Machine Learning-based Differentiation of Bipolar Disorder-II and Unipolar Disorder Using Actigraph Data: A Comparative Analysis of Classical Models and literature survey

Sambridhi Deo
Email: deosambridhi@gmail.com

Abstract*—Bipolar disorder, affecting over 46 million people globally, is the sixth leading contributor to disability according to the 1990 World Health Organization figures. Patients diagnosed with bipolar disorder have a higher risk of mortality, with a lifetime suicide risk of 20 times greater than the general population. Furthermore, the risk and prevalence of ischemic heart disease are also higher in both male and female patients with bipolar disorder compared to the general population. Diagnosis and early prediction of bipolar disorder are crucial to prevent or delay its onset and improve clinical outcomes. Traditional diagnostic approaches, such as clinical interviews and physical evaluations, often lead to misdiagnosis due to the absence of clear biomarkers or blood tests. Machine learning algorithms applied to objective markers can potentially overcome prognosis uncertainty and improve diagnostic accuracy. In this study, we utilized science databases, did literature survey, and investigated the application of various classical algorithms to differentiate between bipolar and unipolar depression using actigraph data. Our analysis encompassed data preprocessing, feature engineering, model evaluation, hyperparameter tuning, and cross validation to identify the most accurate and reliable approach for bipolar disorder prediction. We also performed an in-depth exploratory data analysis to gain insights into the features of the dataset and the relationships between variables. Our results demonstrate the effectiveness of machine learning algorithms and provided promising results in distinguishing between bipolar disorder II and unipolar depression. This study provides a foundation for future work, including integrating with potential biomarkers. Ultimately, our research contributes to the differentiating of more effective diagnostic tool to prevent misdiagnosis between Unipolar Disorder and Bipolar Disorder II, potentially improving the lives of millions of patients worldwide.

Keywords: Machine Learning, Classification algorithms, Bipolar Disorder.

I. Introduction

Bipolar disorder (BD) is a major mood disorder characterized by recurrent episodes of depression and (hypo)mania. The earliest mentions of Bipolar in medical literature were made by Hippocrates, also referred to as the father of medicine. Hippocrates was the first person to document two moods- extremely high and extremely low [1]. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) described bipolar disorder as a group of brain disorders that cause extreme fluctuation in a person's mood, energy, and ability to function. Later on, DSM-5 states that in bipolar disorder, increased impulsivity or inattention is accompanied by elevated mood, grandiosity, and other specific bipolar features [2]. DSM-5 categorizes Bipolar disorder into three main subtypes–Bipolar I, Bipolar II, and Cyclothymic disorder. The two main subtypes are BD-I (manic episodes, often combined with depression) and BD-II (hypomanic episodes, combined with depression) while Cyclothymic disorder is a cyclic disorder that causes brief episodes of hypomania and depression [3].

According to mental health America, 1 in 40 American adults which is almost 4% of the population live with Bipolar disorder. With over 46 million people diagnosed with Bipolar Disorder globally [4], people with a first-degree relative diagnosed with BD are said to have a higher risk according to American Psychiatric Association(APA council reports, 2017). Bipolar Disorder is not only considered to be chronic but also disabling. Although it is less common than depression and anxiety, it is the 6th leading cause of disability worldwide [5]. The risk of significant mortality in patients diagnosed with any Bipolar disorder type is higher as the lifetime risk of suicide is 20 times more than the general population [6]. The risk and prevalence of Ischemic heart disease are also higher in Bipolar disorder patients than in the general population in both sexes, especially at younger ages [7]. This illness also carries high economic costs and was ranked as the sixth leading contributor to disability in 1990 World Health Organization figures [8]. A study concluded that when a woman is diagnosed with bipolar disorder at the age of 25, she might lose nine years of life, 12 years of normal health, and 14 years of effective functioning [9].

Bipolar disorder is characterized by two mood swings including emotional highs (mania or hypomania) and lows (depression). Though Mania and Hypomania are two distinct episodes, a person having these episode experiences similar types of symptoms. Both Mania and Hypomania include a few of these symptoms [10]:

These symptoms are associated with high morbidity and mortality. A study entitled "Suicide attempts in Bipolar Patients" did an analysis of a sample of 169 patients of Hospital

TABLE I: Bipolar Disorder symptoms

Mania/Hypomania	Depressive Episodes
Increase activity, or agitation	Depressed mood
Insomnia	Fatigue
Unusual talkative behavior	Decreased ability to think
Illogical racing thoughts	Less or excessive-sleep
Euphoria	Suicidal thoughts
Distractability	significant weight loss

Santiago Apostol diagnosed as having a bipolar I disorder (DSM-III-R), out of which 56 patients (33%) had 1 or more suicide attempts. The study concluded that Suicide attempts are highly prevalent in bipolar patients and are related to drug abuse, family history of affective disorders, misdiagnosis, and severe depressive episodes [11]. Thus Diagnosis and Prediction of Bipolar Disorder preferably at the Early stage are really significant in order to reduce the burden of bipolar disorder and improve the long-term outcome for patients [12]. Early detection of bipolar disorder can prevent or delay the onset of the disorder, and improve clinical outcomes in people who develop it [13]. The traditional way for Bipolar Disorder diagnosis was a combination of four major steps–suspecting based on depressive symptoms, following up with clinical interviews, and a case-based finding tool, and confirming the diagnosis [14]. Bipolar disorder is often misdiagnosed. Two surveys, one taken in 1994 and one taken in 2000, reveal little change in the rate of misdiagnosis [15].

