KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 4, February 2024, Hal 2119-2128
DOI 10.30865/klik.v4i4.1703
https://djournals.com/klik

Comparative Analysis of Neural Network Architectures for Mental Health Diagnosis: A Deep Learning Approach

Gregorius Airlangga

Engineering Faculty, Information System Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: gregorius.airlangga@atmajaya.ac.id
Email Corresponding Author: gregorius.airlangga@atmajaya.ac.id

Abstract−Mental health conditions present a complex diagnostic challenge due to the subtlety and diversity of symptoms. This study provides a comprehensive analysis of various neural network architectures, including Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory networks (LSTM), and Dense Neural Network (DNN), in their ability to classify mental health conditions. Utilizing a rich dataset of symptoms and expert diagnoses, we preprocessed the data to address class imbalances and trained each model to evaluate its diagnostic performance. Our results are presented through confusion matrices that reveal the accuracy, precision, recall, and F1-scores for each model. The MLP and DNN models demonstrated high accuracy in identifying distinct conditions but struggled with overlapping symptoms. LSTM and RNN models captured temporal patterns to some extent yet required further optimization for improved accuracy. CNN models showed robust feature detection capabilities, with the CNN 1D model excelling in specificity for certain conditions. However, a common challenge across all models was the differentiation between conditions with similar symptom presentations. Our findings suggest that while individual models have their strengths, an ensemble approach may be necessary for enhanced diagnostic precision. Future work will focus on integrating models, refining feature extraction, and employing explainable AI to increase transparency and trust in model predictions. Additionally, expanding the dataset and conducting clinical trials will ensure the models' effectiveness in real-world settings. This research moves us closer to achieving nuanced, AI-driven diagnostics that can support clinicians and benefit patient outcomes in mental healthcare.

Keywords: Mental Health; Machine Learning; Comparison; Deep Learning; Neural Network

1. INTRODUCTION

The escalating global incidence of mental health disorders presents a formidable challenge to healthcare systems worldwide [1]–[3]. Mental health conditions, such as depression, anxiety, and schizophrenia, not only exert a profound impact on the quality of life of individuals but also impose substantial economic burdens on societies [4]–[6]. Traditional diagnostic methods in psychiatry, which primarily rely on patient interviews and psychological assessments, are subjective and may lead to inconsistent diagnoses [7]–[9]. The subjective nature of these assessments, coupled with the complexity and heterogeneity of psychiatric disorders, underscores the critical need for innovative diagnostic approaches [10]–[12]. In this context, artificial intelligence (AI) and machine learning (ML) technologies emerge as potent tools capable of transforming the landscape of psychiatric diagnostics through the objective analysis of complex, multidimensional data [13]–[15]. Extensive research has been conducted to explore the application of AI in the diagnosis and treatment of mental health disorders. Early studies focused on leveraging machine learning algorithms to analyze structured clinical data, yielding promising results in identifying potential biomarkers and diagnostic patterns [16]. As the field has evolved, the focus has shifted towards more sophisticated deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), applied to diverse data sources including neuroimaging, genetic profiles, and digital phenotyping [17]–[19]. These advanced models have demonstrated a remarkable ability to discern intricate patterns within data, offering insights into the biological and behavioral underpinnings of psychiatric conditions. Despite these advancements, significant challenges remain, including issues related to data quality and availability, model interpretability, and the ethical implications of deploying AI in clinical settings [20].

The urgency of addressing mental health issues has never been more pronounced. The WHO has highlighted the growing prevalence of mental disorders globally, emphasizing the need for more effective diagnostic and treatment strategies [21]. The COVID-19 pandemic has further exacerbated mental health issues, making the development of scalable and objective diagnostic tools an imperative [22]. Accurate and timely diagnosis is crucial for effective intervention, yet the current reliance on subjective assessments presents significant barriers to achieving this goal [23]. AI-driven diagnostic tools have the potential to revolutionize mental health care, providing a means to supplement clinical expertise with data-driven insights [24]. The state-of-the-art in AI-assisted psychiatric diagnosis is characterized by a rich tapestry of approaches, each with its strengths and limitations. Deep learning models have been at the forefront of recent advances, demonstrating superior performance in tasks ranging from predictive modeling based on electronic health records to the analysis of complex neuroimaging data [25]. These models, however, are not without their challenges. The black-box nature of many deep learning algorithms, along with concerns regarding data privacy, model generalizability, and the potential for algorithmic bias, poses significant hurdles to the clinical adoption of AI technologies in psychiatry [26]. In addition, a detailed examination of the literature reveals several gaps that this research seeks to fill. Firstly, there is a notable lack of studies comparing the efficacy of different deep learning architectures within the same dataset, a gap that hinders the identification of optimal models for specific diagnostic tasks [27]. Secondly, the challenge of data imbalance, a common issue in medical datasets, is often inadequately addressed, affecting the accuracy and fairness of diagnostic models [28]. Furthermore, the ethical and practical considerations surrounding the deployment of AI in clinical settings require more thorough investigation.

