Mental Disorder Classification

Abdulrahman Emad
Systems and Biomedical Engineering
Cairo University
Cairo, Egypt
abdulrahman.masoud02@eng-st.cu.edu.eg

Mourad Magdy
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
murad.ibrahim02@eng-st.cu.edu.eg

Youssef Ashraf Mohammed
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
youssef.aziz02@eng-st.cu.edu.eg

Mariam Ahmed Said
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
maryam.abdulmajeed01@eng-st.cu.edu.eg

Abstract*—Mental health disorders pose significant challenges for accurate and timely diagnosis. In this paper, we explore the application of machine learning techniques for classifying mental health conditions based on a dataset collected from a private psychology clinic comprising 120 psychology patients exhibiting 17 essential symptoms. Our objective is to distinguish between Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, and individuals considered normal, who utilize therapy for personal development and minor mental health issues. We implemented three machine learning models: Random Forest, Support Vector Machine (SVM), and Logistic Regression. The Random Forest model was our best model as it achieved a training accuracy of 94.79% and a testing accuracy of 91.67%, with an average f1-score of 0.86. Our results demonstrate the efficacy of these models in the classification of complex mental health conditions.The model can benefit not only healthcare providers but also employers in the tech industry. Tech employers can utilize this classification model to predict their employees' mental health status and provide targeted resources to address relevant symptoms. The larger-scale impact is a focused effort on reducing mental illness across industries and workplaces, with a broader goal of decreasing mental illness globally.

Keywords—Mental Health, Healthcare, Mental Health Classification, Machine Learning, Depression, Bipolar Disorders, AI in medical Field

I. Introduction

Mood is known as a conscious state of mind or a sustained feeling that influences a person's thoughts, behavior, and interactions. Mood disorders are disruptions in emotions, either "lows," which are known as depression, or "highs," known as hypomania and mania.

Mood disorders are classified into different categories based on the symptoms; In this paper, we will just focus on major depressive disorder, mania bipolar disorder (Bipolar I), and depressive bipolar disorder (Bipolar II).

Major Depressive Disorder, also known as unipolar depression, is diagnosed according to the DSM by the presence of five of nine symptoms of a sad mood with other symptoms like recurrent suicidal ideation or acts of self-harm that should exist for at least 2 weeks.

Bipolar Disorder is different from Major Depressive disorder, because it has two extremes of highs and lows, hence the name, in order to be diagnosed with Bipolar disorder the patient must show signs of both "Mania" which is an extreme heightened state and "Depression" which is an extreme low state, the symptoms of mania can be summarized into an elevated ,or in some cases irritable mood with other symptoms that have to last for at least a week or have to be so severe that the patient needs hospitalization, which leads us to the subcategories of the disorder, like Bipolar Disorder I or Manic Bipolar, which is characterized by full manic episodes along with the depression symptoms, and Bipolar Disorder II, which is characterized with milder, non-psychotic form of mania which is hypomania that lasts for 4 days on average and more severe depressive episodes.

The factors contributing to mood disorders aren't just psychological, which is what's commonly known; to mention a few, neuroimmunological factors like Nitric Oxide contribute to inflammatory processes that contribute to the development of mood disorders. Enlargements in the brain regions responsible for emotions and feelings in the brain, which are the amygdala and the orbitofrontal cortex, were found in the brain images of the patients that were diagnosed by a professional. As for biological factors, they are typically due to an imbalance in neurotransmitters, where low levels of serotonin, dopamine, and norepinephrine account for symptoms of depression and high levels of dopamine account for symptoms of mania. [1]

Mental disorders are as serious as physical disorders, where the social and economic effects of mood disorders include functional impairments, disabilities, or the inability to be productive.[2]

According to the World Health Organization, "an estimated 3.8% of the population experience depression, including 5% of adults (4% among men and 6% among women) and 5.7% of adults older than 60 years. Approximately 280 million people in the world have depression." These statistics are a cause for concern and a motive to develop more effective methods to diagnose and treat mood disorders.[3]

With the advances in technology and artificial intelligence, specifically machine learning the, idea of using machine learning as a asset to health professional to increase the quality of healthcare became more popular by the day, thus the incorporation of AI into the field of mental health might account to mitigating the accounts of misdiagnosis of bipolar disorder, as according the NIH "69 percent of patients with bipolar disorder are misdiagnosed initially and more than one-third remained misdiagnosed for 10 years or more"[4], which is a cause for concern as the treatment of Bipolar disorder differs from that of major depression disorder, one example being that bipolar disorders are prescribed mood regulators, while patients which major depressive disorders are prescribed antidepressants medications.