II. Literature Survey

A. *Misdiagnosis of Bipolar Disorder: An Overview and related works*

Signs and symptoms of Bipolar Disorder overlap with other conditions such as Unipolar disorder, ADHD, substance use disorder, and anxiety disorders.

1) ADHD(attention-deficit/hyperactivity disorder) - A chronic condition including attention difficulty, hyperactivity, and impulsiveness [16].
2) Unipolar Disorder- A mental health disorder characterized by persistently depressed mood or loss of interest in activities, causing significant impairment in daily life [17].
3) Substance Use Disorder- A substance use disorder (SUD) is a mental disorder that affects a person's brain and behavior, leading to a person's inability to control their use of substances such as legal or illegal drugs, alcohol, or medications [18].
4) Anxiety Disorder - A mental health disorder characterized by feelings of worry, anxiety, or fear that are strong enough to interfere with one's daily activities [19]

Due to this reason, a correct diagnosis of Bipolar disorder is often challenging and might lead to misdiagnosis. Misdiagnosis of bipolar disorder can have serious consequences for patients, including delayed or inappropriate treatment, exacerbation of symptoms, and even harm due to inappropriate medication. For example, a person with bipolar disorder who is misdiagnosed with depression and treated with an antidepressant alone may experience a worsening of manic symptoms or rapid cycling. Similarly, a person with bipolar disorder who is misdiagnosed with ADHD and treated with stimulant medication may experience a worsening of manic or hypomanic symptoms [20] [21].
Related papers discussing the misdiagnosis of Bipolar disorder have been mentioned below:

1) The Compelling and Persistent Problem of Bipolar Disorder Disguised as Major Depression Disorder: An Integrative Review [22] : Ghaemi and associates [23]recently reported that 40 percent of a group of patients with bipolar disorder had previously received an incorrect diagnosis of major depression. Major Depression Disorder is a mood disorder that causes a persistent feeling of sadness and loss of interest. It affects how you feel, think, and behave and can lead to a variety of emotional and physical problems. The purpose of this research was to describe the current state of the science of the misdiagnosis of bipolar disorder, with the ultimate goal of improving psychiatric diagnostic workups including screening. To do this, An integrative review was conducted using standard criteria for evaluating research articles, and Clinically oriented, reliable, and valid screening tools for bipolar disorder also were reviewed.

2) Misdiagnosis of bipolar disorder in children and adolescents: A comparison with ADHD and major depressive disorder [24]: The objective of the research was to highlight the underdiagnosis vs overdiagnosis of bipolar disorder (BD) in children and adolescents and compare it with diagnoses of attention-deficit/hyperactivity disorder (ADHD) and major depressive disorder (MDD). For this Sixty-four children and adolescents (ages 7 to 18) treated in a community setting were systematically assessed for diagnostic and treatment histories. The best estimate consensus diagnosis was made using DSM-IV criteria. The results mentioned that ADHD was overdiagnosed as all the patients with ADHD had received the diagnosis, but this was on a small sample of the population. So, the results might not be as accurate as this for a large population. Majorly, The results showed that 33% of patients with BD were incorrectly diagnosed with MDD. Even in a such small sample, if 33% misdiagnosis is existing, that's a big number and needs to be solved. To further this research, Studies with larger samples need to replicate the method. While this paper couldn't list the strategies to reduce misdiagnosis, it effectively utilized DSM-5 criteria.

3) Strategies to reduce misdiagnosis of Bipolar Depression [23]: The author discusses complexities in the diagnosis of bipolar disorder, especially in distinguishing bipolar from unipolar depression. The author also classifies between major depression and bipolar disorder by highlighting that more mood lability, more motor retardation, and greater time spent sleeping are associated with Bipolar Disorder. Early age of onset, a high frequency of depressive episodes, a greater percentage of time ill, and a relatively acute onset or offset of symptoms are also suggested to be signifying bipolar disorder rather than major depression. When Bipolar Disorder patients are misdiagnosed with major depressive disorder and treated with antidepressants, it can exacerbate hypomania, and mania, and can be dangerous.

B. *Classifying Bipolar Disorder and Major Depressive Disorder*

1) A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data [25]: Around 37% of patients with BD who present after their first manic/hypomanic episode are nonetheless misdiagnosed as having MDD [25]. This research aims to develop a diagnostic approach to reduce the misdiagnosis of Bipolar Disorder as a major depressive disorder by utilizing online questionnaires and blood biomarker data. Based on the mental health questionnaire-9 response, researchers recruited a few participants to provide them with a dried blood sample for biomarker analysis. The dry blood sample was analyzed for neuropsychiatric biomarker levels using a validated targeted proteomic approach. And, Data was used from a delta study [26], an investigator-led study conducted by the Cambridge Centre for Neuropsychiatric Research (CCNR) at the University of Cambridge. The participants also went through clinical interviews for the diagnosis process. Then, Extreme Gradient Boosting and nested cross-validation were used to train and validate diagnostic models differentiating BD from MDD in participants who self-reported a current MDD diagnosis. The research was successful as the diagnostic algorithm accurately identified patients with BD in various clinical scenarios.

