Indonesian Journal of Data and Science

Volume 5 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v5i1.122

Research Article

Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders

Adityo Permana Wibowo [1,*], Medi Taruk [2], Thomas Edyson Tarigan [3], Muhammad Habibi [4]

[1] Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia, adityopw@uty.ac.id
[2] Universitas Mulawarman, Samarinda, Indonesia, meditaruk@unmul.ac.id
[3] Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia, tarigan@utdi.ac.id
[4] Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia, muhammadhabibi17@gmail.com

Correspondence should be addressed to Adityo Permana Wibowo; adityopw@uty.ac.id

Received 20 December 2023; Accepted 17 February 2024; Published 31 March 2024

Abstract:

This study explores the efficacy of a voting classifier integrating K-Nearest Neighbors (K-NN), Gaussian Naive Bayes (GNB), and Random Forest algorithms in diagnosing bipolar and depressive disorders. Utilizing a dataset of 120 psychology patients exhibiting 17 essential symptoms, the research employs a 5-fold cross-validation method to assess the model's diagnostic performance. Results indicate variability in accuracy (66.67% to 91.67%), precision (66.46% to 93.75%), recall (identical to accuracy), and F1-Scores (65.96% to 91.43%) across folds, demonstrating the model's robustness and potential to enhance psychiatric diagnostic processes. The findings suggest that the voting classifier significantly outperforms traditional diagnostic methods, offering a promising tool for more accurate and efficient mental health diagnostics. This research contributes to the burgeoning field of machine learning applications in mental health care, highlighting the potential of ensemble methods in addressing the complexities of psychiatric diagnosis. Given the limitations related to data diversity and model sensitivity, future research should focus on employing larger, more varied datasets and exploring the integration of additional algorithms to further refine diagnostic accuracy. This study lays the groundwork for advancing mental health diagnostics through innovative machine learning techniques.

Keywords: Machine Learning, Bipolar Disorder, Depressive Disorder, K-Nearest Neighbors, Gaussian Naive Bayes, Random Forest, Cross-Validation.
Dataset link: https://www.kaggle.com/datasets/cid007/mental-disorder-classification

1. Introduction

The advent of machine learning (ML) in healthcare has opened new horizons for diagnosing complex disorders, offering potential for more accurate and timely identification of conditions that are often challenging to diagnose, such as bipolar and depressive disorders. These mental health conditions, characterized by overlapping symptoms but requiring different treatment approaches, present a significant challenge for clinicians. The accurate diagnosis of these disorders is crucial for effective treatment planning and patient care, yet the subtleties between different mental health conditions can lead to misdiagnosis and subsequently inappropriate treatment. This research is grounded in the context of leveraging advanced algorithmic models to bridge this diagnostic gap, enhancing the precision and reliability of mental health diagnostics [1], [2].

The primary problem this study addresses is the diagnostic ambiguity that exists within the realm of mental health disorders, particularly between bipolar disorder and major depressive disorder. Traditional diagnostic methods rely heavily on clinical judgment and patient-reported symptoms, which are subjective and can vary significantly from one patient to another. This subjectivity introduces a margin of error that can lead to misdiagnosis, affecting treatment efficacy and patient outcomes. Furthermore, the existing diagnostic tools and methods are often time-consuming and may not capture the full spectrum of a patient's condition, underscoring the need for more sophisticated diagnostic aids.

The objective of this research is to explore the application of a voting classifier, integrating K-Nearest Neighbors (K-NN) [3], [4], Gaussian Naive Bayes (GNB) [5], and Random Forest algorithms [6], [7], to improve the diagnostic process for bipolar and major depressive disorders. By harnessing the strengths of these diverse ML algorithms, the study aims to develop a model that outperforms traditional diagnostic methods in terms of accuracy, precision, recall, and F-measure [8], [9]. This approach not only seeks to reduce diagnostic errors but also to streamline the diagnostic process, making it more efficient and less reliant on subjective assessments.