This research aims to address the challenges by developing a comprehensive AI framework for the diagnosis of mental health disorders. By conducting a systematic comparison of various deep learning architectures, including CNNs, RNNs, LSTM networks, and traditional neural networks, this study endeavors to identify the most efficacious models for psychiatric diagnosis. The research is particularly focused on overcoming the issues of data imbalance and enhancing model interpretability, with the goal of facilitating the clinical integration of AI diagnostic tools. This study also makes several contributions to the field of psychiatric diagnostics. It offers a comparative analysis of multiple deep learning architectures, providing insights into the optimal models for diagnosing mental health disorders. It introduces innovative preprocessing and data augmentation strategies to tackle the issue of data imbalance, thereby enhancing the accuracy and reliability of diagnostic models. Moreover, the study incorporates interpretability mechanisms into the AI models, improving their transparency and trustworthiness. Finally, it discusses the ethical considerations and potential clinical applications of AI in psychiatry, paving the way for the responsible use of AI technologies in mental health care. The article is organized as follows: Section 2 describes the methodology, including the dataset, preprocessing techniques, model architectures, and evaluation metrics. Section 3 presents the results, detailing the performance of each model and offering a comparative analysis. In addition, it discusses the implications of these findings, focusing on the potential of AI to enhance diagnostic accuracy and addressing the ethical considerations of AI in clinical practice. Section 4 concludes the article, summarizing the key findings and their implications for the advancement of psychiatric diagnostics through AI.

2. RESEARCH METHODOLOGY

In this research study, as presented in the figure 1, we navigate through a meticulous workflow designed to leverage the strengths of various deep learning models for the classification of mental health conditions. Initially, we acquire a comprehensive dataset, "Dataset-Mental-Disorders.csv", which is then meticulously prepared to ensure quality and relevance. Following preparation, we engage in a thorough preprocessing phase where the data undergoes cleaning, imputation for missing values, encoding of categorical variables, normalization to standardize numerical inputs, and balancing to correct for any class imbalances. With a clean and structured dataset, we proceed to the selection and optimization of advanced neural network models such as CNNs, RNNs, LSTMs, and DNNs, each chosen for their ability to process complex data patterns. The performance of these models is rigorously evaluated using a series of metrics and confusion matrices to ascertain their diagnostic accuracy. To facilitate these processes, we employ a suite of Python libraries, including Pandas for data manipulation, Matplotlib and Seaborn for visualization, and TensorFlow/Keras for building and refining our neural network models, with the Imbalanced-learn library addressing any issues of class imbalance. This streamlined yet robust approach encapsulates our commitment to applying cutting-edge AI techniques to the nuanced task of mental health diagnosis.

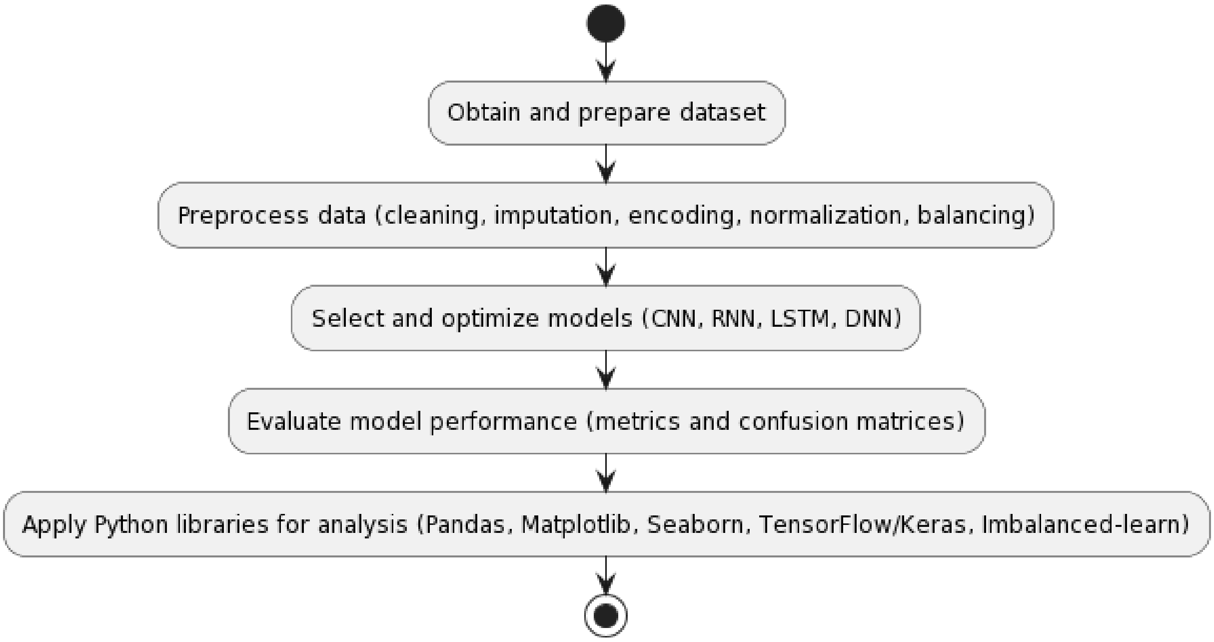

Figure 1. Research Methodology Workflow

2.1 Dataset Preparation

Our research employs the "Dataset-Mental-Disorders.csv" dataset and can be collected from [29], which has been carefully curated to cover a broad range of mental health conditions along with their symptomatology. This comprehensive dataset encompasses various attributes, including demographic details (such as age and gender), clinical manifestations (like mood fluctuations, levels of anxiety, thought processes, etc.), and diagnoses as assessed by mental health experts. It consists of numerous cases, each accurately annotated with a diagnosis by psychiatric professionals. The dataset's richness and breadth are crucial for simulating the complexity encountered in diagnosing mental health issues in clinical settings. This, in turn, serves as a solid base for deploying and assessing the performance of artificial intelligence models in our study.