C. *Predicting Bipolar Disorder Risk Factors*

1) Clinical risk factors for bipolar disorders: A systemic review of prospective studies [27]: This research evaluates the prevalence, duration, clinical features, and predictive value of non-affective psychopathology as clinical risk factors for bipolar disorder. It was done by screening PubMed, CINAHL, PsycINFO, and other scientific databases. 16 published reports meeting selection criteria were found with varying study designs. Despite the difference in methods, findings about clinical factors of Bipolar disorder were consistent. Some of those were early-onset panic attacks and disorders, separation anxiety, and generalized anxiety disorders. Researchers concluded that clinical factors for bipolar disorder typically arise before syndromal onset and include anxiety and behavioral disorders. Further evaluation of the predictive value of such clinical risk factors can help identify the population at increased risk. This research should be further carried out by including interviews with doctors, their insights should be added.

2) Predicting Bipolar Disorder Risk Factors in Distressed Young Adults from Patterns of Brain Activation to Reward: A Machine Learning Approach [28]: This study was to apply multivariate pattern recognition to prodigy the severity of behavioral traits and symptoms of Bipolar disorder from patterns of whole-brain activation. Researchers acquired functional neuroimaging data from two independent samples of transdiagnostically recruited adults. Pattern recognition model performance in each sample was measured using correlation and mean squared error. The first sample predicted the severity of specific hypo/mania symptoms and other fed features. The region with the highest contribution to the model was the left ventrolateral prefrontal cortex. These findings and emerging technologies like machine learning could be further improved to provide neural biomarkers to aid the early identification of bipolar disorder risk, especially in young adults.

3) Predicting bipolar disorder on the basis of phenomenology implications for prevention and early intervention [29]: The delay between the first onset of symptoms and diagnosis is often worse for Bipolar disorder patients. Early detection has the potential to reduce the risk factors and save lives. It is not possible to anticipate who will develop bipolar disorder solely based on early phenomenology. This study was a literature review of the Bipolar prodrome using the medicine, Web of Sciences, and a hand search of relevant factors, and predictors. The most cited prodromal features according to the results of this study were lability, nonspecific,non-mood symptoms, and cyclothymic temperament. This study furthermore concluded that for accurate characterization and prediction of bipolar disorder, high-quality, research studies should be conducted with adequate control groups. Apart from that, the utilization of emerging technologies for studying biomarkers and early predictions would definitely facilitate the prompt.

D. *Signs and symptoms of Bipolar Disorder*

1) Symptoms and signs of the initial prodrome of bipolar disorder: a systematic review [30]: Signs and symptoms are extremely important to not ignore when it comes to illnesses like Bipolar Disorder. Misdiagnosis is often the result of not studying signs and symptoms correctly. This study utilizes the databases like PsycINFO, PubMed, EMBASE, etc to review the symptoms and signs of initial bipolar prodrome with an aim to identify potential clinical targets for early intervention. Most of the common symptoms were irritability and aggressiveness, sleep disturbances, depression and mania symptoms/signs, hyperactivity, anxiety, and mood swings. However, not every person who develops Bipolar disorder will experience every mentioned symptom. The interesting data highlighted was that the mean duration of prodrome ranges from 1.8 to 7.3 years. Though this study clustered the common symptoms and highlighted them nicely, the utilization of more datasets could make the study much better.

2) A factor analysis of signs and symptoms of Mania [31]: For analysis of signs and symptoms of mania, researchers first conducted a literature review on manic subtypes. Twenty signs and symptoms relevant to classic mania and the mixed bipolar state were selected and each of them was scored from 0 to 5. Some of them were decreased sleep, pressured speech, increased motor activity, euphoric mood, etc. The study cohort was a group of 237 patients with DSM-III-R-defined. 4 psychiatrists evaluated 95% of the patients between 2 to 5 days of hospitalization and rated them based on the features from 0 to 5. In this study, five independent factors representing dysphoric mood, psychomotor pressure, psychosis, increased hedonic function, and irritable aggression were identified. However, the conventional view of symptom factors of mania wasn't confirmed.

III. About Machine Learning

A. Introduction to Machine Learning

Machine learning, a subset of artificial intelligence, enables computers to learn from and make predictions based on data [32]. It has shown great promise in various fields, including the medical domain, where it contributes to improved diagnostics, personalized treatment, and the resolution of complex problems [33]. Applying machine learning to bipolar disorder research has the potential to enhance our understanding of the condition, improve diagnostic accuracy, and inform more effective treatment strategies. Our aim with this study is to reduce misdiagnosis.

B. Machine Learning in Medical Research

The medical field has adopted machine learning techniques to address a wide range of challenges, from predicting patient outcomes and identifying disease patterns to optimizing treatment plans and enhancing patient care. From medical diagnosis and prognosis to medical treatment, Machine learning has had promising results. Various machine learning techniques, including supervised learning, unsupervised learning, and reinforcement learning, have been employed in medical research to tackle specific tasks, such as image classification, natural language processing, and decision-making under uncertainty [34].

There are mainly three types of machine learning. In our study, we have focused on supervised learning.

- Supervised learning
- Semi-supervised learning
- Unsupervised learning

C. Supervised Learning and Classification

Supervised learning is a widely used machine learning technique, in which a model is trained on a dataset with labeled examples, learning to predict the output (or "label") based on the input features. According to a famous book in the Machine learning world "Pattern Recognition and Machine Learning", Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems [35]. Classification, a type of supervised learning, aims to assign input data to one of several predefined categories. In this type of learning, the model is fully trained using the training datasets and evaluated based on the test dataset before using it to perform on unseen data [36]. In the context of bipolar disorder research, classification models can be trained to differentiate between bipolar and unipolar depression using various data sources, such as actigraphy data or electronic health records.

