Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 2, November 2024, pp. 1294~1300
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i2.pp1294-1300

Hybrid logistic regression support vector model to enhance prediction of bipolar disorder

Nisha Agnihotri, Sanjeev Kumar Prasad
School of Computing Science and Engineering, Galgotias University, Greater Noida, India

Article Info

Article history:

Received Apr 21, 2024
Revised Aug 5, 2024
Accepted Aug 11, 2024

Keywords:

Bipolar disorder
Hybrid model
Logistic regression
Supervised machine learning
Support vector machine

ABSTRACT

Bipolar disorder has become one of the major mental health issues due to stressed life around the world. This is the major reason for suicides these days as these people are unable to convey their feeling and emotions to others. This proposed research shows the logistic regression and support vector machine hybrid model to predict bipolar disorder in patients is to develop an accurate and reliable model that can effectively predict the presence of bipolar disorder in patients based on their clinical and demographic data. The purpose is to make a framework that can help healthcare professionals diagnose bipolar disorder early, thereby enabling timely and appropriate treatment to be provided. The model should take into account various patient-specific features, such as age, gender, family history, medication use, and other medical conditions, in addition to relevant clinical and demographic variables. The aim is to create a model that can accurately classify patients with bipolar disorder and non-bipolar disorder patients while minimizing false-positive and false-negative classifications. The work shows improvement in evaluation detection in performance with hybrid logistic support vector regression (LSVR) to detect disorder and protect them to avoid worse situation.

Corresponding Author:

Nisha Agnihotri
School of Computing Science of Engineering, Galgotias University
Greater Noida, Uttar Pradesh, India
Email: nisha.agnihotri_phd20@galgotiasuniversity.edu.in

1. INTRODUCTION

Machine learning has shown promising results in predicting bipolar disorder in humans by analyzing large datasets of clinical and biological markers. By using algorithms to learn patterns in these datasets, machine learning models can identify early signs and risk factors associated with bipolar disorder before the onset of symptoms. The hybrid logistic support vector regression (LSVR) model is a machine learning algorithm designed to predict bipolar disorder in patients. It combines two different types of models, namely the support vector regression (SVR) and the logistic regression (LR) model, to leverage their respective strengths in predicting the presence of bipolar disorder in patients.

Using machine learning techniques, they prepared a combined model framework which helps to classify mental health features that have occurred in social media communities [1]. The dataset also undergoes through empirical validation process and the results exceeded the recent state of the art baselines created by this model [2]. The study in this later shows a multimodal machine learning model which is created on different features such as visual, acoustic and textual through performing a cross modality correlation methodology.

The study showcased that deep learning and machine learning model helps in detecting bipolar disorder by involving a methodology that captures the increased binary decision diagram (BDD) accuracy