2.2 Dataset Preprocessing

During the early phases of our study, we undertook an extensive data cleansing operation to remove any elements that might introduce bias, with a particular emphasis on excluding non-essential or identifying information such as the 'Patient Number'. This action was essential to prevent the analysis from being influenced by data irrelevant to our predictive goals. Simultaneously, we tackled the common problem of missing entries often encountered in real-world datasets. To address this, we employed imputation strategies, specifically using the mode of each column to replace missing values as described in the provided equation (1). This approach was chosen to uphold the dataset's integrity and ensure its completeness, laying a robust groundwork for our analysis to be built upon fully populated data.

$$x_{\text{missing}} = \text{mode}(X) \qquad (1)$$

In this context x_{missing} signifies the absent data point within dataset attribute X, while mode(X) identifies the attribute's most commonly occurring value. As the preprocessing progressed, we transformed the dataset's categorical attributes using Label Encoding, as depicted in equation (2). This method translated textual categories into numerical values, making the dataset compatible with machine learning algorithms. Acknowledging the potential for significant variation in the scale of numerical values, we also undertook a normalization procedure. This adjustment rescaled the numerical values to a uniform range, a crucial step for models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) that are sensitive to the magnitude of features. This normalization, as outlined in equation (3), guarantees that the model's performance is not unduly affected by the scale of any individual feature, ensuring an equitable assessment across all variables.

$$L(c_i) = k \qquad (2)$$

In this scenario, c_i represents the specific categorical value within the dataset, where L is the function used for label encoding, and k is the numerical representation assigned to to c_i. The dataset encountered notable issues with class imbalance, where some mental health conditions appeared more frequently than others. To address the potential bias this imbalance could introduce, favoring more prevalent conditions, we employed the RandomOverSampler method as indicated in equation (4). This technique adjusted the class distribution by augmenting the representation of less common conditions, ensuring every diagnosis was fairly represented within the dataset. Such a preprocessing measure is vital for the integrity of the study, guaranteeing that the machine learning models develop an unbiased ability to recognize all conditions with consistent precision, instead of being skewed towards detecting more frequent diagnoses. Through diligent data preprocessing, we laid the groundwork for an unbiased and thorough assessment of machine learning algorithms in the identification of mental health issues.

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)} \times (b - a) + a \qquad (3)$$

Where X_{norm} is the normalized value, X is the original value, min(X) and max(X) are the minimum and maximum values observed in the dataset, respectively, and a and b represent the scale's desired range, typically 0 and 1.

$$N'_{\text{minority}} = N_{\max} \qquad (4)$$

Where N'_{minority} is the new number of instances in each of the minority classes after applying RandomOverSampler, aiming for all classes to have a uniform number of instances $N_{\max}$.

2.3 Model Selection and Optimization

This investigation seeks to rigorously explore the application of deep learning methodologies for the detection and diagnosis of mental health disorders. Leveraging a comprehensive set of deep learning techniques, our study delves into the nuanced potential these models hold for such complex classification challenges. Specifically, we employed advanced neural network models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Dense Neural Networks (DNNs), each selected for its inherent strengths in processing and analyzing the intricate patterns within our dataset. The selection of these models was strategic, aimed at harnessing the diverse computational approaches each model brings to the analysis of mental health conditions. For instance, CNNs were chosen for their proficiency in identifying spatial hierarchies in data, whereas RNNs and LSTMs were utilized for their ability to process sequences of data, capturing temporal dynamics which are crucial in understanding patterns in patient symptoms over time. DNNs were incorporated to assess their capability in learning complex representations of the data through their deep, interconnected structures. This varied model lineup was intended to encapsulate a wide spectrum of deep learning architectures, offering a rich, comparative insight into their efficacy in classifying mental health disorders. The experiment involved preprocessing the "Dataset-Mental-Disorders.csv" to ensure optimal model input, including the encoding of categorical variables, handling of missing values, and normalization of data to address class imbalance through techniques like RandomOverSampler. This preparatory work set the stage for a thorough evaluation of each model's performance, assessed through metrics such as accuracy, precision, recall, and F1 scores, alongside the generation of confusion matrices for a visual understanding of model efficacy.

To optimize these models to their peak performance, the research utilized RandomizedSearchCV, an efficient technique aimed at exploring the extensive range of potential hyperparameters available for configuring these models. This approach diverges from the exhaustive nature of traditional grid search techniques, which systematically test every possible combination of parameters. RandomizedSearchCV, as outlined in equation (15), selects a random subset of parameters for evaluation in each iteration. This approach significantly accelerates the optimization process by limiting the number of iterations needed for evaluation while still ensuring comprehensive exploration of the parameter space. This method increases the chances of discovering the most efficacious settings for the models. Such an optimization approach is crucial as it balances computational efficiency with the improvement of model accuracy, facilitating the identification of hyperparameters that are finely adjusted for accurately diagnosing mental health issues with utmost precision and reliability.

2.4 Evaluation Metrics

The performance evaluation of the machine learning models in our research was comprehensively carried out using a set of crucial metrics, each providing distinctive insights into the models' ability to diagnose accurately. The metrics utilized—accuracy, precision, recall, F1-score, as outlined in equations (5)-(9), alongside the confusion matrix—are vital for an all-encompassing review, shedding light on both the general accuracy of the models' predictions and their proficiency in correctly identifying individual classes while maintaining a balance between sensitivity and specificity. Accuracy is determined by the proportion of correct predictions (true positives and true negatives) over all predictions made, serving as a direct indicator of the models' overall performance. Precision evaluates the models' exactness by calculating the percentage of true positive predictions out of all positive predictions made, highlighting the models' dependability in identifying positive cases. Recall, also known as sensitivity, measures the models' capacity to detect all actual positive cases, determined by the fraction of true positives out of the total actual positives. The F1-score, which is the harmonic mean of precision and recall, acts as a singular metric to gauge the equilibrium between precision and recall, proving indispensable when a balanced importance is attributed to both metrics. Complementing these metrics, the confusion matrix provides a visual and quantitative analysis of the models' accuracy, delineating the precise counts of true positives, true negatives, false positives, and false negatives.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

