Exploring Alternative Approaches to Diagnosing Mental Health Disorders Using Neural Networks

Zachary A. Cherry

University of Wisconsin - Madison

ABSTRACT

This project investigates the potential application of feed-forward neural networks in reversing the current diagnostic process for mental health disorders, with a specific focus on Mania Bipolar Disorder, Depressive Bipolar Disorder, and Major Depressive Disorder. This investigation is done utilizing a dataset comprising 120 psychology patients to determine the feasibility of accurately diagnosing these mental health conditions based on symptom patterns and behaviors not explicitly covered by the DSM-5 diagnostic criteria.

Keywords: Feed-Forward, Machine-Learning, Diagnosis, Bipolar, Depression

INTRODUCTION

Mental health diagnosis traditionally relies on standardized criteria outlined in the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders). This method, focusing primarily on categorical symptoms and behaviors aligned with specific diagnostic criteria, has provided a structured framework for understanding and treating mental health conditions. Despite its effectiveness, this approach has limitations, particularly in diagnosing conditions with atypical or less clearly defined symptom profiles. There has been a growing interest in leveraging more advanced computational techniques to address these limitations and enhance diagnostic accuracy.

Recent advancements in machine learning, especially neural networks, offer a promising alternative for improving mental health diagnoses. Neural networks can learn complex relationships and patterns in data, which traditional diagnostic methods may miss. This project explores whether a feed-forward neural network (FNN) can classify mental disorders based on a broader range of symptoms and behaviors than those covered by the DSM-5. The primary goal is to assess the feasibility of using neural networks to diagnose conditions such as Mania Bipolar Disorder, Depressive Bipolar Disorder, and Major Depressive Disorder. A control group was also added to determine if the model can discern individuals without a formal diagnosis.

The dataset utilized in this study includes information from 120 psychology patients, with recorded symptoms encompassing mood swings, suicidal thoughts, exhaustion, euphoria, and sleep disorders, among others. Analyzing these symptoms through a machine learning model aims to determine if researchers can develop a more comprehensive diagnostic tool. Neural networks, with their ability to process relationships in complex data through multiple layers, offer several advantages for this task. This project seeks to integrate these advanced techniques into mental health diagnostics, potentially offering new insights and improving the overall accuracy of patient diagnoses.

METHODS AND MATERIALS

The dataset used in this study is derived from a private psychology clinic at Harvard University, where Dr. Hengameh Karbalaeipour collected the data. The data collection focused on mental disorder classification, encompassing data from 120 psychology patients. Key variables include binary indicators for symptoms such as mood swings, suicidal thoughts, and anorexia. Multi-class variables, ranked as Most Often, Usually, Sometimes, or Seldom, include measures of sadness, euphoria, exhaustion, and sleep disorders. Ordinal variables, ranked from 1 to 10, assess levels of sexual activity, concentration, and optimism. Additionally, the dataset includes an identifier for each patient and an expert diagnosis label, which categorizes patients into Bipolar Type-1 (n = 28), Bipolar Type-2 (n = 31), Depression (n = 31), or No Diagnosis (n = 30).

The dataset was randomly split into training and validation sets to evaluate the model's performance. Specifically, 80% of the data (n = 96) was used for training the model, while the remaining 20% (n = 24) was reserved for validation purposes. The split was stratified to ensure that the proportions of each diagnostic category were consistent across both the training and validation sets. This approach helps to maintain the representation of the smaller classes and ensures that the model is exposed to a balanced distribution of diagnoses during training and validation.

Neural Network Model

A feed-forward neural network (FNN) was chosen for this classification task due to its effectiveness in handling structured input data and its capacity to model complex relationships between input features and output classes. FNNs are well-suited for tasks where the data is organized in a tabular format, such as in this case, with 17 input features. The architecture's simplicity and flexibility make it an appropriate choice for tasks that do not require sequential or spatial data processing, such as time series or image data.

In our network, the hidden layers use the Rectified Linear Unit (ReLU) activation function. ReLU is defined as $f(x) = \max(0, x)$, where the output is the input value if it is positive, and zero otherwise. The choice of ReLU is motivated by several factors:

- **Non-Linearity:** ReLU introduces non-linearity into the model, allowing it to learn complex patterns and representations in the data. Without non-linearity, the network would function as a linear model, which is inadequate for capturing intricate relationships.
- **Avoiding Vanishing Gradient Problem:** ReLU mitigates the vanishing gradient problem that affects traditional activation functions like sigmoid or tanh, where gradients become very small and hinder learning. ReLU provides a constant gradient of 1 for positive inputs, facilitating effective weight updates during training.
- **Computational Efficiency:** ReLU is computationally efficient due to its simple operations (comparison and thresholding), which leads to faster training times compared to more complex activation functions.

