Jurnal Riset Sains dan Teknologi
Volume 9 No. 2 September 2025, 101-110

e-ISSN: 2549-9750
p-ISSN: 2579-9118

Leveraging Linear Discriminant Analysis for Early Mental Health Disorder Identification

Pemanfaatan Analisis Diskriminan Linier untuk Identifikasi Dini Gangguan Kesehatan Mental

Deden Iwan Setiawan[1], Marselina Endah Hiswati[2*], Sriwidodo[3], Mohammad Diqi[4], Luh Putu Erikawati[5], Rahayu Cahya Ariani[6]
[1,2,4,5]Universitas Respati Yogyakarta, Yogyakarta, Indonesia
[3,6]Universitas Respati Indonesia, Jakarta, Indonesia
Jl. Bambu Apus I No.3, Kec. Cipayung, Kota Jakarta Timur, DKI Jakarta 13890, Indonesia

*Corresponding author: marsel.endah@respati.ac.id

DOI:
10.30595/jrst.v9i2.23053

Histori Artikel:

Diajukan:
10/07/2024

Diterima:
28/05/2025

Diterbitkan:
08/09/2025

ABSTRACT

Mental health disorders pose a significant global challenge, with early identification playing a crucial role in effective intervention and treatment. However, existing diagnostic methods often rely on subjective assessments, leading to potential misdiagnosis and delayed treatment. This study aims to address these limitations by exploring the application of Linear Discriminant Analysis (LDA) for early identification of mental health disorders, specifically focusing on Bipolar Type-1, Bipolar Type-2, Depression, and Normal conditions. Utilizing a publicly available dataset from Kaggle comprising 120 records and 17 attributes, this study applies LDA to classify mental health conditions. The preprocessing steps included handling missing values, encoding categorical data, and normalizing the dataset to enhance model performance. The classification performance was evaluated using a confusion matrix and classification report metrics, demonstrating high accuracy, precision, recall, and F1-scores, particularly for Bipolar Type-1 and Depression, while slightly lower for Bipolar Type-2 and Normal conditions. The novelty of this research lies in the application of LDA to a nuanced mental health dataset, emphasizing its potential as a computational diagnostic tool to complement traditional assessment methods. However, findings suggest that larger, more diverse datasets and the incorporation of objective clinical assessments are necessary to further improve classification accuracy. This study underscores the potential of LDA as a practical and interpretable approach for early mental health diagnosis, providing a foundation for future research to enhance its robustness and clinical applicability.

Keywords: *Linear Discriminant Analysis; Mental Health Diagnosis; Classification Model; Dataset Preprocessing; Machine Learning*

ABSTRAK

Gangguan kesehatan mental merupakan tantangan global yang signifikan, dengan identifikasi dini memainkan peran krusial dalam intervensi dan pengobatan yang efektif. Namun, metode diagnostik yang ada seringkali bergantung pada penilaian subjektif, yang dapat menyebabkan diagnosis yang salah dan penundaan pengobatan. Penelitian ini bertujuan untuk mengatasi keterbatasan tersebut dengan mengeksplorasi penerapan Analisis Diskriminan Linier (LDA) untuk identifikasi dini gangguan kesehatan mental, khususnya pada Bipolar Tipe-1, Bipolar Tipe-2, Depresi, dan kondisi normal. Menggunakan dataset yang tersedia secara publik dari Kaggle yang terdiri dari 120 catatan dan 17 atribut, penelitian ini menerapkan LDA untuk mengklasifikasikan kondisi kesehatan mental. Langkah-langkah prapemrosesan meliputi penanganan nilai yang hilang, pengkodean data kategorikal, dan normalisasi dataset untuk meningkatkan kinerja model. Kinerja klasifikasi dievaluasi menggunakan matriks kebingungan dan metrik laporan klasifikasi, menunjukkan akurasi, presisi, recall, dan skor F1 yang tinggi, terutama untuk Bipolar Tipe-1 dan Depresi, sementara sedikit lebih rendah untuk Bipolar Tipe-2 dan kondisi Normal. Keunikan penelitian ini terletak pada penerapan Analisis Diskriminan Linier (LDA) pada dataset kesehatan mental yang kompleks, menyoroti potensinya sebagai alat diagnostik komputasional untuk melengkapi metode penilaian tradisional. Namun, temuan menunjukkan bahwa dataset yang lebih besar dan beragam, serta integrasi penilaian klinis objektif, diperlukan untuk meningkatkan akurasi klasifikasi. Studi ini menyoroti potensi LDA sebagai pendekatan praktis dan dapat diinterpretasikan untuk diagnosis dini kesehatan mental, memberikan landasan bagi penelitian masa depan untuk meningkatkan ketahanan dan penerapan klinisnya.