$$\theta^* = \arg\max_{\theta \in \Theta} f\,(\theta) \tag{9}$$

2.5 Software Tools

Our investigation leverages Python's powerful programming capabilities, renowned for its extensive library ecosystem and strong community support, especially beneficial in the fields of data science and deep learning. For this study, we carefully selected several Python libraries, each known for its specific strengths that complement our research objectives. The Pandas library was instrumental in data manipulation and preparation, facilitating efficient data cleaning, transformation, and structuring to prime the dataset for deep learning model application. Visualization of data and model outcomes was adeptly handled by Matplotlib and Seaborn, providing rich visualization options that greatly aided in exploratory data analysis and in making the results comprehensible and visually engaging. These tools were crucial in uncovering data patterns, anomalies, and in visually summarizing the performance of various deep learning models. For model construction, training, and evaluation, we employed TensorFlow and its high-level API, Keras, which offered a streamlined pathway for developing sophisticated neural network architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks. These frameworks allowed for the flexible design of deep learning models tailored to the nuances of mental health diagnosis, facilitating the implementation of models capable of capturing complex patterns in the data.

The research also addressed the challenge of class imbalance through the use of the Imbalanced-learn library, which provided strategies like RandomOverSampler to achieve a balanced class distribution in the training data. This was vital for ensuring that our models could equally learn to identify various mental health conditions without bias toward more prevalent labels.In integrating these tools, from data preprocessing with Pandas to model development and evaluation with TensorFlow/Keras, and addressing data imbalance with Imbalanced-learn, our methodology encapsulated a holistic approach to applying deep learning in the realm of mental health diagnostics. This comprehensive use of Python libraries not only facilitated the effective analysis of the dataset but also ensured the robustness and reliability of our deep learning models in diagnosing mental health conditions.

3. RESULT AND DISCUSSION

3.1 Result Explanation

Table 1 presents a comparative analysis of various deep learning and machine learning models in terms of their diagnostic performance on a mental health dataset. Each model is evaluated based on four key metrics: accuracy, precision, recall, and F1-score. Starting with the CNN Sequence model, it exhibits moderate performance with an accuracy of approximately 0.4583. This suggests that it correctly predicts the diagnosis in around 45.83% of cases. Its precision, at 0.5099, indicates that when it predicts a diagnosis, it is correct about 50.99% of the time. However, its recall and F1-score are both aligned with its accuracy, suggesting a balanced but moderate ability to identify true positives and a fair trade-off between precision and recall. The LSTM model's accuracy and recall are identical, standing at 0.4167, meaning it correctly identifies the diagnosis in approximately 41.67% of cases. Its precision is notably higher at 0.6580, suggesting that while it may not predict as many true positives as desired, when it does predict a case as positive, it tends to be more reliable. However, its F1-score is the lowest among the models at 0.3991, indicating a disparity between precision and recall.

The DNN model shows a substantial improvement in performance across all metrics compared to the previous two models, with an accuracy and recall of 0.7083. This indicates a correct diagnosis in about 70.83% of cases. Its precision is also high at 0.7393, and the F1-score stands at 0.7178, suggesting a good balance between precision and recall and a strong overall performance. The CNN 1D model demonstrates excellent performance, with the highest accuracy and recall of 0.75. Its precision at 0.8646 indicates high reliability in its positive predictions, and the F1-score of 0.7573 mirrors the high accuracy and recall, pointing to a very effective model. The RNN model, on the other hand, has lower performance metrics, with an accuracy and recall equal to 0.4167. Its precision is moderate at 0.5972, and the F1-score is 0.4508, which indicates that this model may struggle with correctly identifying true positives and negatives. The CNN model's performance is strong, with an accuracy and recall of 0.7083, similar to the DNN model. Its precision is also at 0.7393, and it has a healthy F1-score of 0.7178, suggesting that it is adept at balancing the identification of true positives and negatives. Lastly, the MLP (Multilayer Perceptron) model aligns with the CNN 1D model in terms of accuracy and recall at 0.75, suggesting that it correctly diagnoses 75% of cases. The precision and F1-score are also high at 0.8646 and 0.7573, respectively, which indicates that the MLP model is highly effective in its diagnostic predictions and maintaining a strong balance between precision and recall.

Table 1. Comparison Results

Name	Accuracy	Precision	Recall	F1-Score
CNN Sequence	0.4583	0.5099	0.4583	0.4502
LSTM	0.4167	0.6580	0.4167	0.3991
DNN	0.7083	0.7393	0.7083	0.7178
CNN 1D	0.75	0.8646	0.75	0.7573
RNN	0.4167	0.5972	0.4167	0.4508
CNN	0.7083	0.7393	0.7083	0.7178
MLP	0.75	0.8646	0.75	0.7573

3.2 Discussion

Figure 2 presents an informative visual examination of how various symptoms correlate with mental health diagnoses established by experts. The figure is composed of a series of histograms overlaid with density plots, each representing the frequency distribution of a particular symptom and its association with diagnoses, which are differentiated by a color code ranging from 0 to 3. These colors facilitate a clear visual distinction of the data by diagnosis, offering a layered perspective on the symptom profiles. In interpreting these histograms, attention is directed toward the x-axis, which measures the severity or occurrence rate of the symptoms, while the y-axis indicates the number of instances for each diagnostic group. These histograms provide a dual function: they quantify the frequency of symptoms and illustrate the relative prevalence of symptoms across various mental health disorders. For example, a symptom like "Euphoria" may display a diverse distribution, with certain diagnostic groups showing a heightened presence at more intense levels of the symptom. This pattern may suggest that "Euphoria" is particularly prominent in those groups and could play a crucial role in the diagnosis.