Below some of the clasification learning algorithms have been listed:

1) Logistic regression: A statistical method used for binary classification tasks, it models the probability of an outcome based on input features. Logistic regression is easy to interpret and computationally efficient, making it a popular choice for many applications [37].

2) Support Vector Machines (SVM): A powerful method for classification and regression tasks that aims to find the optimal hyperplane separating two classes. SVMs can handle non-linear decision boundaries using kernel functions and are known for their excellent generalization ability [38].

3) Random Forest: An ensemble learning method that constructs a multitude of decision trees and combines their predictions to improve overall accuracy and reduce overfitting. Random Forests are robust to noise, outliers, and handle missing data well(What Is Random Forest? — IBM, n.d.).

4) K-Nearest Neighbors (KNN): A non-parametric method that classifies instances based on the majority vote of their k nearest neighbors. KNN is simple to understand and can be effective for small datasets. However, it can be computationally expensive for large datasets [39].

5) Naive Bayes: A probabilistic classifier based on Bayes' theorem that assumes independence among features, making it simple and efficient. Naive Bayes is often used for text classification and natural language processing tasks [40].

6) Artificial Neural Networks (ANN): A computational model inspired by the structure and function of biological neural networks, which can learn complex patterns and representations in data. ANNs consist of interconnected neurons organized in layers and can be trained using gradient-based optimization algorithms [41].

7) Gradient Boosting Machines (GBM): An ensemble learning method that sequentially builds weak learners (decision trees) and combines their predictions to improve overall performance. GBM is known for its excellent predictive accuracy and has become a popular choice for many machine learning competitions [42].

8) Deep Learning methods (e.g., Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory): Advanced machine learning techniques that can automatically learn hierarchical feature representations from raw data. Deep learning methods have been particularly successful in tasks involving images, text, and time-series data [43].

D. Feature Engineering and Data Preprocessing

Feature engineering plays a crucial role in the success of machine learning models, as it involves extracting relevant features from raw data to improve model performance [44]. In medical research, feature engineering can involve the extraction of meaningful information from complex data sources, such as medical images, time-series data, and patient demographics. Data pre-processing is another essential step in the machine learning pipeline which includes manipulation and dropping of the data, as it ensures that the input features are on a similar scale and handles missing or noisy data to improve model performance. This can be done with exploratory data analysis(EDA). The quality of data directly affects the accuracy and efficiency of the machine learning models. Thus, this is a crucial step.

E. Model Selection and Evaluation

A wide range of machine learning algorithms has been employed in medical research to address various classification tasks, including logistic regression, random forest, support vector machines (SVM), k-nearest neighbors (KNN), decision trees, and gradient boosting, among others [45]. These algorithms have unique strengths and weaknesses, and their performance can vary depending on the specific problem and dataset. Model selection should consider factors such as interpretability, complexity, and generalization performance. When evaluating machine learning models, it is essential to use appropriate performance metrics, such as accuracy, precision, recall, F1-score, and area under the curve (AUC), to ensure reliable and interpretable results [46]. In medical research, metrics like sensitivity and specificity are often used to assess the performance of diagnostic models. Additionally, techniques like cross-validation, hyperparameter tuning, and bootstrapping can be employed to obtain more robust estimates of model performance.

F. Feature Importance and Interpretability

Understanding the importance of different features in the model can provide valuable insights into the factors contributing to the differentiation between bipolar and unipolar depression. Techniques such as permutation importance and tree-based feature importance can be used to rank features by their contribution to the model's performance. This information can help identify key factors that influence the classification and inform further research and clinical practice.

Machine learning has the potential to significantly impact the medical field, particularly in the diagnosis and treatment of complex disorders such as bipolar disorder. By employing various machine learning techniques and carefully selecting and evaluating models, researchers can uncover meaningful insights and patterns that might not be easily discernible through traditional methods. The application of machine learning in bipolar disorder research can lead to improved diagnostic accuracy, more personalized treatment strategies, and a better understanding of the underlying factors that contribute to the condition. As machine learning techniques continue to advance, it is essential to explore and leverage their potential in medical research to drive innovation and improve patient outcomes.

IV. Data Preparation

A. About our Data

The dataset we have utilized for training and evaluating our model is "The depression dataset"([47]. Depression is a common and significant term when it comes to Bipolar Disorder. Depression is characterized by symptoms like sadness, feelings of emptiness, anxiety, sleep disturbance, etc. The severity of depression is determined by the number of symptoms, their seriousness, and duration, as well as the impacts on social and occupational function. Most frequently actigraph recordings of motor activity are considered an objective method for observing depression. Actigraph recordings are the result of actigraphy which is a method of measuring sleep parameters and average motor activity over a period of days to weeks using a noninvasive accelerometer. Often, the accelerometer is housed in a small device worn like a wristwatch [48].

The dataset contains two folders: one for the condition group and another for the control group. For each patient, a CSV file has been provided containing the actigraph data collected over time. The columns are timestamp (one-minute intervals), date (date of measurement), and activity (activity measurement from the actigraph watch). In addition, the MADRS scores are provided in the file "scores.csv". MADRS is the most commonly used measure in antidepressant efficacy trials [49]. It contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1:

working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

B. Exploratory Data Analysis

The depression dataset was imbalanced. The figure below shows the data distribution between Bipolar disorder II, Unipolar Disorder, and Bipolar disorder I patients.