Kata Kunci: Analisis Diskriminan Linier; Diagnosis Kesehatan Mental; Model Klasifikasi; Pra-pemrosesan Data; Pembelajaran Mesin

1. INTRODUCTION

Mental disorders pose a severe challenge to public health that requires special attention (Gao et al., 2020). In recent years, awareness of the importance of early identification of mental disorders has increased due to the significant impact they can have on the mental health of individuals and society as a whole (William & Suhartono, 2021).

Mental disorders can affect various aspects of an individual's life, including interpersonal relationships, employment, and overall quality of life (Srivastava et al., 2018). Early identification of mental disorders has a positive impact on providing more effective and timely support to individuals who may be experiencing mental health issues (Christensen et al., 2020).

Swift recognition of signs and symptoms can open the door to more effective interventions, reduce the risk of complications, and improve prospects for recovery (McDaid et al., 2019). Furthermore, mental disorders also have a significant economic impact, both through healthcare costs and their effects on worker productivity (Kersemaekers et al., 2018).

Early identification is not only beneficial for the individuals affected but can also reduce the economic burden arising from long-term care or loss of productivity (Samartzis & Talias, 2019). Early identification techniques for mental disorders have been explored in previous research.

One approach is the use of mechanical capacity models to detect or predict internal mental health (Reddy & Satvika, 2023). Another technique is "brain fingerprinting," which uses functional connectivity profiles to determine individual uniqueness of brain activity or structure (Hermens, 2023).

Social media posts and machine learning have also been used for mental assessment, although the availability of sufficient data remains a challenge (Dalal et al., 2023). Additionally, a framework based on visibility graphs has been proposed to recognize mental disorders using electroencephalogram (EEG) data, showing promising results in improving recognition accuracy (Zhang et al., 2023).

These techniques offer advantages such as early detection and intervention, which can help prevent the progression of mental health problems. Previous studies have compared the effectiveness of various techniques for early identification of mental disorders.

Artificial intelligence (AI) and machine learning (ML) approaches have shown promise in detecting early symptoms of mental health issues with optimized accuracy. Additionally, the use of machine learning models on speech sounds is a promising pathway for improving the diagnosis of speech-based diseases (Reddy & Satvika, 2023).

However, it is essential to note that the effectiveness of these techniques may depend on factors such as culture, language, and content of speech, as well as gender, age, and accent (Lavanya et al., 2023). While there is no indication of a superior technique for detecting early symptoms of mental disorders in the abstracts provided, the use of AI and ML approaches and the analysis of speech signals have shown potential in this area of research.

Linear Discriminant Analysis (LDA) is used for dimension reduction and classification. It is widely studied in statistics, machine learning, and pattern recognition. LDA is designed to find an optimal transformation to extract discriminant features that characterize different classes of objects (Zhao et al., 2020).

LDA can also be viewed as a promising dimensionality reduction technique, especially when there is a group structure in the data (Trendafilov & Gallo, 2021). LDA is used in the early identification of mental disorders by reducing the dimensions of data interpretation, making it easier to analyze grouped data and facilitating dimensionality reduction (Sudibyo et al., 2020).

The critical aspects of utilizing LDA in the context of early identification of mental disorders are its interpretability and feature selection capabilities. LDA allows for the analysis of separability between classes in a multidimensional feature space, providing insights into the results of machine learning models (Reddy & Satvika, 2023).

It helps identify the essential features that distinguish between different groups, such as mean node strength, clustering coefficient, and the number of edges (Андреев et al., 2023). This feature selection approach ensures that the identified features are relevant and contribute to the accuracy of the classification model (Bai, 2023).

By focusing on interpretable features, LDA improves the quality of the identification process by providing insights into the underlying patterns and relationships between variables (Dalal et al., 2023). Additionally, LDA can achieve high accuracy while ensuring the interpretability of the results, making it a valuable tool for early identification of mental disorders (Zhang et al., 2023).

The research problem identified in this study is the limitations or deficiencies in the quality of Early Identification of Mental Disorders (Christensen et al., 2020). In this context, there are challenges related to the methods and approaches currently used to recognize the signs and symptoms of mental disorders at an early stage (Csillag et al., 2017).

These limitations may include low accuracy, speed, or sensitivity in the identification process, which, in turn, can hinder efforts to provide timely support to individuals in need (Silove, 2021). This research aims to address these limitations and enhance the quality of Early Identification of Mental Disorders through a more innovative and practical approach (Nunes et al., 2019).