The relationship between symptom manifestation and expert diagnosis becomes apparent through the density of specific colors at certain intervals in the histograms. If a particular color predominates at the higher spectrum of the "Exhaustion" histogram, this signifies a strong link between that symptom and a specific mental health condition. Such visual indicators are instrumental in pinpointing which symptoms are characteristic of particular disorders and may also uncover symptoms that are prevalent across several conditions, suggesting a common underlying pathology or symptom intersection. The density plots that accompany the histograms offer a smoothed outline of the data distribution, clarifying the concentration of data points and accentuating the most frequent symptom intensities associated with each diagnosis. These curves are invaluable for identifying central tendencies and the dispersion of data, which are essential for understanding both the standard and atypical symptom presentations. Analyzing these visual plots side by side can unveil whether symptoms are specific to certain conditions or shared across them. For instance, a symptom like "Suicidal Thoughts" might be closely grouped within a particular diagnostic category, indicating a strong linkage to that condition. In contrast, a symptom such as "Sleep Disorder," if dispersed across various diagnostic groups, could denote a general symptom that is not exclusive to any single disorder. The form of the histograms may also imply data skewness; an elongation towards the more severe end of symptom intensities suggests a skew towards more extreme manifestations of symptoms. Furthermore, the breadth of the data within each histogram reflects the range of symptom severity; wider bins with higher frequencies denote a spectrum of symptom intensities within certain diagnoses, pointing to variability in patient experiences and symptom reporting.

Outlying data points, appearing as isolated bins away from the central distribution, might indicate atypical presentations of symptoms. Such anomalies may arise from diverse causes, such as data collection errors, unique individual experiences, or diagnostic inaccuracies, and necessitate additional scrutiny to comprehend their origins and impact. For healthcare practitioners, the intricate visual details in these plots can aid in refining diagnostic precision by underscoring symptom trends strongly linked to specific mental health issues. Moreover, the recognition of symptom overlaps across different diagnoses depicted in these plots is crucial for identifying comorbid conditions. From a research perspective, these symptom distributions are immensely beneficial for designing treatment strategies. Symptoms that exhibit a marked association with certain mental health diagnoses could be targeted with greater specificity, fostering more individualized and efficacious treatment approaches.