rate [3]. To understand and deduce the systematic review, the study applied the PRISMA methodology which included research of total 30 articles in their initial reviewing and screening stage [4]. The researched focus on identifying mental health disorder in patients using a correlation in their textual, visual and audio modalities [5]. They implemented recurrent neural networks methodology to include the temporary information in the machine learning process and based on that deploying a model which extract the features based on dynamic evolution. They created a diagnostic algorithm which helped in identifying bipolar disorder patients in different clinical scenarios with efficient accuracy and helps to perform accurate clinical treatment to patients [6]. The study proposed a recurrent decision tree which helped in contributing towards the prediction of bipolar disorder [7]. The results predicted that the proposed methodology outperforms the existing decision tree model. The study found out that their algorithm in comparison to the preexisting machine learning models which follow Bayesian linear discriminant technique, decision tree or Gaussian Naïve Bayes (NB) resulted to accuracy around 97% [8]. The hybrid NBTree showed that the results have better accuracy and precision as compared to the other two approaches. In terms of response time the NB exceeds the decision tree and NBTree algorithms [9]. The best thing about machine learning algorithms is that once the model is trained in reference to the outputs of data, it can work on automation [10]. This paper provided a brief review and gave a future prospect of how machine learning solves different problems. The research helps in finding a relative performance study of supervised algorithms of machine learning in predicting various diseases [11]. For understanding classification studies, a hybrid logistics vector trees classifier (LVTrees) model is performed and analyzed logistic regression, support vector classifier, and extra tree classifier. LVTrees models has exceeded other model implementing ADASYN and Chi2 techniques with an accuracy result of 100%. Conducting a T-test on the research helps to find the efficacy of the proposed algorithm and the k- fold cross-validation also helps to prove the superiority of the model [12]. The aim of the study is to find out how much mental stress does college students undergo at different phases of college life. The pressure they undergo during exams time or the stress that occurs during placement time usually goes unnoticed [13]. A hybrid support vector machine (SVM) model was applied to find out the distinct features and calculated high - dimensional inputs with proper accuracy [14]-[16]. This study found out that SVM has comparatively best performance in classification as compared to logistic regression method as compared to the different types and the background of these datasets [17]-[19]. This study reflected that SVM model has better performance in classification as compared to logistic regression model in both training and testing datasets. This paper researched about the comparative study in the performance of different models by implementing different types of background of the datasets taken in consideration and transformation of these models according to the categorical scale in these predicted variables. Further these models were evaluated on the basis of Press'Q statistic and apparent error rate (Aper) [20]. This study compared the convolutional neural network - bidirectional long short-term memory (CNN-bi LSTM) model with the existing CNN and recurrent neural networks (RNN) with the baseline approaches. The accuracy of neural networks can be enhanced by improving the optimization using hyper parameters [21]. In this research, two different models are employed in the forecasting of data, Gaussian process regression and exponential smoothing. None of them provided an improvement in persistent baseline [22]. The experimented results predicted that this model helps to provide better predictive performance as calculated through various performance metrics. This finally concluded that the depression time series based on the bipolar disorder patient's dataset are fairly heterogeneous and resulted in constraints of the accuracy of different forecasting based on automated mood forecasting in the patients [23], [24]. Deep learning using temporal graphics of clinical history is a new and promising analytical approach for mental health outcome prediction [25]. This review evaluated the magnitude of the risk of developing bipolar disorder I-II in individuals at clinical high-risk [26].

To analyze the performance more significantly, various visualization approaches and statistical techniques were employed which proves a great difference between non-depressive content and linguistic representation of depressive. As shown in Figure 1 young one is suffering with depression anxiety and other mental illness. All the above related study also depicts the percentage of people suffering with mental illness and their types.

2. METHOD

The programmed algorithms of machine learning are used to optimize and learn the operations by input data analysis in order to make predictions in an acceptable limit or range. Using new data programmed algorithms give more accurate results and prediction performance. Machine learning algorithms are grouped according to inputs and purpose, they are classified in three broad categories. They are supervised learning, unsupervised learning, semi-supervised learning. Supervised machine learning uses labelled training dataset which is first used to train the dataset in underlying algorithm. The outcome of this trained dataset is given to unlabeled test dataset which is categorized into similar groups. The supervised learning is of two types:

classification problems and regression problems. Here we are using classification problem, which uses underlying output variable is discrete. The variable is classified in different categories or parts like 'bipolar disorder' and 'Controlled', and BD-I or BD-II. The general steps involved in building a hybrid LSVR-model to predict bipolar disorder in patients.

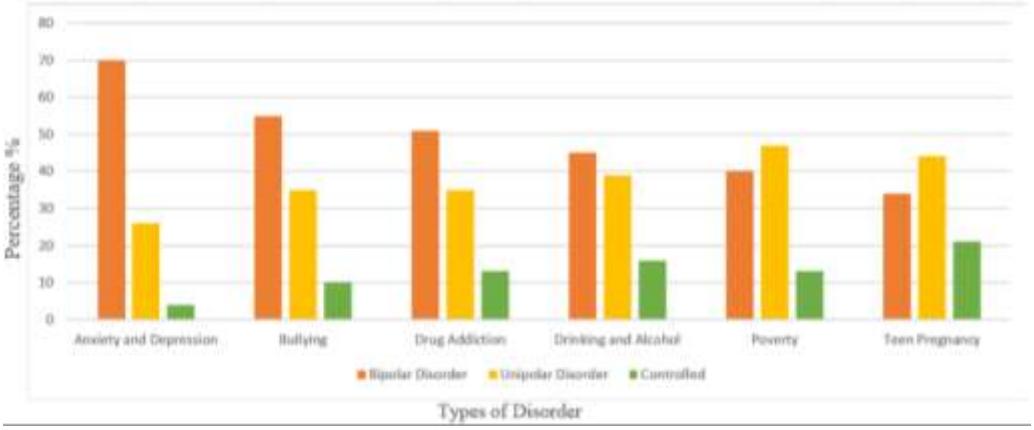

Figure 1. A survey of U.S. showing BD as a major problem in the young population in 2018 [20]

2.1. Dataset acquisition and preprocessing

Collect data on patients with and without bipolar disorder, including demographic information, medical history, and symptoms. Preprocess the data by removing duplicates, handling missing values, and transforming the data into a format suitable for modeling. The dataset taken in consideration is from the data source "theory of mind in remitted bipolar disorder" of participants. This is collected through MiniPons, which was based on interpersonal accuracy in recognition of dynamic nonverbal signals.