This research aims to utilize LDA as an analytical tool to support the early identification of mental disorders. By focusing on this method, the study aims to develop a more sophisticated approach to recognizing patterns and characteristics associated with mental disorders at an early stage.

By utilizing LDA, it is anticipated that this research can improve the accuracy and speed of identification, helping to mitigate limitations in existing early identification approaches. Thus, the study seeks to provide a foundation for enhancements in the effort to identify mental disorders, ultimately enhancing the effectiveness of interventions and support for individuals in need of early mental health attention.

This research makes significant contributions both in the scientific and practical domains. Scientifically, the utilization of LDA to support the early identification of mental disorders provides new insights and a profound understanding of patterns related to mental health conditions.

These findings can serve as a foundation for further research in this field, opening opportunities for further development in early identification methods. On a practical level, this research has positive implications for enhancing the quality of mental health services by providing a more effective tool in the early identification process.

It is hoped that the results of this study can assist healthcare providers in offering earlier support to individuals who may require mental health attention, leading to improved prognosis and quality of life for them.

2. RESEARCH METHOD

Figure 1 presents the research workflow, outlining the sequential steps taken in this study, from data collection to result interpretation.

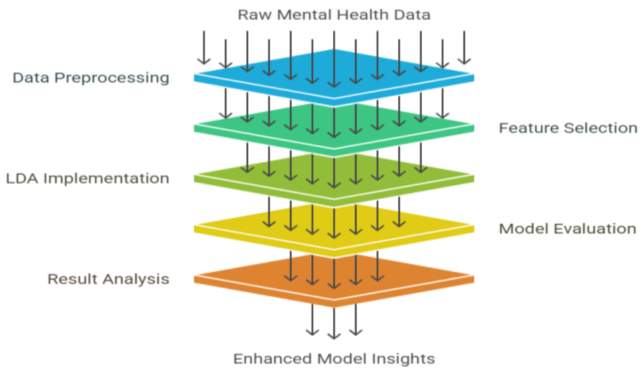

Figure 1. Research flow diagram

2.1 Dataset

The dataset used for this study is sourced from Kaggle and contains data relevant to mental health assessment. It comprises 18 columns and 120 records, providing a comprehensive overview of various mental health attributes.

The dataset includes 17 attributes: Sadness, Euphoric, Exhausted, Sleep disorder, Mood Swing, Suicidal thoughts, Anorexia, Authority Respect, Try-Explanation, Aggressive Response, Ignore & Move-On, Nervous Breakdown, Admit Mistakes, Overthinking, Sexual Activity, Concentration, and Optimism.

Each attribute represents a specific aspect of mental health, measured through various means to capture the complexity of mental health conditions. The target variable in the dataset classifies the mental health condition of individuals into four categories: Bipolar Type-2, Depression, Bipolar Type-1, and Normal.

These classifications are used to train and evaluate the LDA model for the early identification of mental health disorders. The structured nature of the dataset, with well-defined attributes and an explicit target variable, makes it suitable for applying LDA to differentiate between the various mental health conditions effectively.

2.2 Preprocessing

Data preprocessing involved several crucial steps to prepare the dataset for analysis. Firstly, handling missing values was essential, as any gaps in the data could significantly impact the model's performance. Missing values were imputed using appropriate statistical methods or by removing incomplete records. Secondly, categorical data encoding was necessary since many attributes were categorical.

Techniques such as one-hot encoding or label encoding were employed to convert categorical variables into a numerical format suitable for the LDA model. Lastly, the data was normalized or standardized to ensure that all features contributed equally to the analysis. This step involved scaling the data to a specific range or standardizing it to have a mean of zero and a standard deviation of one, thus improving the model's accuracy and convergence.

2.3 Linear Discriminant Analysis (LDA)

LDA is a statistical method used to find a linear combination of features that best separates two or more classes of objects or events. It is commonly utilized in classification and dimensionality reduction. The main objective of LDA is to maximize the separation between different classes while minimizing the variance within each class by finding a linear projection that maximizes the ratio of between-class variance to within-class variance.

In classification, LDA identifies the projection that best separates different classes, such as finding the axis that maximizes the distance between the means of two classes while minimizing the spread within each class. Additionally, LDA can reduce the dimensionality of the data by projecting it into a lower-dimensional space that maintains class separation.