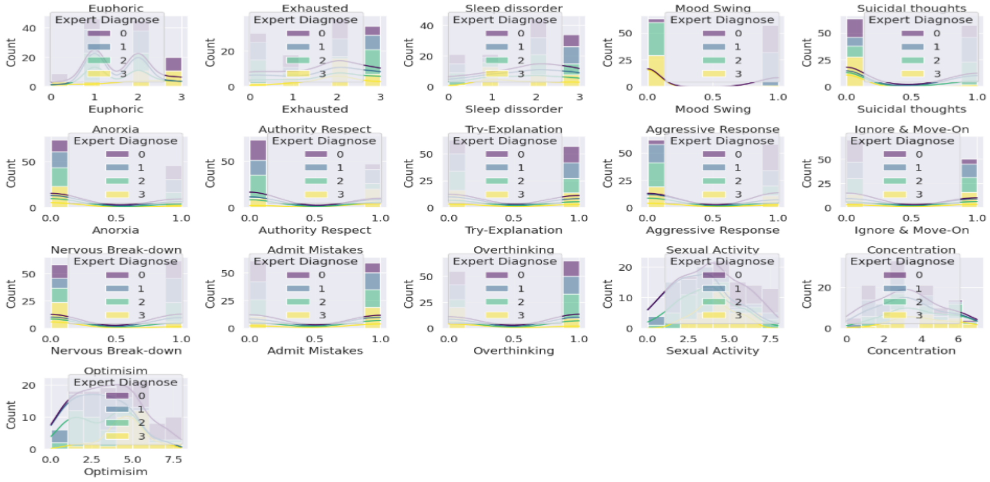

Figure 2. Histogram of Mentality Disorder Symptoms

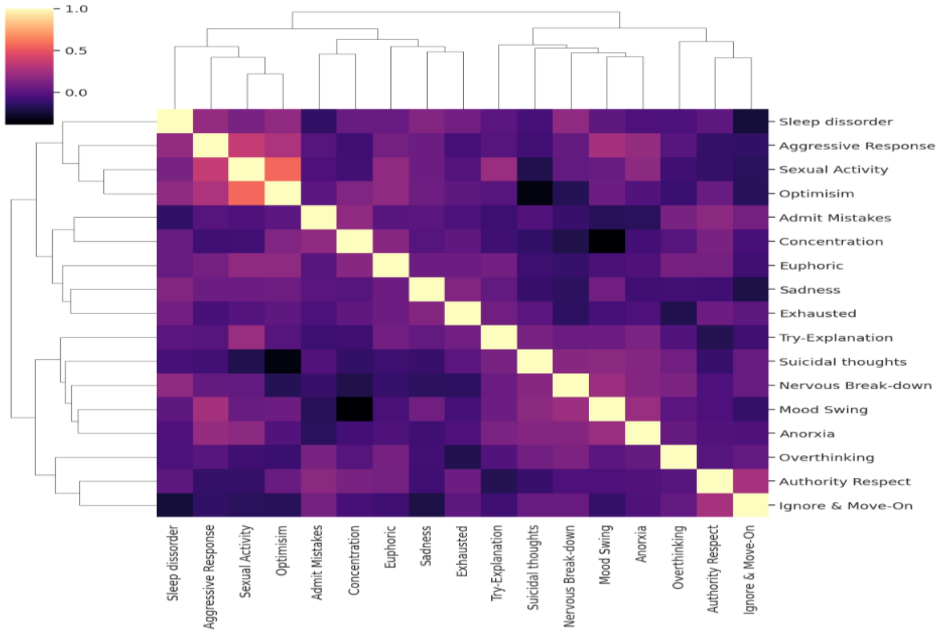

Figure 3. Heatmap of hierarchical clustering

Figure 3 showcases a heatmap enhanced by hierarchical clustering, offering a detailed and nuanced visualization of the interplay between various symptoms that are frequently assessed in mental health diagnostics. This form of visual representation excels at demonstrating how closely related different symptoms are, using a spectrum of colors to denote the intensity and direction of the correlations between pairs of symptoms. Lighter colors like yellow on the heatmap indicate a strong positive correlation, suggesting that an increase in one symptom often coincides with an increase in another. On the other hand, darker colors signal a negative correlation, pointing to an inverse relationship where one symptom's increase may correspond to a decrease in another. The heatmap is carefully labeled with a range of psychological symptoms or behavioral indicators such as "Sleep Disorder," "Euphoric," or "Aggressive Response." These labels are not arbitrary; they are strategically organized along the rows and columns to align with the dendrograms—tree-like structures that extend along the axes. These dendrograms perform a critical function by categorizing symptoms into clusters based on the similarity of their correlation profiles, potentially hinting at common etiologies or shared characteristics within psychiatric disorders. They offer a visual hierarchy, depicting the degree of relatedness among the symptoms as revealed by the dataset.

To fully interpret the heatmap, one must consider both the individual colors and the overall pattern they create. For example, if "Sadness" and "Exhaustion" consistently appear in similar light hues, this pattern may indicate a symptom cluster that could be characteristic of a particular mental health condition. The heatmap thus provides practitioners a valuable perspective on the symptom patterns they encounter, underscoring pertinent connections that may inform diagnostic assessments or treatment plans. Beyond its clinical implications, the heatmap is a rich resource for mental health research. It can inform the formation of hypotheses for empirical studies by illustrating patterns of clustering and correlation. For instance, if "Optimism" is frequently depicted in high contrast to other symptoms, it might lead researchers to investigate its role as a potential buffer against mental health issues. However, it is crucial to acknowledge the heatmap's constraints. While it effectively displays correlations, it does not imply causation. A strong correlation does not establish a cause-and-effect relationship between symptoms; it may merely reflect that they are symptoms of the same underlying issue. Furthermore, the data represented in the heatmap is a reflection of a particular sample and might not universally apply to all demographic groups. Therefore, while the heatmap is a powerful tool for mapping out symptom correlations, its findings should be interpreted within the broader context of clinical expertise and additional research.