In this paper, we are working on a machine learning model to classify major depressive disorder, manic bipolar, which is bipolar I, and depressive bipolar, which is bipolar II.

II. Related Work

A. Introduction

A review of the literature and research survey examining the applications of machine learning for mental health shows significant growth in this area in recent times. Current performance measures about machine learning research outcomes such as accuracy, precision, recall, and F1 scores-are widely used in these studies. key motivators for the use of machine learning in predicting mental health include access to behavioral data, the benefits of automated data processing, and data-driven assessments, all of which can enhance clinical health outcomes and decision-making.

B. Supervised Learning Techniques for Mental Health Classification

Supervised learning, which uses labeled data for training models, is essential for predicting mental health disorders. Algorithms like Support Vector Machines (SVM), Decision Trees, Naïve Bayes, K-nearest neighbor (KNN), and Logistic Regression have successfully categorized mental health states using questionnaire data. For example, Bayesian networks from health record questionnaires have shown varied accuracy in predicting depression in the elderly, showcasing their adaptability in mental health analysis.

Key Supervised Learning Techniques:

- **Support Vector Machines (SVM):** Suitable for binary classification, separating classes with an optimal margin, with kernel selection being crucial.
- **Gradient Boosting Machine (GBM):** Combines weak learners iteratively to form a strong model, achieving high accuracy but is computationally demanding.
- **Random Forest:** An ensemble of decision trees improving robustness and reducing overfitting, though large ensembles are computationally intensive.
- **Naïve Bayes:** Assumes feature independence and performs well in specific scenarios, valuable for mental health predictions.
- **K-Nearest Neighbor (KNN):** Labels data based on nearest neighbors, with no training phase required but computationally demanding for large datasets.

  These models have demonstrated varying levels of classification accuracy, with the highest achieving 83.51% correctly classified instances and an AUC of 0.798.

III. Dataset and Features

A. Dataset Description

The dataset we used is a collection of 120 Psychology patients with 17 essential symptoms collected from a private psychology clinic and the source was **Harvard Dataverse.** We used it to diagnose Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, and normal individuals. The Normal category refers to the individuals using therapy time for specialized counseling, personal development, and life skill enrichment. While such individuals may also have minor mental problems, they differ from those suffering from Major Depressive Disorder and Bipolar Disorder.

TABLE I. DATASET FEATURES DESCRIPTION

No	Feature	Description
1	Sadness	The feeling of unhappiness or sorrow,
2	Euphoric	Intense feelings of happiness, excitement, or joy
3	Exhaustion	Extreme tiredness or fatigue, often resulting from physical or mental exertion.
4	Sleep Disorder	Disruption in normal sleep patterns, which can manifest as difficulty falling asleep, staying asleep, or experiencing restful sleep.
5	Mood Swing	Rapid and intense changes in emotional states
6	Suicidal Thoughts	Persistent thoughts about ending one's own life, which can range from fleeting to obsessive.
7	Anorexia	An Eating disorder characterized by an extreme fear of gaining weight and a distorted body image.
8	Authority Respect	Acknowledgment and deference towards figures of authority, such as parents, teachers, or bosses.
9	Try-Explanation	A willingness or effort to provide justification or clarification for one's actions or decisions.
10	Aggressive Response	Reacting with hostility, anger, or forcefulness towards perceived threats or challenges.
11	Ignore & Move-On	Choosing to disregard or overlook something perceived as negative or unimportant, in order to focus on more productive or positive aspects of life.
12	Nervous Breakdown	Severe mental or emotional collapse, often characterized by overwhelming stress, anxiety, or depression.
13	Admit Mistakes	Acknowledging errors or faults in one's actions or judgment is often a step toward personal growth, learning, and accountability.
14	Overthinking	Excessive rumination or analysis of thoughts, events, or situations.
15	Sexual Activity	Engagement in intimate or sexual behaviors

16	Concentration	Ability to focus one's attention and mental resources on a specific task, topic, or goal.
17	Optimism	A positive outlook or attitude towards life.