LDA involves several key steps, each incorporating specific mathematical calculations to achieve the desired data transformation for classification or dimensionality reduction. Below are the detailed steps in LDA, along with the associated mathematical equations:

1. Calculate Class Means and Overall Mean

First, compute the mean vector for each class k. For a class k with n_k samples, the mean vector μ_k is calculated in Equation 1.

$$\mu_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i^{(k)} \qquad (1)$$

where $x_i^{(k)}$ represents the feature vector of the i-th sample in class k. Next, compute the overall mean vector μ using all N samples, as Equation 2.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2)$$

2. Compute Within-Class Scatter Matrix S_W

The within-class scatter matrix measures the spread of samples within each class. It is computed in Equation 3 by summing the scatter matrices of all classes.

$$S_W = \sum_{k=1}^{c}\sum_{i=1}^{n_k}\left(x_i^{(k)} - \mu_k\right)\left(x_i^{(k)} - \mu_k\right)^T \qquad (3)$$

Here, c is the number of classes, and the term $\left(x_i^{(k)} - \mu_k\right)\left(x_i^{(k)} - \mu_k\right)^T$ represents the outer product of the deviation of each sample from the class mean.

3. Compute Between-Class Scatter Matrix S_B

The between-class scatter matrix captures the dispersion of class means relative to the overall mean.

$$S_B = \sum_{k=1}^{c} n_k(\mu_k - \mu)(\mu_k - \mu)^T \quad (4)$$

In Equation 4, n_k is the number of samples in class k, and $(\mu_k - \mu)(\mu_k - \mu)^T$ is the outer product of the deviation of each class mean from the overall mean.

4. Solve the Generalized Eigenvalue Problem

To find the optimal projection directions, solve the generalized eigenvalue problem for the matrix $S_W^{-1}S_B$, calculated in Equation 5.

$$S_W^{-1}S_B w = \lambda w \quad (5)$$

This equation finds eigenvectors w and corresponding eigenvalues λ. The eigenvectors corresponding to the largest eigenvalues provide the directions that maximize the separation between classes.

5. Select the Projection Matrix *W*

Construct the projection matrix W from the selected eigenvectors. Typically, the eigenvectors corresponding to the largest eigenvalues are chosen to form this matrix.

6. Project the Data onto the New Space

Finally, project the original data X onto the new space defined in Equation 6 by the projection matrix W.

$$Y = XW \quad (6)$$

Here, Y is the transformed data in the new lower-dimensional space that maximizes class separability. By following these steps, LDA transforms the original feature space into a new space where the classes are well-separated, facilitating better classification and reducing the dimensionality of the data while preserving class discrimination.

2.4 Evaluation Metrics

The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known. The matrix displays the actual versus predicted classifications, providing insight into the model's accuracy.

TP (True Positive) refers to correctly predicted positive cases, FN (False Negative) denotes actual positive cases incorrectly predicted as negative, FP (False Positive) indicates actual negative cases incorrectly predicted as positive, and TN (True Negative) refers to correctly predicted negative cases.

The evaluation metrics for this model include the confusion matrix and the classification report. The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known. The matrix displays the actual versus predicted classifications, providing insight into the model's accuracy. Using the values from the confusion matrix, several important metrics can be derived: accuracy, precision, recall (sensitivity), and F1-score, as calculated in Equations 7-10. The classification report complements the confusion matrix by providing a more detailed breakdown of the performance of each class. It includes metrics such as precision, recall, and the F1 score for each class, offering a comprehensive overview of the model's effectiveness. The classification report is beneficial for understanding the model's performance across different classes identifying any imbalances or areas where the model may need improvement.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

3. RESULT AND DISCUSSION

3.1 Data Analysis Process Using LDA

Linear Discriminant Analysis (LDA) was applied in this study to classify mental health conditions by transforming the dataset into a more interpretable space while preserving essential class distinctions. The first step in the LDA process involved calculating within-class and between-class scatter matrices, which quantify how data points vary within and across different categories.

Through eigenvalue decomposition, two key discriminant functions were identified, representing the directions in which the mental health conditions differ most significantly. The first discriminant function was highly effective in distinguishing Bipolar Type-1 and Depression, indicating that these two conditions exhibit strong and contrasting patterns in the dataset.

Meanwhile, the second discriminant function helped differentiate Bipolar Type-2 and Normal conditions, although with greater overlap, suggesting that these conditions share more subtle similarities. Once transformed using these discriminant functions, the dataset was projected into a two-dimensional space, allowing for better visualization and interpretation.

The results revealed well-separated clusters for Bipolar Type-1 and Depression, demonstrating that these conditions have unique feature distributions that facilitate accurate classification. However, Bipolar Type-2 and Normal conditions exhibited some overlapping regions, indicating that their distinguishing characteristics are less pronounced.

This overlap explains the misclassifications observed in the performance metrics, as LDA struggled to clearly separate individuals with mild symptoms or those whose experiences fluctuated between these two categories. Despite this limitation, the LDA transformation provides valuable insight into how different mental health conditions relate to one another, reinforcing its potential as a diagnostic tool that can complement traditional assessments.