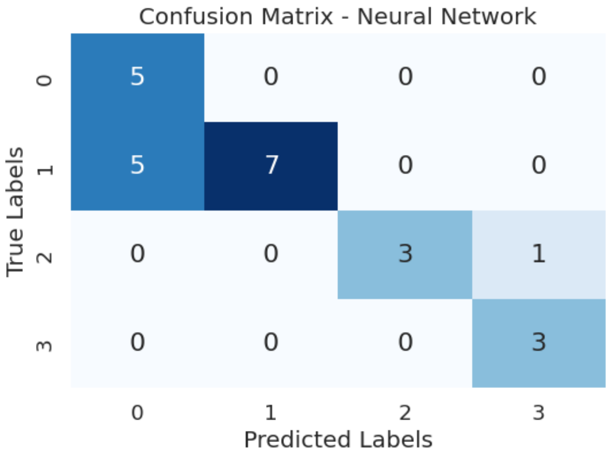

Figure 4. Confusion Matrix of MLP

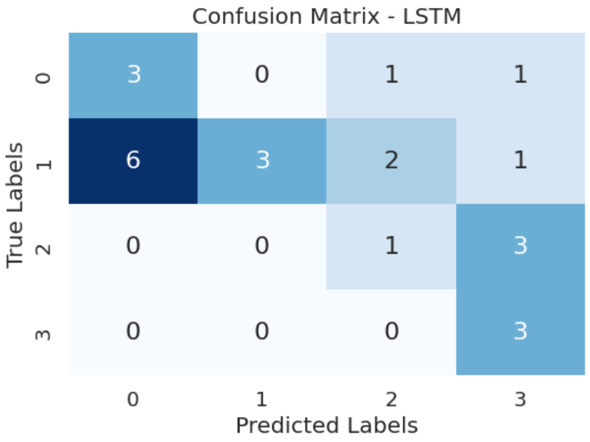

Figure 5. Confusion Matrix of LSTM

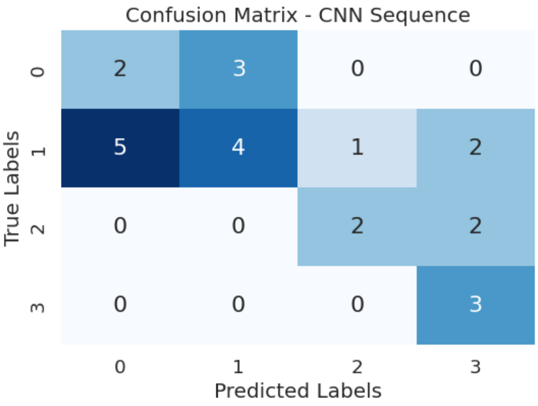

Figure 6. Confusion Matrix of CNN Sequence

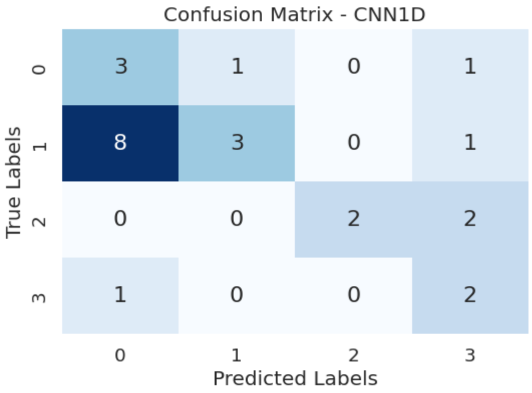

Figure 7. Confusion Matrix of CNN 1 D

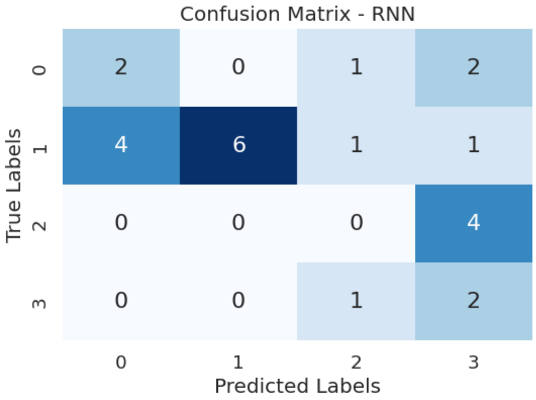

Figure 8. Confusion Matrix of RNN

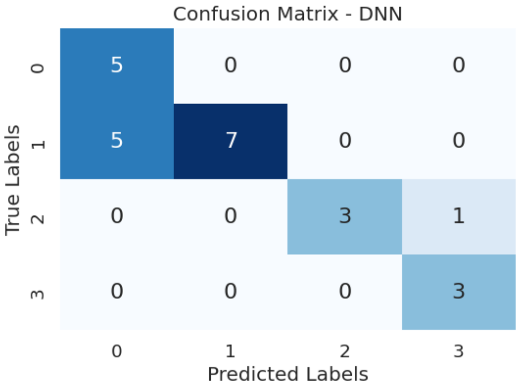

Figure 9. Confusion Matrix of DNN

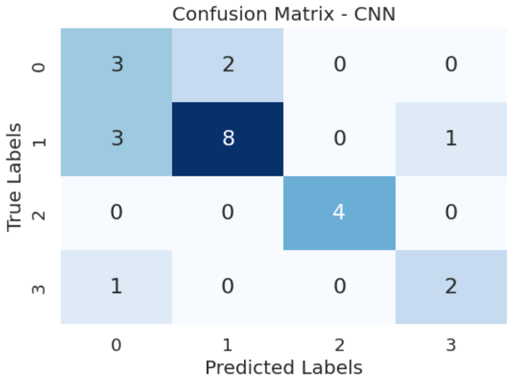

Figure 10. Confusion Matrix of CNN

As presented in the figure 4-10 about confusion matrix, The MLP/Neural Network matrix shows it performs flawlessly in predicting the first and fourth conditions, yet it confuses the first condition with the second, as evidenced by five instances incorrectly classified. This pattern suggests the MLP might be grappling with distinguishing between these two conditions due to overlapping symptoms or features that it interprets as similar. Moving on to the LSTM, its confusion matrix portrays a balanced predictive ability across all conditions, but it particularly excels in identifying the second and fourth conditions. Misclassifications, however, are notably present, especially when distinguishing the first condition, suggesting the LSTM might benefit from further refinement to better capture the temporal dynamics inherent in the symptom data. The confusion matrix for the CNN Sequence model indicates a dispersion of correct predictions alongside several misclassifications, signaling that while the model has learned certain patterns, it has not achieved a high level of clarity in its predictions. The spread of misclassifications across different conditions suggests challenges with the model's ability to consistently discern the sequential patterns unique to each condition. In contrast, the CNN 1D model shows a commendable job in accurately predicting the second condition but less precision with others. This may indicate that while the CNN 1D model is adept at identifying specific features of certain conditions, its generalizability across a broader spectrum of conditions is somewhat limited. The RNN's confusion matrix reveals a model well-tuned to the second condition but shows distributed errors with others, hinting that the RNN may not fully utilize its sequential data processing capabilities to differentiate between the conditions effectively.

The DNN's matrix resembles that of the MLP/Neural Network, with precise identification of the first and fourth conditions but confusion between the first and second. This suggests that while the DNN can recognize some conditions with high accuracy, it shares a common area of confusion with the MLP in differentiating between conditions that may present with similar features. Lastly, the CNN model's confusion matrix displays a good balance of correct predictions, particularly with the second condition, but also reveals some misclassifications, especially in distinguishing the first condition. The CNN demonstrates a better ability to differentiate between conditions compared to the MLP, yet it still shows room for improvement. To sum up, while each model exhibits specific strengths, particularly in identifying certain conditions, they all share challenges in differentiating between conditions with similar symptom presentations. The MLP and DNN models are strong in certain areas but share common points of confusion. The LSTM and RNN models, designed to process sequential data, show promise but require further tuning. The CNN models, including the 1D and Sequence, show capability in feature extraction but struggle with generalizing across all conditions. These insights suggest that an integrated approach, possibly an ensemble of models, could enhance diagnostic accuracy, and point to a need for additional features or data to help the models distinguish more effectively between overlapping symptoms.