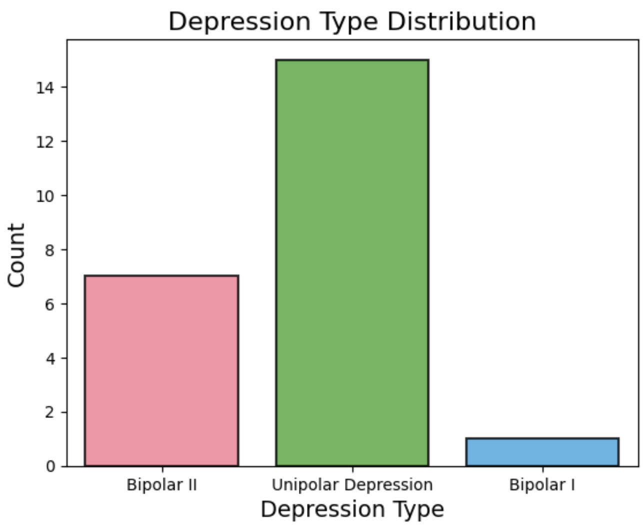

Fig. 1: Data Distribution based on Depression type.

Below are the figures showing distribution before data cleaning has been performed. Thus, they are imbalanced and might have missing values. Below are other charts showing the data distribution based on different key features.

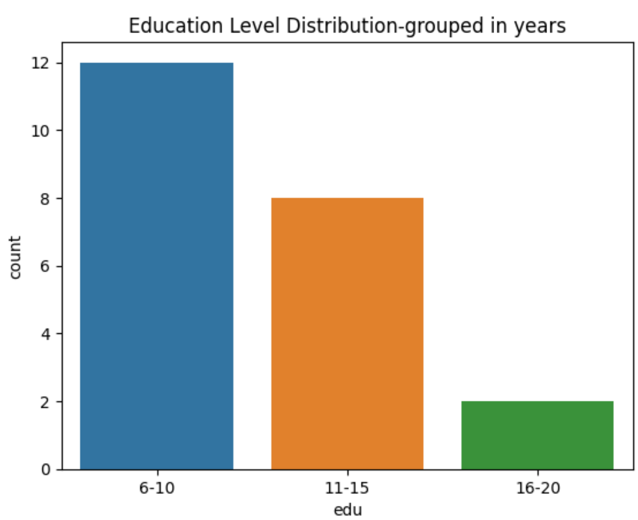

Fig. 2: Data Distribution based on Education.

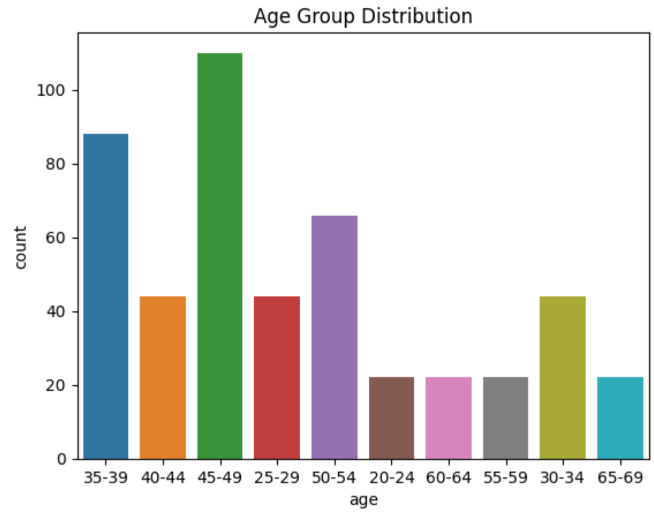

Fig. 3: Data Distribution based on Age.

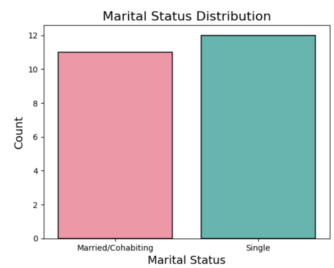

Fig. 4: Data Distribution based on Marriage.

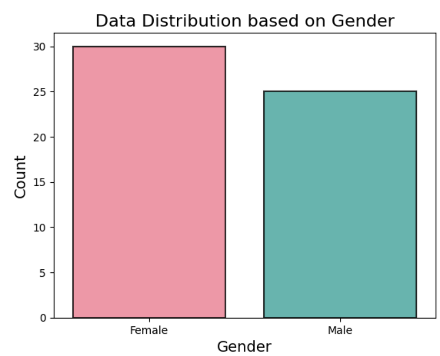

Fig. 5: Data Distribution based on Gender.

1) The MADRS score: MADRS score is the most commonly known and used measure in antidepressant efficacy trials. It stands for Montgomery–Åsberg Depression Rating Scale (MADRS) is a widely used clinician-rated measure of depressive severity [50]. The MADRS scoring instructions indicate that a total score ranging from 0 to 6 indicates that the patient is in the normal range (no depression), a score ranging from 7 to 19 indicates "mild depression," 20 to 34 indicates "moderate depression," a score of 35 and greater indicates "severe depression," and a total score of 60 or greater indicates "very severe depression [51]. Our Dataset contained the MADRS score and the chart below plots the MADRS score of the patients. In the figure, the MADRS1 is MARDRS score when measurement started, and the MADRS2 is MARDS score when measurement ended.

