressor and Gradient boosting regressor wrapped in Multioutput regressor gave the best results according to the Bipolar-disorder clinical assessments. Now for the graph of difference between actual and predicted values, for this, we selected Random Forest Regressor as It is the best model by so far and we chose Residual Plot for the Comparison as shown in figure2.

Table 2. Metrics performance for various regressor models

REGRESSOR	MSE	MAE	R-SQUARED
RGR	0.037	0.030	99.25%
GBR	0.0816	0.114	98.74%
LR	1.46	0.76	77.69%
SVR	0.82	0.37	87.24%
KNN	0.75	0.34	88.49%

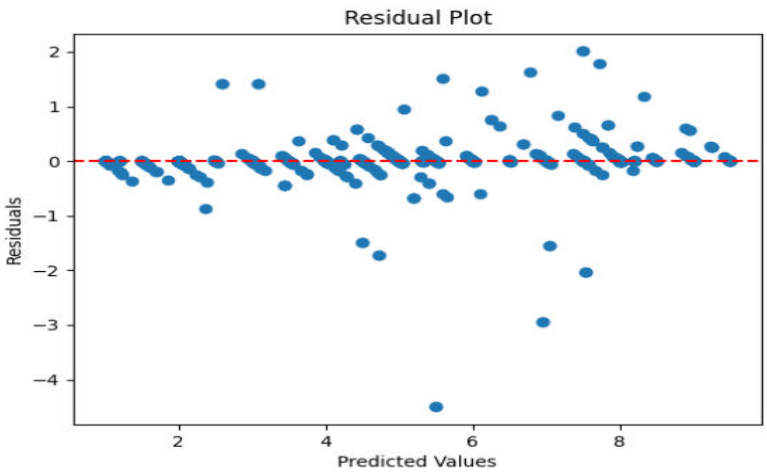

Fig. 2. Residual Plot diagram of Random Forest Regressor

5 Conclusion and Future Work

In our work, we have concluded that the application of machine learning to the field of bipolar disorder signifies a substantial advancement in our knowledge and treatment of this complicated illness. With the ability to provide early diagnosis, individualized treatment plans, and ongoing monitoring, machine learning algorithms hold the potential to completely transform our understanding of bipolar disorder and greatly enhance the lives of those who are impacted. The use of a Multioutput Regressor in combination with a dataset of assessment tools for bipolar disorder has shown promise in improving our knowledge and treatment of this complex mental health issue.

References

[1] W. A. Campos-Ugaz et al., “An Overview of Bipolar Disorder Diagnosis Using Machine Learning Approaches: Clinical Opportunities and Challenges,” 2023.

[2] M. Montazeri, M. Montazeri, K. Bahaadinbeigy, M. Montazeri, and A. Afraz, “Application of machine learning methods in predicting schizophrenia and bipolar disorders: A systematic review,” Health Sci Rep, vol. 6, no. 1, Jan. 2023, doi: 10.1002/hsr2.962.

[3] P. Mainali, T. Rai, and I. H. Rutkofsky, “From Child Abuse to Developing Borderline Personality Disorder Into Adulthood: Exploring the Neuromorphological and Epigenetic Pathway,” Cureus, Jul. 2020, doi: 10.7759/cureus.9474.

[4] U. Voderholzer, M. Favreau, S. Schlegl, and J. B. Hessler-Kaufmann, “Impact of comorbid borderline personality disorder on the outcome of inpatient treatment for anorexia nervosa: a retrospective chart review,” Borderline Personal Disord Emot Dysregul, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40479-021-00149-7.

[5] T. R. F. Phillips et al., “Multi-Output Regression with Generative Adversarial Networks (MOR-GANs),” Applied Sciences (Switzerland), vol. 12, no. 18, Sep. 2022, doi: 10.3390/app12189209.

[6] Patel RS, Manikkara G, Chopra A. Bipolar Disorder and Comorbid Borderline Personality Disorder: Patient Characteristics and Outcomes in US Hospitals. Medicina (Kaunas). 2019 Jan 14;55(1):13. doi: 10.3390/medicina55010013. PMID: 30646620; PMCID: PMC6358827.

[7] Sonkurt, H. O., Altınöz, A. E., Çimen, E., Köşger, F., & Öztürk, G. (2021). The role of cognitive functions in the diagnosis of bipolar disorder: A machine learning model. International Journal of Medical Informatics, 145, 104311. doi: 10.1016/j.ijmedinf.2020.104311

[8] https://www.kaggle.com/datasets/peerbasha/prediction-of-bipolar-disorder-students/data, last accessed 2016/11/21.