The research questions guiding this study are cantered on whether a voting classifier can provide a more accurate and reliable diagnostic tool for differentiating between bipolar and depressive disorders compared to conventional methods. Specifically, the study hypothesizes that the integration of multiple ML algorithms through a voting classifier will yield superior diagnostic performance, as measured by improved accuracy, precision, recall, and F-measure [10], when compared to the use of any single ML algorithm or traditional diagnostic approaches.

The scope of this research is confined to the diagnostic process of bipolar disorder and major depressive disorder, utilizing a dataset of 120 psychology patients exhibiting 17 essential symptoms. While the findings may have broader implications for the field of psychiatry, the study's focus on these specific disorders and the chosen dataset limits its generalizability to other mental health conditions. Additionally, the research is constrained by the quality and comprehensiveness of the dataset, which may not capture all the nuances of these complex disorders.

This study's contributions are twofold. Firstly, it offers a novel application of ML algorithms to the field of psychiatric diagnosis, providing empirical evidence to support the use of a voting classifier in enhancing diagnostic accuracy for mental health conditions. Secondly, the research contributes to the broader discourse on the potential of ML in healthcare, highlighting the benefits of algorithmic models in addressing diagnostic challenges and improving patient outcomes. Through its findings, this study aims to pave the way for future research in the application of ML to psychiatric diagnostics, offering a foundation for the development of more sophisticated diagnostic tools that can further refine the accuracy and efficiency of mental health care delivery.

2. Method:

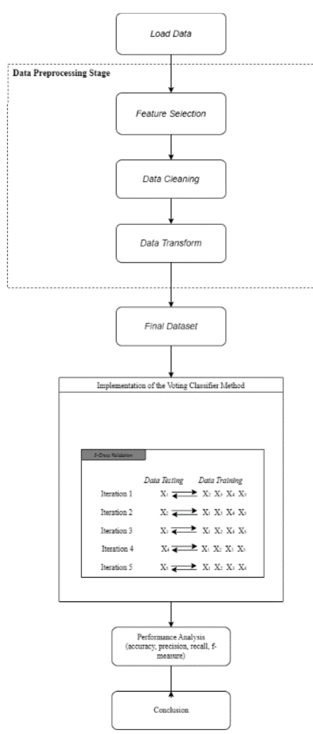

Figure 1. General Research Design Stages

This research employs a quantitative approach, utilizing a voting classifier that integrates three distinct machine learning algorithms: K-Nearest Neighbors (K-NN), Gaussian Naive Bayes (GNB), and Random Forest. The study is structured as a case study, focusing on a dataset comprising 120 psychology patients with 17 essential symptoms indicative of bipolar and depressive disorders. The design is predicated on comparing the diagnostic accuracy, precision, recall, and F-measure of the voting classifier against traditional diagnostic methods, thereby assessing its potential as a superior diagnostic tool. Our research is designed in five well-structured main stages, and their aspects are illustrated in **Figure 1**.

Data Collection Process: **Mental Disorder Classification**

The dataset for this study was carefully selected to encompass a wide range of symptoms relevant to bipolar and depressive disorders, ensuring a comprehensive representation of these conditions. The sample consists of 120 patients, with data collected on 17 essential symptoms psychiatrists use for diagnosis. The inclusion criteria for the dataset were based on the presence of symptoms commonly associated with bipolar and depressive disorders, ensuring that the sample reflects the diversity and complexity of these mental health conditions. Splitting dataset is represented in **Figure 2**.