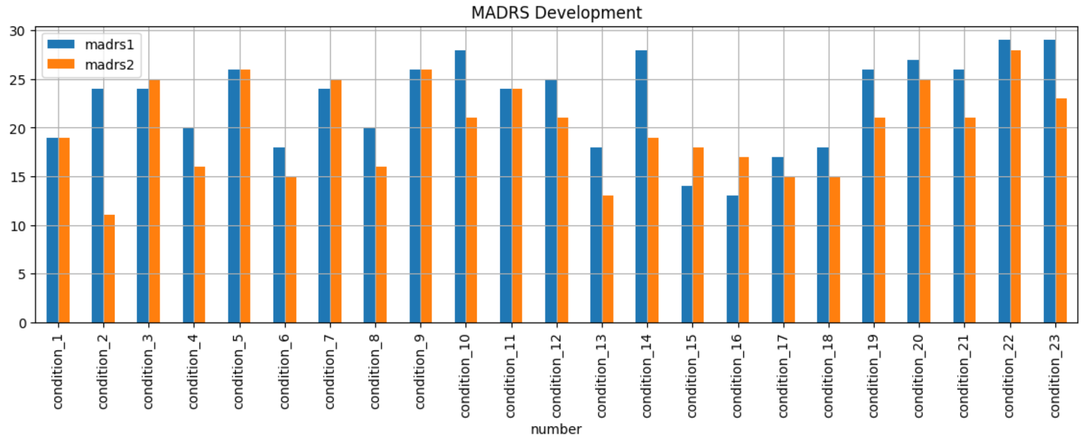

Fig. 6: Data Distribution based on the MADRS score.

V. Applied Classification Algorithms

The gap between technology and classification of unipolar disorder and Bipolar Disorder II is must to prevent misdiagnosis. The table below shows the summary of existing researches and applied classification learning algorithms for differentiating between the Unipolar disorder and Bipolar Disorder II. Thus, further research can be expanded focusing on the gaps.

TABLE II: Classification algorithms for Bipolar vs Unipolar classification

Classification Algorithm	**Bipolar vs Unipolar**
Logistic Regression	yes
Support Vector Machines	yes
Random Forest	yes
K-Nearest Neighbors	yes
Naive Bayes	yes
Artificial Neural Networks	yes
Gradient Boosting Machines	yes
Deep Learning Methods	X
Decision Trees	X
AdaBoost	X
XGBoost	X
LightGBM	X
CatBoost	X
Gaussian Processes	X

VI. Proposed Methodologies

Our aim was to develop a reliable model to classify Bipolar disorder and Unipolar disorder using Machine learning techniques. Given the size of the data, deep learning model wouldn't be effective here. We follow a systematic approach to ensure the model's effectiveness and generalizability. Here is a detailed outline of the steps involved in our methodology:

A. Data Collection

We collected data from reliable sources, including research articles and public datasets. The data comprises various features and labels indicating Bipolar and Unipolar disorders.

B. Feature Engineering

We performed feature engineering to extract relevant features and reduce the dimensionality of the dataset. We conduct correlation analysis to identify highly correlated features and remove them to avoid multicollinearity.

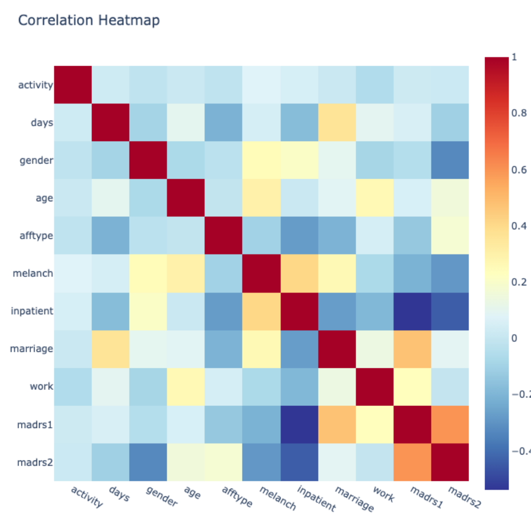

Fig. 7: Correlation heatmap among different features

C. Handling Missing value

We encountered missing values in some of our datasets. Missing values can create issues when training machine learning models, so we needed to handle them before proceeding with our analysis. We decided to use the IterativeImputer class from scikit-learn to impute the missing values. This class uses a machine learning technique to fill in missing values based on other features in the dataset.

We applied the imputer separately to the numerical columns and then concatenated the imputed numerical data with the non-numerical columns. By using this method, we were able to handle missing values in a way that preserved the integrity of the data and allowed us to perform further analysis and modeling. It was done with IterativeImputer class from scikit-learn.

D. Handling Imbalanced Data

Imbalance data are handled by using techniques such as oversampling, undersampling, or a combination of both, which helps improve the performance of our model on minority classes. In our case, we applied the Synthetic Minority Oversampling Technique (SMOTE) to balance the class distribution in our data. SMOTE generates synthetic samples of the minority class by interpolating between existing samples, which helps to increase the size of the minority class and make it more representative of the majority class.

Also, since our research is differentiation between bipolar disorder II and Unipolar Disorder, we dropped the data for Bipolar disorder I.