B. Data Preprocessing

To make sure our dataset was high-quality and useful, we went through a number of preparation processes before starting the model training process. Below is an outline of these steps. Initially, we looked for any missing values in the dataset. Null values can have a detrimental effect on the performance of the model and cause inaccurate conclusions to be drawn. We were able to proceed without the necessity for imputation or record removal because our investigations showed that the dataset had no missing values. We then evaluated the dataset's balance. Training robust and unbiased models requires approximately equal numbers of samples in each class, which is ensured by a balanced dataset. We verified the data set's balance, which reduced the possibility of model bias toward any specific class. We transformed the category variables into numerical representations in order to get the categorical data ready for analysis. For algorithms used in machine learning, this stage is essential. We implemented two different encoding strategies:

A. **One-Hot Encoding:** In this approach, each category of a categorical variable is represented as a binary vector. For instance, a categorical variable with three categories would be transformed into three binary features. This method is particularly useful when the categorical variable does not have an inherent order. The resulting dataframe from this process is referred to as the "one-hot encoded dataframe."

B. **Multiclass Encoding:** Here, each category is assigned a unique integer value. This technique, also known as label encoding, is efficient in terms of memory usage and is suitable for algorithms that can naturally handle ordinal relationships. The dataframe generated using this method is termed the "multiclass encoded dataframe."

C. Exploratory Data Analysis (EDA)

This figure illustrates the distribution of sadness scores for individuals across four categories: Normal, Bipolar I, Bipolar II, and Depression.

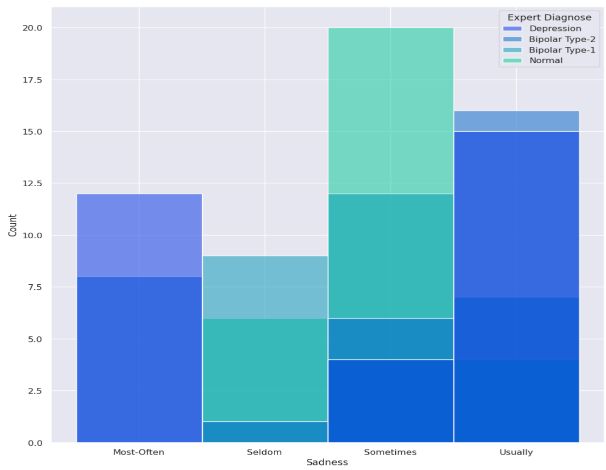


Figure: Sadness feature distribution

Normal: Individuals in the normal category have sadness scores that are generally lower, with most scores clustered towards the lower end of the scale, indicating low levels of sadness.

Bipolar I: The distribution of sadness scores for Bipolar I shows a broader range, reflecting the extreme mood swings characteristic of this disorder. The scores are more spread out, indicating variability in sadness levels.

Bipolar II: Similar to Bipolar I, Bipolar II also displays a wide distribution of sadness scores, but with a slightly higher concentration in the mid to upper range, indicating moderate to high levels of sadness during depressive episodes.

Depression: The distribution for depression is heavily skewed towards higher sadness scores, indicating that individuals with depression report consistently high levels of sadness.

IV. Methods

A. Our Learning Algorithms

In this paper, we implemented three machine learning algorithms: Random Forest, Support Vector Machine (SVM), and Logistic Regression. Each of these algorithms was applied to the dataset to classify mental health conditions based on the 17 essential symptoms**,** The one with the best accuracy in the end was selected.

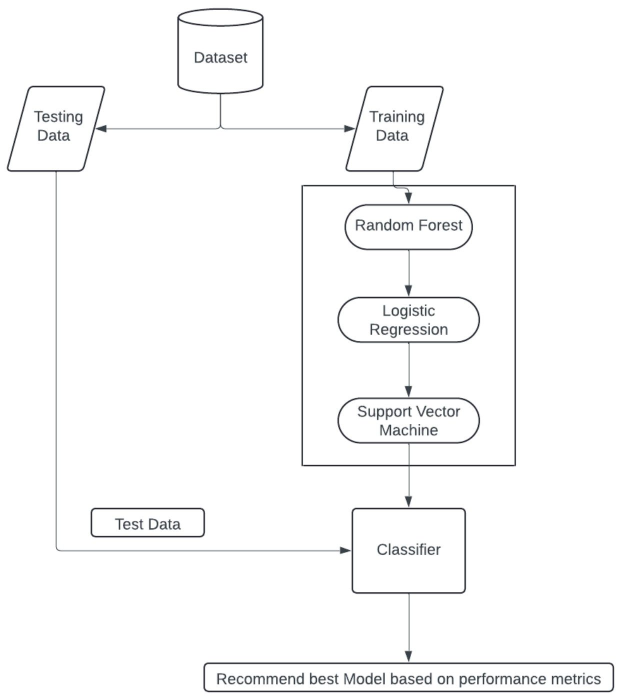
*1) **Random Forest**: An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. The algorithm operates by training each decision tree on a different subset of the training data and then aggregating their predictions.*