In addition to classification, LDA provided insights into the most influential features in distinguishing mental health conditions. Attributes such as Sadness, Mood Swings, and Overthinking played a crucial role in separating Depression and Bipolar Type-1 from other classes. Similarly, Optimism and Concentration were strong indicators for distinguishing Normal individuals from those with Bipolar Type-2.

Moreover, Aggressive Response and Suicidal Thoughts were among the most significant factors for identifying Bipolar Type-1, reflecting the severity of emotional instability associated with the condition. These findings emphasize the interpretability of LDA, as it not only enables classification but also highlights the underlying patterns and relationships between symptoms, offering meaningful contributions to mental health research and potential clinical applications.

3.2 Model Performance

Figure 2 presents the confusion matrix, which provides a detailed summary of the classification model's performance on the test data. This matrix illustrates the actual versus predicted classifications for the four mental health condition categories: Bipolar Type-2, Depression, Bipolar Type-1, and Normal.

The diagonal elements represent the correct predictions for each class, while the off-diagonal elements indicate misclassifications. For instance, the model correctly predicted eight cases of Bipolar Type-2, with no misclassifications. It also accurately classified eight Depression cases but misclassified one as Bipolar Type-2 and one as Normal.

Similarly, eight Bipolar Type-1 cases were correctly identified without error, and nine Normal cases were correctly predicted, with one misclassified as Depression. This confusion matrix highlights the model's effectiveness and pinpoints areas needing improvement in classifying mental health conditions.

The confusion matrix for the classification model reveals significant insights into its performance across the four mental health condition categories: Bipolar Type-2, Depression, Bipolar Type-1, and Normal. The model demonstrates excellent accuracy in identifying Bipolar Type-2 and Bipolar Type-1 cases, correctly classifying all eight instances of each condition with no misclassifications.

This indicates that the model effectively captures the unique features of these conditions, making it reliable for distinguishing them from other mental health disorders. However, the model's performance with Depression and Normal categories shows room for improvement. While it correctly classified eight out of ten Depression cases, one case was misclassified as Bipolar Type-2 and another as Normal.

Similarly, the model accurately identified nine out of ten cases for the Normal category but misclassified one as Depression. These misclassifications suggest that some features of Depression overlap with those of Bipolar Type-2 and Normal conditions, leading to occasional confusion. To enhance the model's accuracy, it would be beneficial to refine the features that differentiate

Depression from other conditions and consider incorporating more advanced techniques such as feature engineering or additional data collection. These steps could help reduce misclassification rates and improve the model's overall performance. Table 1 presents the classification report, providing a detailed breakdown of the model's performance across the four mental health condition categories.