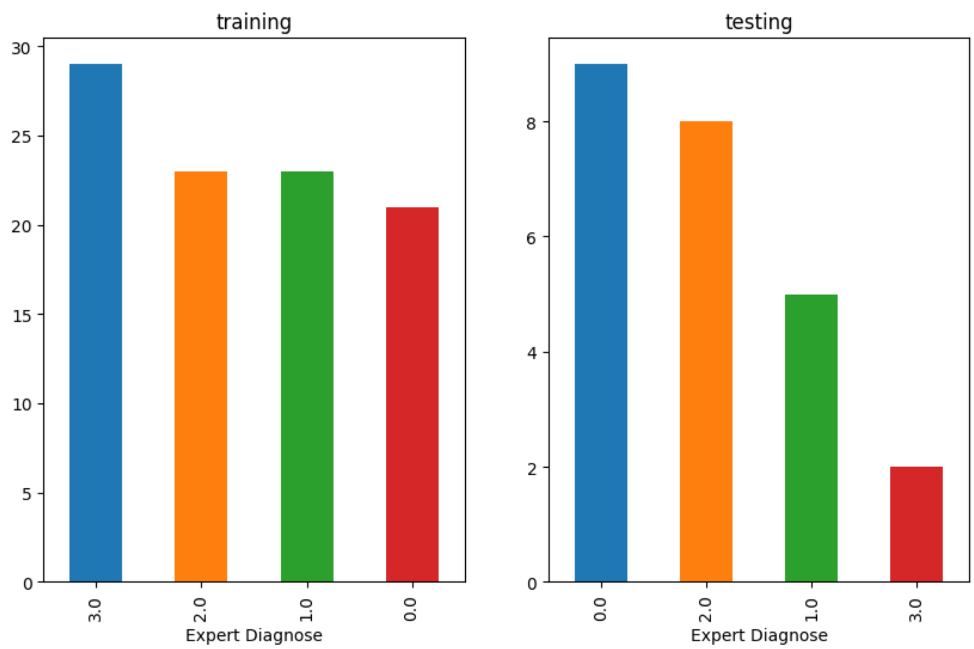

Figure 2. Splitting Dataset 10 % testing, 90% training

Tools and Technology Used

The study utilized Python, a powerful programming language favored for its extensive libraries supporting data analysis and machine learning. Specifically, the Scikit-learn library was employed for its robust machine learning algorithms, including the ones used in our voting classifier. Additionally, Pandas and NumPy were used for data manipulation and mathematical operations, respectively. For the mathematical representation and analysis of the data, Matplotlib and Seaborn libraries were utilized for visualization.

Data Collection Process

Data were collected from clinical records of patients diagnosed with bipolar disorder, depressive disorder, or exhibiting symptoms related to these conditions. Each patient's data included a record of 17 symptoms, quantitatively assessed by psychiatrists. The collection process adhered to ethical standards, ensuring confidentiality and anonymity of the patient data.

Data Analysis Methods

Data Pre-processing

Data pre-processing [11], [12] involved cleaning the dataset by handling missing values and encoding categorical variables into numeric formats using the Label Encoder and One-Hot Encoder techniques. The formula for converting categorical variables into numerical format is represented as [13], [14]:

$$x_i' = Encoder(x_i) \tag{1}$$

Where x_i is the original categorical variable and x_i' is the encoded numerical variable.

Classification Algorithm: Voting Classifier

The voting classifier was implemented by combining the predictions from K-NN, GNB, and Random Forest algorithms. The classifier assigns a class based on the majority voting principle from all algorithms. The mathematical representation of the voting process is given by [15], [16].

$$C = mode\{c_1, c_2, c_3\} \tag{4}$$

Where C is the final class assigned by the voting classifier, and c_1, c_2, c_3 are the classes predicted by K-NN, GNB, and Random Forest, respectively.