*2) **Support Vector Machine**: SVM is a supervised learning algorithm that finds the hyperplane which best separates the classes in the feature space. The algorithm works by maximizing the margin between the closest points of the classes (support vectors). In this paper, we used the radial basis function (RBF) kernel, which maps the input features into higher-dimensional space to handle non-linear relationships.*

*3) **Logistic Regression**: Logistic Regression is a linear model used for binary classification. It estimates the probability that a given input belongs to a certain class. The model uses the logistic function to transform linear combinations of the input features into probabilities.*

B. Methodology

We divide the dataset into 80% training data and 20% testing data . The data is fed into each model individually to determine which model results in higher performance, we also use cross validation to determine the best set of hyperparameters for each model. The block diagram as shown in Fig, shows the methodology used for finding the best model for mental health classification.



C. Mathematical Notations

a) Logistic Regression:

$$P(Y = 1 | x) = (1)/(1 + e^{-(wx+b)})$$

b) Random Forest:

$$y = mode(\{f\,1\,(x), f\,2\,(x), ..., f\,n\,(x)\})$$

c) Support vector Machine:

$$min(w, b)\, 0.5\, ||W|^{2} + C \sum_{i=1}^{N} N\, e(i)$$

Experiments & Results

The data were entirely categorical, which needed transformation into numerical, as machine learning algorithms, in their basic form, are mainly designed to operate on mathematical principles and deal with numbers, not categories, so we had to transform the data with an appropriate method so that the model would interpret the data better. As previously mentioned, we experimented with one hot encoding and multiclass encoding and experimented with three models: SVM, Random Forest, and Logistic Regression. With each of the resulting encoded data, we chose these three because they are designed to work for categorical data. This section summarizes the results of the experiments and their accuracy.

A. Experimenting with Hyperparameters

We experimented with hyperparameters on all models using grid search and all of them are using 5 K-Fold. the following table summarizes the finding along with the chosen hyperparameter for each model

Random Forest	Logistic Regression	Support Vector Machine
One Hot Encoded Data		
Best Max Depth = None, and Best Number of estimators = 200	Best C value = 0.01	Best C value = 0.1 and Best gamma = 1
Best Min Samples leaf = 1, and Best Min Samples Split = 10	Best Solver = liblinear	Best Kernel = poly
Multi-Class Encoded Data		
Best Max Depth = None, and Best Number of estimators = 100	Best C value = 0.1	Best C value = 100 and Best gamma = 0.001
Best Min Samples leaf = 1, and Best Min Samples Split = 10	Best Solver = newton-cg	Best Kernel = rbf

B. Evaluation Metrics

The Metrics used for evaluation were accuracy, precision, recall, f1-score, and AUC-ROC, we mainly focus on the AUC-ROC and f1-score as they give the best insights on how the model performs, as we can't just depend solely on the accuracy metric.

Accuracy: The proportion of correctly classified instances out of the total instances.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Population}$$

Precision: The proportion of positive predictions that are actually correct.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall: The proportion of actual positives that are correctly identified by the model.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F1-Score: The harmonic mean of precision and recall, balancing both metrics.

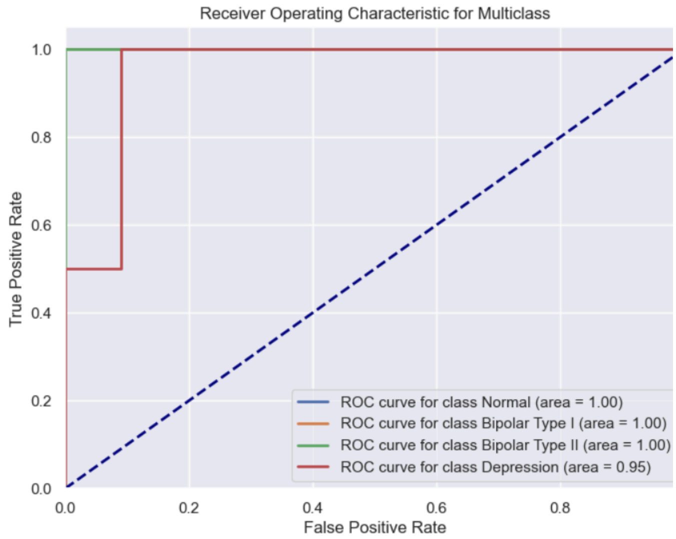
$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The Following Table Summarizes the performances of the models