2.2. Dataset is divided in two parts training and test dataset

Split the data into two parts as shown in Figure 2, one for model training and other for testing the accuracy. The training set should be larger than the testing set. Training data (estimate the parameters for the machine learning models) and test data (evaluate how well the machine learning models work).

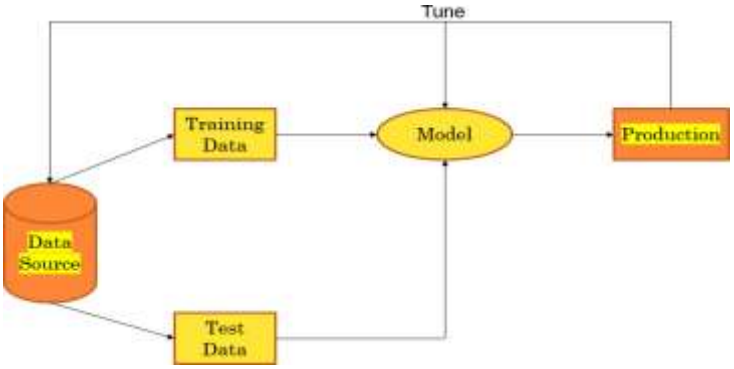

Figure 2. Training and test data

2.3. Combine the models

Combine the SVR and LR models into a hybrid model. One way to combine them is to use the output of the SVR as an input to the LR model. The research proposes LR and SVM for ensemble based on their performance on bipolar disorder dataset s. The most predicted class is indicated as the final predicted motel on the basis of maximum voting criteria by combining two models. The proposed algorithm for final predicted model is the combination of two.

2.4. Performance evaluation of the model

Evaluate the hybrid model on the testing set to determine its accuracy. Adjust the model's hyper parameters necessary to improve the model's accuracy. The evaluation of performance of model is based on performance matrices. Once the algorithms are applied, it is important to validate and undergo performance evaluation onto finally conclude which model is the best [15]. Performance evaluation is done by calculating recall, precision, F1-score, and accuracy. For each model and the model with highest value fits the best.

2.4.1. Deploy the model

Once the model has been evaluated and found to be accurate, deploy it for use in clinical settings to predict bipolar disorder in patients Hybrid-LSVR. We deploy a hybrid classification to calculate inputs in high dimension with accurate manner and to discover specific features.

3. RESULTS AND DISCUSSION

Once the algorithms are applied, it is important to validate and undergo performance evaluation to finally conclude which model is the best. Performance evaluation is done by calculating recall, precision, F1-score, and accuracy for each model, and the model with highest value fits the best. In machine learning, the problem of statistical classification uses confusion matrix, it is a special layout for data visualization for evaluating the performance of algorithms. Instances of a class is represented by each row while columns represent the instance of each predicted class each row of the matrix represents the instances of a true class while each column represents the instances in predicted class. The confusion matrix predicts four values true positive (TP), false positive (FP), false negative (FN), true negative (TN) in predicted and actual class, where the rows show the value of true classes and the column shows the value of predicted classes respectively as shown in Table 1. The rate of success can be calculated as:

$$r = \frac{TN+TP}{FP+FN}$$

It determines the sensitivity and specificity as:

- Sensitivity- what % of patients with BD were correctly identified.
- Specificity-what % of patients without BD were correctly identified.