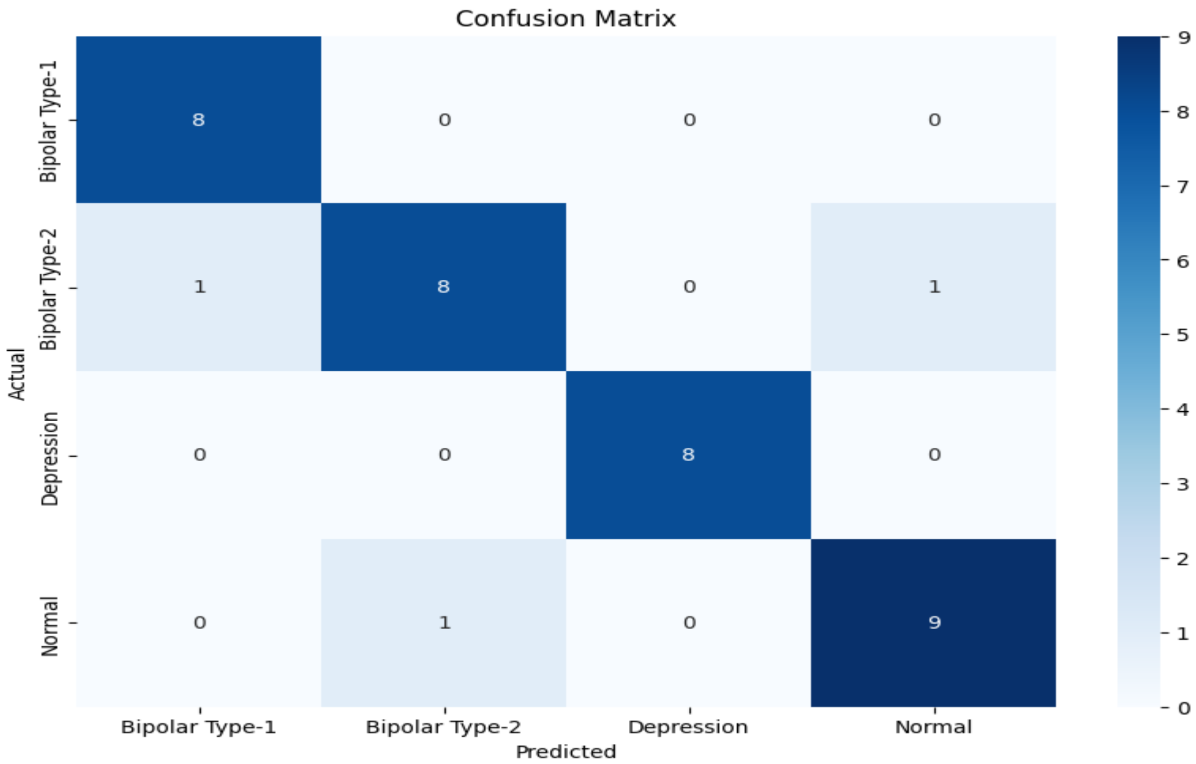

Figure 2. Confusion matrix

Table 1. Classification report

	precision	recall	f1-score	support
Bipolar Type-1	1.00	1.00	1.00	8
Bipolar Type-2	0.90	0.90	0.90	10
Depression	1.00	1.00	1.00	8
Normal	0.90	0.90	0.90	10
accuracy			0.94	36
macro avg	0.95	0.95	0.95	36
weighted avg	0.94	0.94	0.94	36

There are: Bipolar Type-1, Bipolar Type-2, Depression, and Normal. The report includes key metrics such as precision, recall, and F1-score for each class, alongside the dataset's number of instances (support). Precision measures the accuracy of positive predictions, recall indicates the ability to identify all actual positive cases, and the F1-score represents the harmonic mean of precision and recall.

The overall accuracy, macro average, and weighted average metrics offer a comprehensive summary of the model's effectiveness, reflecting its ability to accurately classify mental health conditions in a balanced and reliable manner. The classification report in

Table *1* highlights the model's overall strong performance in distinguishing between the four mental health conditions.

For Bipolar Type-1 and Depression, the model achieved perfect precision, recall, and F1-score scores, each 1.00. This indicates that the model accurately identified all instances of these conditions without any false positives or false negatives, demonstrating its high reliability and effectiveness in classifying these specific mental health conditions.

The perfect scores also suggest that the features selected for these classes are highly distinctive, allowing the model to separate them clearly from the other conditions. In contrast, the performance for Bipolar Type-2 and Normal categories, while still high, shows slight imperfections with precision, recall, and F1-scores at 0.90.

This means that the model had difficulty distinguishing these conditions, leading to a few misclassifications. For Bipolar Type-2 and Normal, the model misclassified 10% of the instances, indicating areas where feature overlap might be confusing.

Despite these minor inaccuracies, the model's overall accuracy remains high at 0.94, with macro and weighted averages of 0.95 and 0.94, respectively, reflecting a balanced performance across all classes. These results suggest that while the model is highly effective, further refinement of features or additional data may be necessary to achieve even higher accuracy, particularly for distinguishing Bipolar Type-2 and Normal conditions.

3.3 Summarization of Key Findings

This research aimed to evaluate the effectiveness of LDA in the early identification of mental health disorders, specifically Bipolar Type-1, Bipolar Type-2, Depression, and Normal conditions. Utilizing a dataset from Kaggle, the study meticulously applied LDA to classify these mental health conditions and assessed the model's performance through confusion matrices and classification reports.

The findings reveal that LDA excels in distinguishing Bipolar Type-1 and Depression with perfect precision, recall, and F1 scores, indicating its robust capability to identify these disorders accurately. However, it demonstrated slightly lesser accuracy for Bipolar Type-2 and Normal conditions, highlighting a potential area for improvement.

The research underscores the potential of LDA as a powerful diagnostic tool for mental health, proving its efficacy but also indicating the need for enhanced feature selection or larger datasets to improve its applicability across a broader spectrum of mental health conditions.

3.4 Result Interpretations

The results of the study indicate distinct patterns in the classification of mental health conditions, where LDA was highly effective for Bipolar Type-1 and Depression, achieving perfect metrics, but less so for Bipolar Type-2 and Normal categories. This suggests that the feature set used in the LDA model captured the characteristics of Bipolar Type-1 and Depression more distinctly than those of Bipolar Type-2 and Normal, possibly due to more apparent behavioral or symptomatic distinctions in the former conditions.