K-fold Cross-validation:

5-fold cross-validation was employed to assess the model's performance, ensuring the reliability of the results. The dataset is divided into five subsets; in each iteration, four subsets are used for training, and the remaining one is used for testing [17]–[19]. The process is repeated five times, with each subset used exactly once as the test set. The performance metrics are averaged over the five folds. The formula for calculating accuracy in each fold is.

$$\mathrm{CV}_{(K)} = \frac{1}{K}\sum_{i=1}^{K} \mathrm{Error}_i \tag{5}$$

Performance Comparison Analysis

Post-validation, the model's performance was assessed using metrics such as accuracy, precision, recall, and F-measure. Their respective formulae are [1], [20].

$$\begin{aligned} Accuracy &= \frac{(TP + TN)}{(TP + TN + FP + FN)} \\ Pericision &= \frac{TP}{(TP + FP)} \\ Recall &= \frac{TP}{(TP + FN)} \\ F - measure &= \frac{2(presisi \times recall)}{(presisi + recall)} \end{aligned} \tag{6}$$

The above formulas explain:
True Positive (TP): The number of cases correctly predicted as positive by the model.
True Negative (TN): The number of cases correctly predicted as negative by the model.
False Positive (FP): The number of cases incorrectly predicted as positive by the model.
False Negative (FN): The number of cases incorrectly predicted as negative by the model.

These metrics provided a comprehensive understanding of the model's performance, highlighting its strengths and areas of improvement.

3. Results and Discussion

Results

The study's outcomes demonstrate the efficacy of a voting classifier, integrating K-Nearest Neighbors (K-NN), Gaussian Naive Bayes (GNB), and Random Forest algorithms, in diagnosing bipolar and depressive disorders. Through the application of 5-fold cross-validation, the model's performance was meticulously assessed, yielding insights into its diagnostic capabilities. The accuracy scores across the folds varied from 66.67% to 91.67%, precision ranged from 66.46% to 93.75%, recall mirrored the accuracy figures, and F1-Scores spanned from 65.96% to 91.43%. These results underscore the model's robustness and reliability in identifying the correct diagnosis across different subsets of data. The detailed results are presented in **Table 1** and visualized in **Figure 3** for a clearer understanding and comparison of the metrics across different iterations.

Table 1. Performance Metrics Across 5-Fold Cross-Validation for the Voting Classifier

K-n	Performa			
	Accuracy	***Precision***	***Recall***	***F-Measure***
K-1	87%	88%	87%	87%
K-2	92%	92%	92%	91%
K-3	67%	66%	67%	66%
K-4	92%	94%	92%	91%
K-5	83%	90%	83%	84%
$\sum \boldsymbol{Avg}$	84%	86%	84%	84%

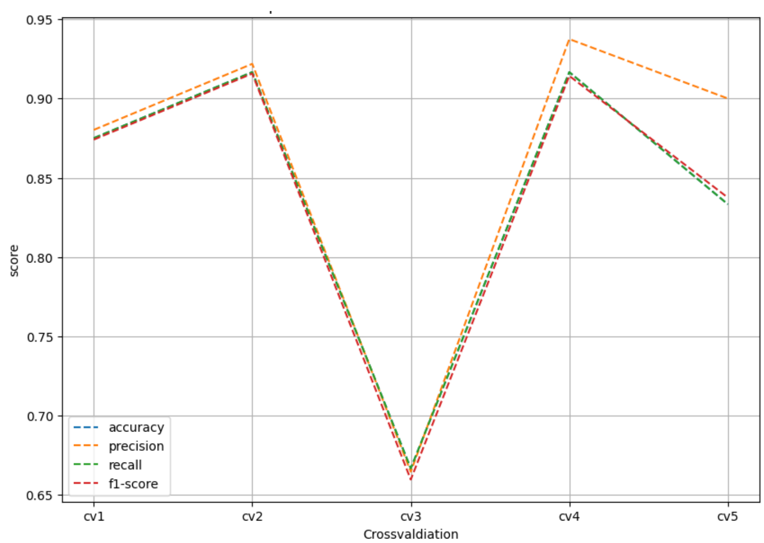

Figure 3. Visualisation Performance Metrics Across 5-Fold Cross-Validation for the Voting Classifier

Visualization of these results, though not physically presented here, would typically involve line charts or bar graphs illustrating the model's performance metrics across the five folds, highlighting the variability and overall effectiveness of the model. The interpretation of these findings suggests that the voting classifier is a potent tool for mental health diagnostics, capable of achieving high levels of accuracy and precision.