Table 1. Confusion matrix

	Has bipolar disorder	Does not have bipolar disorder
Has bipolar disorder	(TP) 1	(FP) 8
Does not have bipolar disorder	(FN) 7	(TN) 84

TP – had BD and predicted with BD
FP – don't had BD but predicted with BD
FN – has BD and not predicted with BD
TN – don't has BD and not predicted with BD
True positive: 1
True negative: 84
False positive: 8
False negative: 7

The performance matrices used to evaluate the performance of machine learning and combined models as shown in Table 2:

1) Accuracy: accuracy helps to calculate the percentage of correct observations out of total observations.

$$Accuracy\ \% = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

2) Precision: precision helps to calculate the relevant number of data points from the test data.

$$Precision\ \% = \frac{TP}{TP+FP} * 100$$

3) Recall: recall helps to calculate the number of TP with respect to total number of observations.

$$Recall\ \% = \frac{TP}{TP+FN} * 100$$

4) F1-score: F1-score values provide us the harmonic mean value between recall and precision.

$$F1 - Score\ \% = \frac{Precision * Recall}{Precision + Recall} * 100$$

Table 2. Comparison table for accuracy, precision, F1-score, and recall

Model/matrices	Accuracy	Precision	Recall	F1-score
SVR	0.946428571	0.933035714	0.946428571	0.93877551
LR	0.160714286	0.127545249	0.160714286	0.121025584
LSVR	0.96	0.910909091	1	0.952380952

Data visualization are important tools for identifying a qualitative understanding of dataset which helps in exploring the data and extract important information or to identify patterns, outliners or corrupt data. In Python various libraries come with lots of different features that enables users to make customized, elegant and interactive plots. In this we have used Matplotlib for easy visualization of data. Each data set may use three types of visualization techniques: i) histogram, ii) heat correlation map, and iii) bar graph.

For our data visualization, a bar graph is used to explore the visual prediction of individual models LR and SVR to be compared with hybrid model-LSVR. The X-axis shows the different matrices of models and Y-axis shows their percentage. The final conclusion of the R2 score on the dataset as shown in Table 3. has provided the performances of the classification-based model [16]. The R2 score value is also known as r-squared or coefficient of determination. It is calculated by checking the amount of various attributes present in the prediction of data. The comparison of machine learning models as shown in Figures 3 predict that LSVR gives more accurate values as compared to individual models LR and SVM.

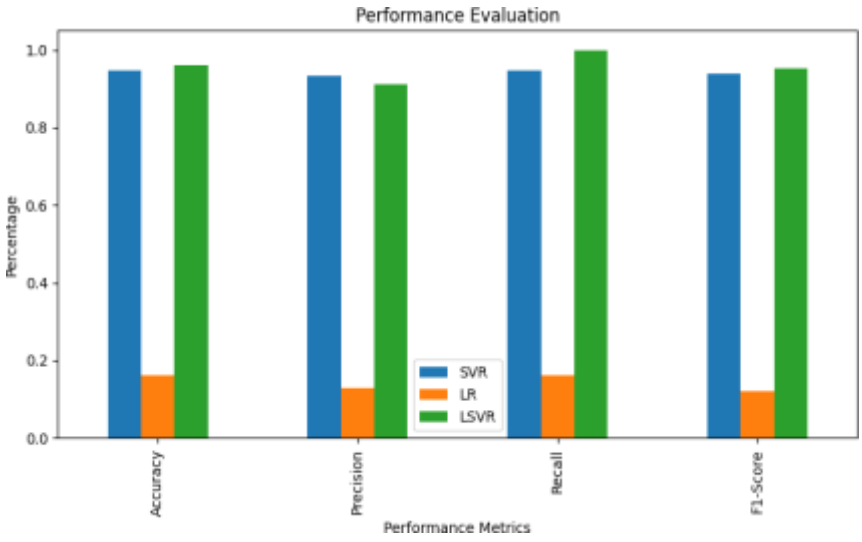

Figure 3. Bar graph of LR, SVR, and LSVR on performance matrices

Table 3. R2-score of LR, SVM, and LSVR

Model	R2-score
LSVR	0.958333333
LR	0.987654322
SVR	0.889765432

After data visualization and performance evaluation, this is concluded that the hybrid model LSVR is more accurately helping in diagnosis as compared to individual classification models. Figure 4 shows a line plot constructed to identifying the frequency of occurrence by the different models in the study. This reflects how the data has been varied according to the different models and further helps in making predictions to get better results.