The FNN architecture features an input layer with 17 nodes, two hidden layers with 64 and 32 units, respectively, each using ReLU activation, and a dropout rate of 0.3 after each hidden layer to prevent overfitting. The output layer consists of 4 units with a softmax activation function, providing class probabilities for classification.

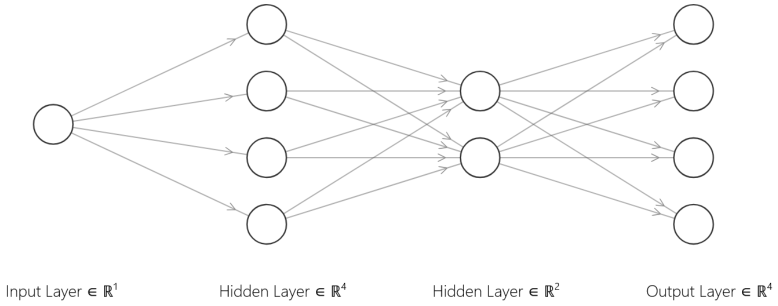
Input Layer ∈ $\mathbb{R}^1$
Hidden Layer ∈ $\mathbb{R}^4$
Hidden Layer ∈ $\mathbb{R}^2$
Output Layer ∈ $\mathbb{R}^4$

Figure 1. Simplified illustration of the neural network architecture, the figure provides a visual representation of the neural network with only 1 input unit.

RESULTS

To optimize the performance of the feed-forward neural network, we utilized the Weights and Biases (WandB) library for hyperparameter tuning. WandB provided comprehensive tools for tracking and managing the hyperparameter search process. The optimal hyperparameters identified included: a learning rate of 0.001, a batch size of 32, and a dropout rate of 0.3. The model was configured to run for 150 epochs and included two hidden layers, the first with 64 units and the second with 32 units. The best-performing neural network model was evaluated on a validation set to assess its generalization capabilities. The model achieved a validation accuracy of 95.83%, indicating that it correctly classified 95.83% of the samples in the validation set. This high accuracy demonstrates the model's effectiveness in predicting the correct class labels for the unseen data.

In addition to accuracy, the model's performance was assessed using the validation loss, which was 0.11. The validation loss quantifies the discrepancy between the predicted probabilities and the actual class labels. A lower validation loss signifies that the model's predictions are closer to the true labels. In this case, a validation loss of 0.11 suggests that the model's predictions are pretty accurate and that the model has a well-calibrated probability distribution across the classes.

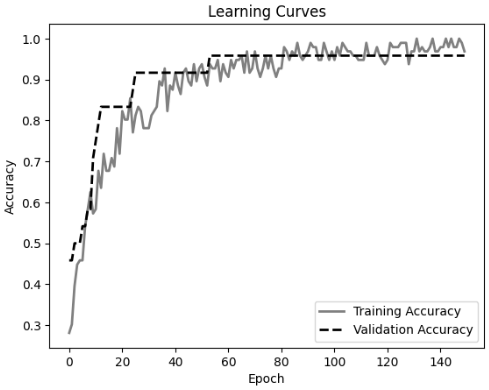

Figure 2. Accuracy plot

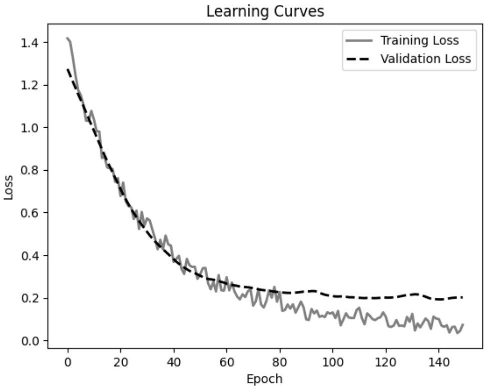

Figure 3. Loss plot

A confusion matrix for the test set was created to assess the performance of the trained neural network. This matrix provides a detailed view of the classification results by showing the counts of true positive, false positive, true negative, and false negative predictions across all classes. The overall accuracy of the model was calculated to be 96%, indicating the proportion of correctly classified instances out of the total number of instances in the test set.