The slightly lower performance for Bipolar Type-2 and Normal may not have met initial expectations for universal high accuracy across all categories. These unexpected results could imply that overlapping symptoms between Bipolar Type-2 and Normal might lead to confusion, suggesting a need for refining the feature extraction or incorporating additional distinctive features.

Alternatively, these findings could point to inherent complexities within these conditions that are not as effectively captured by the current model's parameters, indicating the potential benefit of integrating more complex or alternative machine learning techniques to differentiate these categories better.

3.5 Research Implications

This research reinforces the relevance of LDA as a significant tool in the early detection of mental health disorders, linking it to the broader discourse on the need for precise and accessible diagnostic methods in the mental health field. By demonstrating high accuracy in diagnosing Bipolar Type-1 and Depression, the findings align with existing literature that underscores the potential of machine learning techniques in enhancing diagnostic processes.

Furthermore, the study introduces new insights into the variable performance of LDA across different mental health conditions, particularly highlighting challenges in distinguishing between Bipolar Type-2 and Normal conditions. This contributes to the ongoing discussion about the complexity of mental health diagnostics and the necessity for tailored approaches that consider the nuanced differences among disorders.

The implications extend to clinical practice, where integrating LDA could significantly aid early and accurate diagnosis, potentially leading to better patient management and outcomes. Additionally, this research paves the way for future studies to explore more sophisticated models or hybrid approaches that could address the identified limitations and improve the robustness of diagnostic tools in mental health.

3.6 Research Limitations

This study effectively demonstrates the utility of LDA in classifying certain mental health conditions. However, it is constrained by a relatively small dataset and the potential biases inherent in self-reported data. These limitations may impact the generalizability of the findings across a broader, more diverse population. They might contribute to the slightly diminished accuracy observed in distinguishing between Bipolar Type-2 and Normal conditions.

Despite these constraints, the results are still valid and valuable for the research question concerning the efficacy of LDA in mental health diagnostics. The high accuracy rates achieved in classifying Bipolar Type-1 and Depression confirm that LDA can be a powerful tool in clinical settings, mainly when precise symptom delineation is feasible.

The study's outcomes provide a solid foundation for further exploration and validation with more extensive and varied datasets, potentially including more objective clinical measures to enhance the robustness and applicability of the findings in real-world scenarios. Thus, while acknowledging the limitations, the study offers significant insights into the capabilities of LDA in mental health diagnostics, supporting its potential for broader clinical implementation.

3.7 Recommendations for Future Research

Future research should focus on expanding the dataset to include a more extensive and diverse population, enhancing the robustness and generalizability of the LDA approach in mental health diagnostics. Practical implementation could benefit from integrating LDA with clinical decision-making processes, particularly in settings where quick and effective diagnostic tools are critical.

Additionally, incorporating objective clinical data alongside self-reported symptoms could mitigate biases and improve the accuracy of diagnostics. Investigating hybrid models that combine LDA with other machine learning techniques, such as neural networks or ensemble methods, might offer improvements in handling overlapping symptoms between conditions like Bipolar Type-2 and Normal.

Moreover, future studies could explore the impact of feature engineering techniques to capture the nuances of mental health conditions better, ultimately leading to more precise classification and better patient outcomes. This multifaceted approach would further validate LDA's utility and expand its applicability in the evolving field of mental health technology.

4. CONCLUSIONS

This study successfully achieves its objective of evaluating the feasibility of Linear Discriminant Analysis (LDA) as a tool for the early identification of mental health disorders. By applying LDA to a structured dataset, the research demonstrates that Bipolar Type-1 and Depression can be accurately classified, confirming the model's strength in identifying conditions with distinct symptom patterns.

However, the misclassification of Bipolar Type-2 and Normal cases indicates the need for further refinement, particularly in addressing overlapping features and subtle symptom variations. These findings validate LDA's potential as a computational diagnostic aid while highlighting areas for improvement. Addressing the dataset's size, diversity, and integration of clinical markers will be essential to enhance classification accuracy and model robustness.

Ultimately, this research contributes to the ongoing development of machine learning applications in mental health diagnostics, reinforcing LDA's role as a promising yet evolving approach that could complement traditional assessment methods in clinical practice.