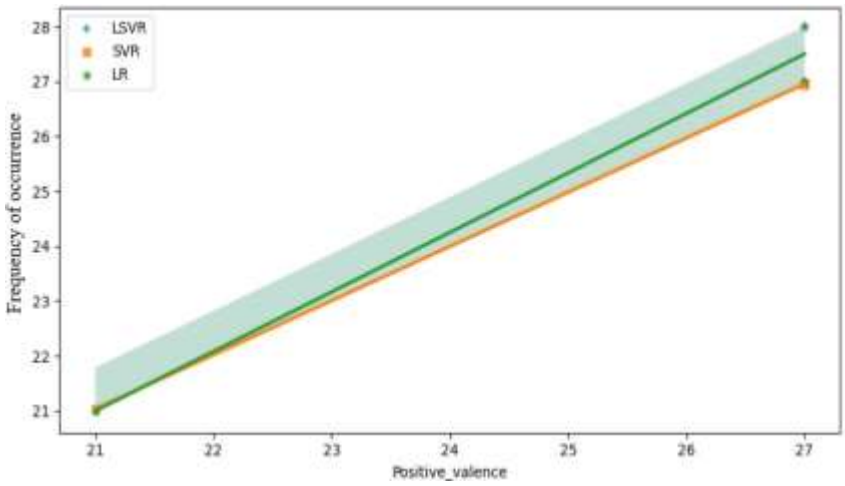

Figure 4. Line plot to check variance

4. CONCLUSION

In conclusion, the hybrid LSVR model can be an effective tool for predicting bipolar disorder in patients. By combining the strengths of LR and SVMs, the hybrid model can provide accurate and reliable predictions while minimizing the risk of false positives and false negatives. A hybrid classification approach has been designed to discover specific features and calculate high dimensional inputs in an accurate manner. Hybrid techniques like preprocessing, and feature selection are used to develop a hybrid LSVR model. Best features are selected using LR and SVM models. Therefore, it is important to carefully evaluate the performance of the model and adjust it as necessary to ensure optimal performance. The hybrid model has the potential to improve the early diagnosis and treatment of bipolar disorder, thereby improving patient outcomes and reducing the burden of this disorder on individuals and society as a whole.

REFERENCES

[1] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A framework for classifying online mental health-related communities with an interest in depression," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016, doi: 10.1109/JBHI.2016.2543741.

[2] L. Sivagnanam and N. K. Visalakshi, "Multimodal machine learning framework to detect the bipolar disorder," in *Advances in Parallel Computing*, no. 41, 2022, pp. 138–146.

[3] R. Saranya and S. Niraimathi, "Bd-Mdl: bipolar disorder detection using machine leanring and deep learning," *Journal of Pharmaceutical Negative Results*, vol. 13, pp. 5892–5905, 2022, doi: 10.47750/pnr.2022.13.S07.715.

[4] J. Chung and J. Teo, "Mental health prediction using machine learning: taxonomy, applications, and challenges," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–19, Jan. 2022, doi: 10.1155/2022/9970363.

[5] F. Ceccarelli and M. Mahmoud, "Multimodal temporal machine learning for bipolar disorder and depression recognition," *Pattern Analysis and Applications*, vol. 25, no. 3, pp. 493–504, 2022, doi: 10.1007/s10044-021-01001-y.

[6] J. Tomasik *et al.*, "A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data," *Translational Psychiatry*, vol. 11, no. 1, p. 41, Jan. 2021, doi: 10.1038/s41398-020-01181-x.

[7] K. A. Yashaswini and A. K. Saxena, "A novel predictive scheme for confirming state of bipolar disorder using recurrent decision tree," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 2, pp. 531–538, 2022, doi: 10.14569/IJACSA.2022.0130262.

[8] M. Á. Luján, A. M. Torres, A. L. Borja, J. L. Santos, and J. M. Sotos, "High-precise bipolar disorder detection by using radial basis functions based neural network," *Electronics*, vol. 11, no. 3, p. 343, Jan. 2022, doi: 10.3390/electronics11030343.

[9] D. Mahmood and D. M. Hussein, "AnalyzingNB, DT and NBTree intrusion detection algorithms," *Journal of Zankoy Sulaimani - Part A*, vol. 16, no. 1, pp. 69–76, 2014, doi: 10.17656/jzs.10285.

[10] B. Mahesh, "Machine learning algorithms - a review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/ART20203995.

[11] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 281, Dec. 2019, doi: 10.1186/s12911-019-1004-8.

[12] V. Rupapara, F. Rustam, W. Aljedaani, H. F. Shahzad, E. Lee, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model," *Scientific Reports*, vol. 12, no. 1, p. 1000, Jan. 2022, doi: 10.1038/s41598-022-04835-6.