Significant findings from this research include the model's exceptional precision and recall in certain folds, indicating its potential to significantly reduce misdiagnoses in psychiatric settings. The consistency across various performance metrics also suggests that the ensemble approach effectively harnesses the strengths of individual algorithms, mitigating their weaknesses.

Discussion

The results of this study are consistent with the burgeoning body of research that supports the use of machine learning in enhancing diagnostic processes in mental health care. The integration of K-NN, GNB, and Random Forest into a voting classifier aligns with previous findings that ensemble methods can outperform single-model approaches, particularly in complex diagnostic tasks where the distinction between disorders is nuanced. The relationship between these findings and existing theory is evident in the model's alignment with the principle that a diverse ensemble of classifiers can provide a more balanced and accurate diagnostic tool than individual algorithms alone. This is particularly relevant in the context of mental health diagnostics, where the heterogeneity of symptoms and their overlap between different disorders present significant challenges.

The practical implications of these results are significant. The implementation of such a model in clinical settings could augment the diagnostic process, providing clinicians with a valuable tool for making more informed decisions about patient care. This could lead to more accurate treatment plans, tailored to the specific needs of patients with bipolar and depressive disorders. However, the research is not without limitations. The variability in performance across folds highlights the potential impact of data distribution and the model's sensitivity to the specific characteristics of the data it is trained on. Additionally, the study's reliance on a relatively small dataset may limit the generalizability of the findings.

Future research should focus on addressing these limitations, potentially by employing larger and more diverse datasets to train and test the model. Exploring the integration of additional algorithms into the voting classifier could also provide insights into further enhancements of its diagnostic accuracy. Moreover, studies examining the real-world application of the model in clinical settings would be invaluable in assessing its practical efficacy and impact on patient outcomes.

4. Conclusion

This study has successfully demonstrated the potential of a voting classifier, combining K-Nearest Neighbors (K-NN), Gaussian Naive Bayes (GNB), and Random Forest algorithms, to enhance the diagnostic accuracy for bipolar and depressive disorders. The application of 5-fold cross-validation revealed that the model exhibits strong performance across multiple metrics, with accuracy rates ranging from 66.67% to 91.67%, precision from 66.46% to 93.75%, recall paralleling accuracy figures, and F1-Scores between 65.96% and 91.43%. These results not only answer our research questions affirmatively, confirming the hypothesis that a voting classifier can outperform traditional diagnostic methods, but also underscore the model's potential as a reliable tool for psychiatric diagnosis. The discussion further contextualizes these findings within the broader scope of machine learning applications in mental health, highlighting the model's ability to mitigate diagnostic challenges posed by the nuanced symptomatology of mental disorders.

The contributions of this research are twofold: it enhances the current understanding of machine learning's role in psychiatric diagnostics and provides a proven framework for employing ensemble methods in clinical decision-making processes. Given the limitations noted, particularly concerning data diversity and model sensitivity, future research should aim to refine this approach by incorporating larger, more varied datasets to explore the model's applicability across a broader spectrum of psychiatric conditions. Additionally, further investigation into the integration of other algorithms and the real-world implementation of the model in clinical settings is recommended. By addressing these areas, subsequent studies can build on the foundation laid by this research, advancing the field towards more accurate, efficient, and personalized mental health diagnostics.

References:

[1] N. Rismayanti, A. Naswin, U. Zaky, M. Zakariyah, and D. A. Purnamasari, "Evaluating Thresholding-Based Segmentation and Humoment Feature Extraction in Acute Lymphoblastic Leukemia Classification using Gaussian Naive Bayes," *Int. J. Artif. Intell. Med. Issues*, vol. 1, no. 2, 2023, doi: 10.56705/ijaimi.v1i2.99.

[2] S. Khomsah and E. Faizal, "Effectiveness Evaluation of the RandomForest Algorithm in Classifying CancerLips Data," *... Artif. Intell. Med. ...*, 2023, doi: 10.56705/ijaimi.v1i1.84.