4. CONCLUSION

Through an exhaustive investigation of neural network architectures, this research has highlighted their diverse capabilities and limitations in classifying mental health conditions. The intricate patterns of symptomatology pose a distinct challenge, one that current models like MLP/Neural Network, LSTM, and various CNNs approach with varying degrees of success. Notably, while MLP/Neural Network and DNN exhibit commendable accuracy, they, like their counterparts, grapple with conditions bearing similar symptoms. This study concludes that while individual architectures have their merits, future endeavors should focus on ensemble approaches and sophisticated feature engineering to refine diagnosis. Enhancing model interpretability and expanding datasets will also be crucial steps toward the goal of deploying these deep learning tools in practical, clinical settings for mental health diagnostics.

REFERENCES

[1] W. H. Organization, "The World Health Report 2001: Mental health: new understanding, new hope," 2001.

[2] S. Bajaj *et al.*, "South Asian women with diabetes: psychosocial challenges and management: consensus statement," *Indian J. Endocrinol. Metab.*, vol. 17, no. 4, p. 548, 2013.

[3] S. Benatar and G. Brock, *Global health: ethical challenges*. Cambridge University Press, 2021.

[4] W. H. Organization and others, "World mental health report: transforming mental health for all," 2022.

[5] D. Lee, Y. Kim, and B. Devine, “Spillover effects of mental health disorders on family members’ health-related quality of life: Evidence from a US sample,” *Med. Decis. Mak.*, vol. 42, no. 1, pp. 80–93, 2022.

[6] Y. Shemesh and A. Chen, “A paradigm shift in translational psychiatry through rodent neuroethology,” *Mol. Psychiatry*, vol. 28, no. 3, pp. 993–1003, 2023.

[7] D. Taliaz and D. Souery, “A new characterization of mental health disorders using digital behavioral data: Evidence from major depressive disorder,” *J. Clin. Med.*, vol. 10, no. 14, p. 3109, 2021.

[8] J. J. Newson, D. Hunter, and T. C. Thiagarajan, “The heterogeneity of mental health assessment,” *Front. psychiatry*, vol. 11, p. 76, 2020.

[9] T. Richter, B. Fishbain, G. Richter-Levin, and H. Okon-Singer, “Machine learning-based behavioral diagnostic tools for depression: advances, challenges, and future directions,” *J. Pers. Med.*, vol. 11, no. 10, p. 957, 2021.

[10] B. J. Coombes *et al.*, “Dissecting clinical heterogeneity of bipolar disorder using multiple polygenic risk scores,” *Transl. Psychiatry*, vol. 10, no. 1, p. 314, 2020.

[11] H. Li, R. Zhang, Y.-C. Lee, R. E. Kraut, and D. C. Mohr, “Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being,” *NPJ Digit. Med.*, vol. 6, no. 1, p. 236, 2023.

[12] R. S. McIntyre *et al.*, “The clinical characterization of the adult patient with bipolar disorder aimed at personalization of management,” *World Psychiatry*, vol. 21, no. 3, pp. 364–387, 2022.

[13] S. Vatansever *et al.*, “Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions,” *Med. Res. Rev.*, vol. 41, no. 3, pp. 1427–1473, 2021.

[14] C. Sarkar *et al.*, “Artificial intelligence and machine learning technology driven modern drug discovery and development,” *Int. J. Mol. Sci.*, vol. 24, no. 3, p. 2026, 2023.

[15] N. Chafai, L. Bonizzi, S. Botti, and B. Badaoui, “Emerging applications of machine learning in genomic medicine and healthcare,” *Crit. Rev. Clin. Lab. Sci.*, pp. 1–24, 2023.

[16] K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh, and J. Zou, “From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment,” *Cell*, 2023.

[17] E. Lin, C.-H. Lin, and H.-Y. Lane, “Deep learning with neuroimaging and genomics in Alzheimer’s disease,” *Int. J. Mol. Sci.*, vol. 22, no. 15, p. 7911, 2021.

[18] C. Su, Z. Xu, J. Pathak, and F. Wang, “Deep learning in mental health outcome research: a scoping review,” *Transl. Psychiatry*, vol. 10, no. 1, p. 116, 2020.

[19] C. Gupta *et al.*, “Bringing machine learning to research on intellectual and developmental disabilities: Taking inspiration from neurological diseases,” *J. Neurodev. Disord.*, vol. 14, no. 1, p. 28, 2022.

[20] A. M. Antoniadi *et al.*, “Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review,” *Appl. Sci.*, vol. 11, no. 11, p. 5088, 2021.

[21] G. B. D. 2019 M. D. Collaborators and others, “Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990--2019: a systematic analysis for the Global Burden of Disease Study 2019,” *The Lancet Psychiatry*, vol. 9, no. 2, pp. 137–150, 2022.

[22] J. Sun *et al.*, “Artificial intelligence in psychiatry research, diagnosis, and therapy,” *Asian J. Psychiatr.*, p. 103705, 2023.

[23] A. Saka *et al.*, “GPT models in construction industry: Opportunities, limitations, and a use case validation,” *Dev. Built Environ.*, p. 100300, 2023.

[24] S. Morande, “Enhancing psychosomatic health using artificial intelligence-based treatment protocol: A data science-driven approach,” *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100124, 2022.

[25] A. Abrol *et al.*, “Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning,” *Nat. Commun.*, vol. 12, no. 1, p. 353, 2021.

[26] A. Grzenda, “Artificial intelligence in mental health,” *Converg. Ment. Heal. A Transdiscipl. Approach to Innov.*, p. 137, 2021.

[27] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, “Deep learning: Systematic review, models, challenges, and research directions,” *Neural Comput. Appl.*, vol. 35, no. 31, pp. 23103–23124, 2023.

[28] Y. Shao, “Imbalance Learning and Its Application on Medical Datasets,” Dissertation, Göttingen, Georg-August Universität, 2021, 2021.

[29] Programmerrdai, “Mental Disorder Classification.” 2021.