[13] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," *Procedia Computer Science*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.

[14] J. Bektaş and T. Ibrikci, "Hybrid classification procedure using SVM with LR on two distinctive datasets," in *Proceedings of the 6th International Conference on Software and Computer Applications*, Feb. 2017, pp. 68–71, doi: 10.1145/3056662.3056717.

[15] A. Widodo and S. Handoyo, "The classification performance using logistic regression and support vector machine (SVM)," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 19, pp. 5184–5193, 2017.

[16] N. Agnihotri and S. K. Prasad, "Predicting the symptoms of bipolar disorder in patients using machine learning," in *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, Dec. 2021, pp. 697–702, doi: 10.1109/SMART52563.2021.9676247.
[17] N. Agnihotri and S. K. Prasad, "Bipolar disorder: early prediction and risk analysis using machine learning," *YMER*, vol. 21, no. 8, pp. 958–980, 2022, doi: 10.37896/YMER21.08/83.
[18] N. Agnihotri and S. Kumar Prasad, "Review on machine learning techniques to predict bipolar disorder," *Turkish Journal of Computer and Mathematics Education*, vol. 13, no. 02, pp. 195–206, 2022.
[19] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Measurement: Sensors*, vol. 25, p. 100587, Feb. 2023, doi: 10.1016/j.measen.2022.100587.
[20] P. J. Moore, M. A. Little, P. E. McSharry, J. R. Geddes, and G. M. Goodwin, "Forecasting depression in bipolar disorder," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2801–2807, 2012, doi: 10.1109/TBME.2012.2210715.
[21] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 23649–23685, 2022, doi: 10.1007/s11042-022-12648-y.
[22] J. M. Horowitz and N. Graf, "Most U.S. Teens see anxiety and depression as a major problem among their peers," *Pew Research Center*, 2019. https://www.pewresearch.org/2019/02/20/most-u-s-teens-see-anxiety-and-depression-as-a-major-problem-among-their-peers/ (accessed Mar. 10, 2021).
[23] S. K. Prasad and N. Agnihotri, "Implementing machine learning techniques to predict bipolar disorder," *International Journal of Medical Engineering and Informatics*, vol. 1, no. 1, p. 1, 2022, doi: 10.1504/ijmei.2022.10052205.
[24] N. Agnihotri and S. K. Prasad, "Accuracy enhancement with artificial neural networks for bipolar disorder prediction," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 3, pp. 1695–1702, 2023, doi: 10.11591/IJEECS.V32.I3.PP1695-1702.
[25] Y. Shao, Y. Cheng, S. Gottipati, and Q. Zeng-Treitler, "Outcome prediction for patients with bipolar disorder using prodromal and onset data," *Applied Sciences*, vol. 13, no. 3, p. 1552, Jan. 2023, doi: 10.3390/app13031552.
[26] G. Salazar de Pablo *et al.*, "Predicting bipolar disorder I/II in individuals at clinical high-risk: results from a systematic review," *Journal of Affective Disorders*, vol. 325, pp. 778–786, 2023, doi: 10.1016/j.jad.2023.01.045.

BIOGRAPHIES OF AUTHORS

Nisha Agnihotri is a Research Scholar in School of Computing Science and Engineering, Galgotias University, Greater Noida, UP, India. Her research area comprises about understanding machine learning and deep learning algorithms and its applications and currently working on diagnosing bipolar disorder with the help of machine learning at early stages which will help the clinicians and technicians in this field. She has published some research papers in this field in international conference and journals. She can be contacted at email: nisha.agnihotri_phd20@galgotiasuniversity.edu.in.

Sanjeev Kumar Prasad is a Professor in School of Computer Science and Engineering, Galgotias University, NCR Delhi, India. He has awarded a Ph.D. in Computer Science from Gurukula Kangri Vishwavidyalaya, Haridwar, Uttrakhand, India, in 2018. He has 19 years of rich teaching experience in the field of computer science. He has published more than 20 research papers in reputed international journals and conferences. His areas of interest include big data, internet of things, mobile ad-hoc networks, blockchain technology and machine learning. He has published book chapters in Springer and he is the reviewer of many renowned journals such as IEEE Access and Springer journals. He can be contacted at email: sanjeevkps2002@gmail.com.