REFERENCES

Bai, Y. (2023). Deep Learning-Based Pre-Diagnosis and Analysis of Psychological Disorders. International Journal of Education and Humanities, 6(3), 126–130. https://doi.org/10.54097/ijeh.v6i3.4763

Christensen, M. K., Lim, C. C. W., Saha, S., Plana-Ripoll, O., Cannon, D., Presley, F., Weye, N., et al. (2020). The Cost of Mental Disorders: A Systematic Review. Epidemiology and Psychiatric Sciences, 29. https://doi.org/10.1017/s204579602000075x

Csillag, C., Nordentoft, M., Mizuno, M., McDaid, D., Arango, C., Smith, J., Lora, A., et al. (2017). Early Intervention in Psychosis: From Clinical Intervention to Health System Implementation. Early Intervention in Psychiatry, 12(4), 757–764. https://doi.org/10.1111/eip.12514

Dalal, S., Jain, S., & Dave, M. (2023). Early Depression Detection Using Textual Cues From Social Data: A Research Agenda, 393–406. http://dx.doi.org/10.1007/978-981-19-9090-8_34

Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., et al. (2020). Mental Health Problems and Social Media Exposure During COVID-19 Outbreak. Plos One, 15(4), e0231924.

https://doi.org/10.1371/journal.pone.0231924

Hermens, D. F. (2023). How Could Brain Fingerprinting Lead to the Early Detection of Mental Illness in Adolescents and What Are the Next Steps? Expert Review of Neurotherapeutics. https://doi.org/10.1080/14737175.2023.2226870

Kersemaekers, W. M., Rupprecht, S., Wittmann, M., Tamdjidi, C., Falke, P., Donders, R., Speckens, A. E. M., et al. (2018). A Workplace Mindfulness Intervention May Be Associated With Improved Psychological Well-Being and Productivity. A Preliminary Field Study in a Company Setting. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.00195

Lavanya, B., KV, R. K., & Nanjappan, D. (2023). The Effectiveness of Self-Instructional Module on Knowledge and Attitude Regarding Mental Illness and Its Early Identification Among Female Junior Health Assistants in Selected Primary Health Centers at Hassan. International Journal of Advanced Psychiatric Nursing, 5(1), 165–170. https://doi.org/10.33545/26641348.2023.v5.i1c.123

McDaid, D., Park, A., & Wahlbeck, K. (2019). The Economic Case for the Prevention of Mental Illness. Annual Review of Public Health, 40(1), 373–389. https://doi.org/10.1146/annurev-publhealth-040617-013629

Nunes, L. C., Pinheiro, P. R., Pinheiro, M. C. D., Filho, M. S., Nunes, R. E. C., & Pinheiro, P. G. C. D. (2019). Automatic Detection and Diagnosis of Neurologic Diseases. Ieee Access, 7, 29924–29941.

Reddy, C. K. K., & Satvika, P. (2023). An Efficient Early Diagnosis and Healthcare Monitoring System for Mental Disorder Using Machine Learning Approach. http://dx.doi.org/10.21203/rs.3.rs-2912222/v1

Samartzis, L., & Talias, M. A. (2019). Assessing and Improving the Quality in Mental Health Services. International Journal of Environmental Research and Public Health, 17(1), 249. https://doi.org/10.3390/ijerph17010249

Silove, D. (2021). Challenges to Mental Health Services for Refugees: A Global Perspective. World Psychiatry, 20(1), 131–132. https://doi.org/10.1002/wps.20818

Srivastava, K., Chaudhury, S., Bhat, P., & Mujawar, S. (2018). Media and Mental Health. Industrial Psychiatry Journal, 27(1), 1. https://doi.org/10.4103/ipj.ipj_73_18

Sudibyo, U., Rustad, S., Andono, P. N., Fanani, A. Z., Purwanto, P., & Muljono, M. (2020). A Novel Approach on Linear Discriminant Analysis (LDA). http://dx.doi.org/10.1109/iSemantic50169.2020.9234274

Trendafilov, N. T., & Gallo, M. (2021). Linear Discriminant Analysis (LDA), 229–268. http://dx.doi.org/10.1007/978-3-030-76974-1_7

William, D., & Suhartono, D. (2021). Text-Based Depression Detection on Social Media Posts: A Systematic Literature Review. Procedia Computer Science, 179, 582–589. https://doi.org/10.1016/j.procs.2021.01.043

Zhang, B., Wei, D., Chang, W., & Yang, Z. (2023). A method of mental disorder recognition based on visibility graph. Journal of Biomedical Engineering, 40(3), 442–449. https://doi.org/10.7507/1001-5515.202208077

Zhao, H., Lai, Z., Leung, H., & Zhang, X. (2020). Linear Discriminant Analysis, 71–85. http://dx.doi.org/10.1007/978-3-030-40794-0_5

Андреев, A. B., Kurkin, S. A., Stoyanov, D., Badarin, A., Paunova, R., & Hramov, A. E. (2023). Toward Interpretability of Machine Learning Methods for the Classification of Patients With Major Depressive Disorder Based on Functional Network Measures. Chaos an Interdisciplinary Journal of Nonlinear Science, 33(6). https://doi.org/10.1063/5.0155567