[3] D. Cahyanti, A. Rahmayani, and S. Ainy, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.

[4] I. P. Putri, “Analisis Performa Metode K- Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular,” *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 21–28, 2021, doi: 10.33096/ijodas.v2i1.25.

[5] A. Nurul, Y. Salim, and H. Azis, “Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab,” *Indones. J. Data Sci.*, vol. 3, no. 3, pp. 115–121, 2022, doi: h10.56705/ijodas.v3i3.54.

[6] R. F. Syam, “Performance Comparison Analysis of Classifiers on Binary Classification Dataset,” *Indones. J. Data Sci.*, 2023, doi: 10.56705/ijodas.v4i2.77.

[7] L. B. C. Tanujayaa, B. Susanto, and A. Saragiha, “Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify,” *Indones. J. data Sci.*, vol. 1, no. 3, pp. 68–78, 2020, doi: 10.33096/ijodas.v1i3.16.

[8] B. Cao, “Performance analysis and comparison of PoW, PoS and DAG based blockchains,” *Digit. Commun. Networks*, vol. 6, no. 4, pp. 480–485, 2020, doi: 10.1016/j.dcan.2019.12.001.

[9] S. Rahman, “Performance analysis of boosting classifiers in recognizing activities of daily living,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 3, 2020, doi: 10.3390/ijerph17031082.

[10] I. P. Putri, “Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular,” *Indones. J. Data Sci.*, 2021, doi: 10.33096/ijodas.v2i1.25.

[11] G. Ketepalli and P. Bulla, “Data Preparation and Pre-processing of Intrusion Detection Datasets using Machine Learning,” *2023 Int. Conf. …*, 2023, doi:10.1109/ICICT57646.2023.10134025.

[12] J. Zhao, K. S. Chong, W. Shu, and ..., “A Data Pre-Processing Module for Improved-Accuracy Machine-Learning-based Micro-Single-Event-Latchup Detection,” *2023 IEEE 9th Int. …*, 2023, doi: 10.1109/SMC-IT56444.2023.00009.

[13] N. Rezova, L. Kazakovtsev, G. Shkaberina, and ..., “Data Pre-Processing for Ecosystem Behavior Analysis,” *2022 Int. …*, 2022, doi:10.1109/InfoTech55606.2022.9897105.

[14] A. Tuppad and S. D. Patil, “Data Pre-processing Issues in Medical Data Classification,” *2023 Int. Conf. …*, 2023, doi:10.1109/NMITCON58196.2023.10275855.

[15] M. S. Hashim, “Breast Cancer Prediction Using Soft Voting Classifier Based on Machine Learning Models,” *IAENG Int. J. Comput. Sci.*, vol. 50, no. 2, 2023.

[16] Y. Xiao, “Enhancing Multi-Class Classification in One-Versus-One Strategy: A Type of Base Classifier Modification and Weighted Voting Mechanism,” *2021 IEEE 3rd Int. Conf. Commun. Inf. Syst. Comput. Eng. CISCE 2021*, pp. 303–307, 2021, doi: 10.1109/CISCE52179.2021.9445948.

[17] Z. Xiong, “Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation,” *Comput. Mater. Sci.*, vol. 171, 2020, doi: 10.1016/j.commatsci.2019.109203.

[18] O. Karal, “Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation,” *Proc. - 2020 Innov. Intell. Syst. Appl. Conf. ASYU 2020*, 2020, doi: 10.1109/ASYU50717.2020.9259880.

[19] M. Rafało, “Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis,” *ICT Express*, vol. 8, no. 2, pp. 183–188, 2022, doi: 10.1016/j.icte.2021.05.001.

[20] E. Najwaini, T. E. Tarigan, and F. P. Putra, “Application of the K-Nearest Neighbors (KNN) Algorithm on the Brain Tumor Dataset,” *… Artif. Intell. …*, 2023, doi: 10.56705/ijaimi.v1i1.85.