Random Forest	Logistic Regression	Support Vector Machine
One Hot Encoded Data		
Train score: 97.92%	Train score: 96.875%	Train score: 100.0%
Test score: 87.50%	Test score: 75.0%	Test score: 83.33%

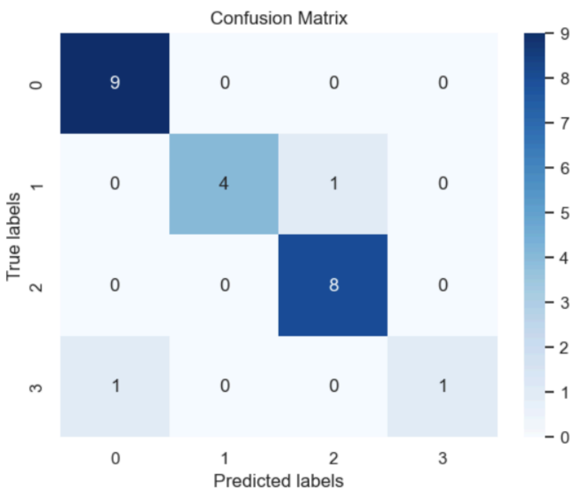
Multi-Class Encoded Data		
Train score: 94.79%	Train score: 94.79%	Train score: 93.75%
Test score: 91.67%	Test score: 87.5%	Test score: 87.5%

AUC



Confusion Matrix

Our best model was Random Forest, here is its confusion matrix



True Positives (TP): The number of instances where the true class and the predicted class are the same.

True Negatives (TN): The number of instances that are correctly predicted as not belonging to a particular class for all classes other than the one being considered.

False Positives (FP): The number of instances that are predicted to belong to a class but actually belong to a different class.

False Negatives (FN): The number of instances that are predicted not to belong to a class but actually belong to that class.

VI. Conclusion and Future Work

This paper highlights the application of machine learning techniques for classifying mental health conditions based on a dataset collected from a private psychology clinic. The Random Forest algorithm emerged as the highest performing model, achieving a training accuracy of 94.79% and a testing accuracy of 91.67%, with an average f1-score of 0.86. The efficacy of Random Forest can be attributed to its ability to handle complex relationships between symptoms and diagnoses, as well as its robustness to noise in the data. While Support Vector Machine (SVM) and Logistic Regression also showed promising results, Random Forest outperformed them, possibly due to its ensemble nature and inherent resilience to overfitting.

For future work, given additional time, team members and computational resources, we would try to expand the dataset to include more patient samples and additional features that could enhance the model's performance, we would also try to experiment with deep learning techniques, such as neural networks which might capture more patterns within the data, this would improve the classification of mental health.

Contributions

All authors contributed equally to this study, collaborating closely from inception to completion. Each member of the team participated in data collection, preprocessing, model selection and training, model evaluation, and result interpretation. The project benefited from the collective expertise and effort of all contributors, ensuring a comprehensive and cohesive approach to the research process.

Acknowledgment

We would like to extend our heartfelt appreciation to Dr. **Inas A. Yassine** and Eng. **Mirna Bibars**, our esteemed instructors in the course. Their guidance, support, and expertise have been invaluable throughout this project. We are deeply grateful for their mentorship, encouragement, and dedication to our learning and growth. Thank you for your unwavering support and inspiration.

References

[1] Sekhon, S., & Gupta, V. (2023, May 8). Mood disorder. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK558911.

[2] Sheline, Y. I. (2003). Neuroimaging studies of mood disorder effects on the brain. Biological Psychiatry, 54(3), 338–352. https://doi.org/10.1016/s0006-3223(03)00347-0

[3] 31). Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression

[4] Singh, T., & Rajput, M. (2006). Misdiagnosis of bipolar disorder. Psychiatry (Edgmont (Pa. : Township)), 3(10), 57–63.

[5] Olatunde, O., Tipu, A., & Falola, B. (2021). Classification of mental health disorders. ResearchGate. https://doi.org/10.13140/RG.2.2.34918.60483

[6] Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S. (2019). Review of Machine Learning Algorithms for Diagnosing Mental Illness. Psychiatry Investigation, 16(4), 262–269. https://doi.org/10.30773/pi.2018.12.21.2

[7] Vaishnavi, K., Kamath, U. N., Rao, B. A., & Reddy, N. V. S. (2022). Predicting Mental Health Illness using Machine Learning Algorithms. Journal of Physics. Conference Series, 2161(1), 012021. https://doi.org/10.1088/1742-6596/2161/1/012021