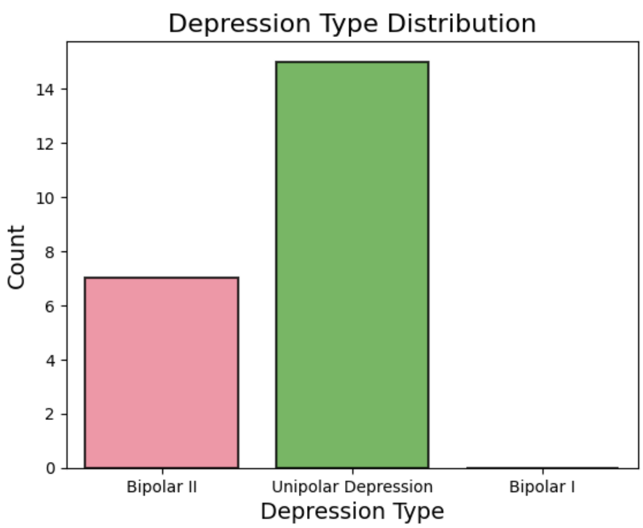

Fig. 8: Data Distribution based on Depression type.

In our dataset, the age of the patients was provided in an interval format, rather than as a single numerical value. To convert this data into a usable format for our analysis, we used the minimum value from each age interval as a representative value for each patient's age.

Additionally, We did this by first extracting the minimum value from each age interval using pandas' apply() function, which allowed us to apply a function to each row of the dataset. Then, we dropped the column containing the age interval data using the drop() function in pandas, since this information was no longer necessary once we had converted the age data to single numerical values.

By converting the age interval data to single numerical values, we were able to use age as a numerical feature in our analysis, which allowed us to investigate its relationship with other variables in the dataset.

E. Model Selection

We explored various machine learning algorithms, including Logistic Regression, Support Vector Machines, Random Forest, K-Nearest Neighbors, Naive Bayes, and others, to select the most suitable model for our classification task. We also researched previously applied classification algorithms and existing research for the differentiation between Bipolar Disorder II and Unipolar Disorder using technology.

F. Data Preprocessing

Before applying machine learning algorithms to our dataset, we performed some preprocessing steps to ensure that the data was in an appropriate format for analysis. Specifically, we applied feature scaling to our numerical data using the StandardScaler class from scikit-learn. This scaling method transforms each feature to have a mean of 0 and a standard deviation of 1, which can help improve the performance of some machine learning algorithms.

We also split our data into training and test sets using the train_test_split() function from scikit-learn. This function randomly splits the dataset into two subsets, one for training the model and one for testing its performance. By doing this, we were able to evaluate how well the model generalizes to new data that it has not seen during training.

G. Visualizing Actigraph Data

Actigraph recordings are an important method for measuring sleep parameters and average motor activity over a period of days to weeks using a noninvasive accelerometer. The accelerometer is often housed in a small device worn like a wristwatch, and recordings from the device can provide valuable information about a patient's condition.

To gain insight into the patient's activity levels over time, we plotted the activity of one of the patients on different dates. This plot helped us to understand the patient's overall activity patterns and to identify any changes or anomalies in their activity levels over time. By visualizing the patient's activity levels, we were able to gain a better understanding of their condition and to identify potential areas for further analysis.

In addition to visualizing the patient's activity levels, we also used several descriptive statistics from the Actigraph recordings, including mean_activity, min_activity, max_activity, and std_activity. These statistics provided us with a quantitative understanding of the patient's activity patterns, which we could use to compare their activity levels to those of other patients or to establish a baseline for their activity levels over time. These were used as an important features for our model.