DISCUSSION

The results of this study demonstrate the potential of feed-forward neural networks (FNNs) in improving the diagnostic process for mental health disorders. The model achieved an impressive validation accuracy of 95.83%, highlighting its effectiveness in classifying the mental health conditions within the dataset. This high accuracy suggests that the FNN can learn complex patterns from the symptom data and accurately predict class labels for new, unseen data.

The validation loss 0.11 further supports the model's robustness, indicating that the predicted probabilities are close to the actual class labels. This low loss value reflects a well-calibrated model, which is crucial for ensuring reliable diagnostic predictions. The confusion matrix, with an overall accuracy of 96%, reveals the model's strong performance across different classes. This accuracy indicates that the model is proficient at distinguishing between Manic-Depressive Disorder, Depressive Bipolar Disorder, and Major Depressive Disorder, as well as individuals without a formal diagnosis. The per-class metrics, including precision, recall, and F1-score, provide additional insights into the model's performance. These metrics help identify areas where the model excels and areas that may require further refinement.

While the results are promising, there are several considerations for future research. First, the dataset used in this study is relatively small, with only 120 patients. A more extensive and more diverse dataset could improve the model's generalization and robustness. Exploring other neural network architectures and hyperparameter settings might yield even better performance. Evaluating the model's performance on external datasets is essential to ensure its applicability to different populations and clinical settings.

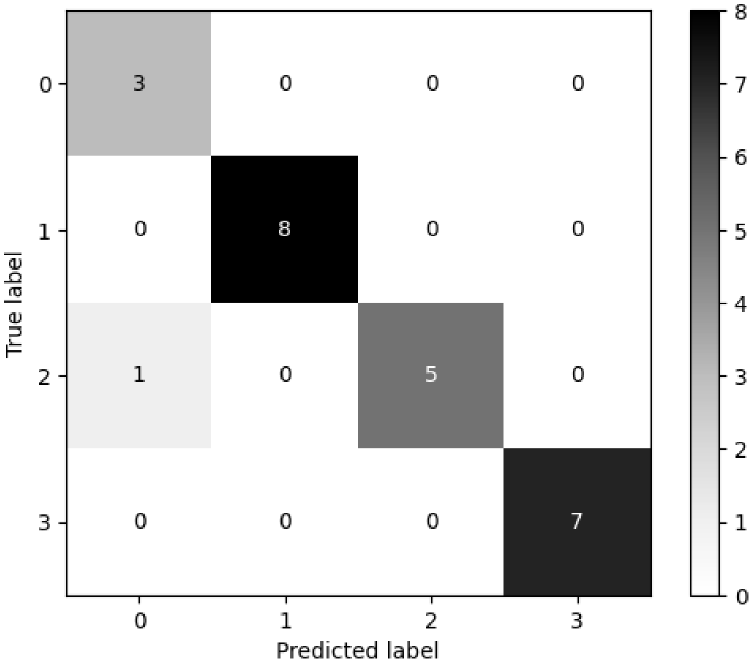

Figure 4. Confusion matrix for the test set, illustrating the performance of the neural network model.

CONCLUSION

This study demonstrates the effectiveness of feed-forward neural networks in diagnosing mental health disorders based on a broad range of symptoms and behaviors. The model's high accuracy and low validation loss indicate its potential as a valuable tool for mental health diagnostics. By leveraging advanced machine learning techniques, this research provides a new perspective on improving diagnostic accuracy and addressing the limitations of traditional diagnostic methods.

The success of the neural network model highlights the promise of incorporating computational techniques into mental health diagnosis. Future work should expand the dataset, explore alternative neural network architectures, and validate the model on external datasets to enhance its applicability and robustness. Overall, this study contributes to the growing body of evidence supporting the integration of machine learning into mental health diagnostics, offering new opportunities for improving patient care and diagnostic precision.

REFERENCES

[1] American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., pp. 123-168). Arlington, VA: American Psychiatric Publishing.

[2] Matravers, M. (2011). Classification, morality and the DSM. *Personality and Mental Health, 5*(2), 152-158.

[3] Pickersgill, M. D. (2014). Debating DSM-5: Diagnosis and the sociology of critique. *Journal of Medical Ethics, 40*(8), 521-525.

[4] Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials, 13*(4), 27-31.

[5] Oweimieotu, A. E., Akazue, M. I., & Edje, A. E. Designing a hybrid genetic algorithm trained feedforward neural network for mental health disorder detection.

[6] Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems, 39*(1), 43-62.