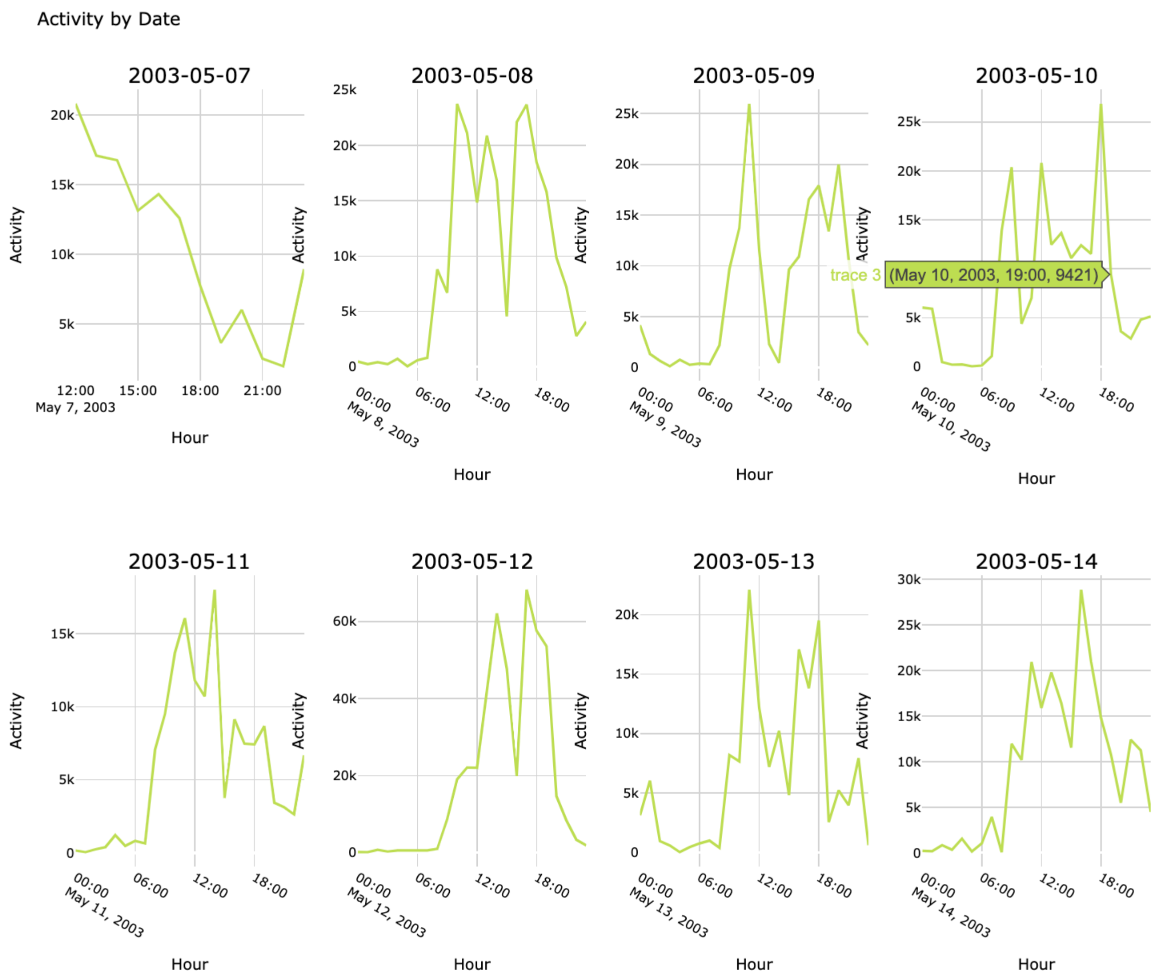

Fig. 9: Activity of a patient on different dates.

H. Applying Machine Learning models

After preprocessing the data, we choose six classification algorithms. Those were

- Logistic Regression
- Random Forest
- Decision Tree
- Support Vector Machine
- K-Nearest Neighbors Boosting

After applying these algorithms, we evaluated their performance using metrics such as accuracy, precision, recall, and F1 score. Based on these evaluations, we decided to perform cross-validation and hyper-parameter tuning on the Random Forest algorithm to improve its performance.

I. Hyper-parameter Tuning

To optimize the performance of our chosen model, we performed hyper parameter tuning using techniques such as Grid Search and Randomized Search. This step helped us to identify the best combination of hyper parameters for our model.

J. Model Evaluation

We used cross-validation to evaluate the performance of our model on unseen data. We calculated various performance metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of our model in classifying Bipolar and Unipolar disorders.

K. Model Interpretation

To gain insights into our model's decision-making process, we analyzed the feature importance and visualized the decision boundaries to understand how the model differentiates between the two disorders.

VII. Results

We applied six algorithms to our prepared dataset. Those were logistic regression, support vector machine, random forest, decision tree, K-nearest neighbors, and gradient boosting. Based on our evaluation matrix, our accuracy for gradient boosting algorithms was higher than all. The accuracy was 83.56% with a precision score of 87, an f1- score of 0.88, and a recall of 90. After doing cross-validation and hyperparameter tuning, we wanted to evaluate our model with new data. Upon completion of this, we found the random forest to be one of the best algorithms for differentiation between bipolar disorder II and unipolar disorder. The accuracy was 78%. Though this is less than our accuracy for the model before cross-validation dn hyperparameter tuning, A decrease in accuracy on the training set may actually be a good sign, indicating that the model is better able to generalize to new data. The precision score was 0.81, the f1-score was 0.85, and the recall was 0.89.

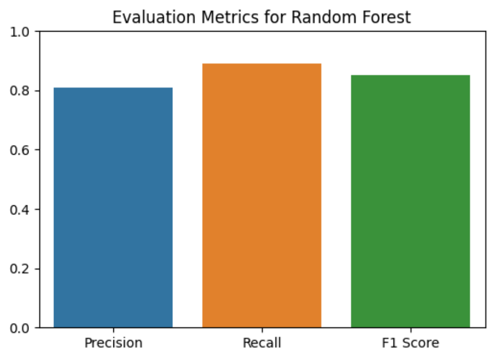

Fig. 10: Evaluation metrics of Random Fores.

VIII. Future work

In future work, we aim to explore the potential of integrating our machine-learning model with wearable technology and bio-markers to improve the early detection and diagnosis of Bipolar and Unipolar disorders. Apart from that improving and updating the model would also be significant part of the future research. Wearable devices, such as smartwatches and fitness trackers, can collect valuable data on physiological parameters like heart rate, sleep patterns, and physical activity, which could provide additional insights into an individual's mental health [52]. Moreover, recent research has shown the potential of biomarkers, such as genetic and epigenetic factors, inflammatory markers, and neuroimaging data, to differentiate between psychiatric disorders and facilitate personalized treatment [53], [54].

In addition to wearable technology and biomarkers, future research could also investigate the potential of incorporating natural language processing (NLP) techniques to analyze patients' speech patterns and linguistic features, which may offer valuable clues for distinguishing between Bipolar and Unipolar disorders [55]. Furthermore, the development of interpretable and explainable machine learning models could enhance clinicians' trust in these tools and facilitate their adoption in clinical settings.

The integration of various data sources and advanced machine learning techniques holds great promise for improving the early detection and diagnosis of Bipolar and Unipolar disorders, ultimately leading to better patient outcomes and more personalized treatment approaches.

References

[1] C. Lovering, "The History Behind Bipolar Disorder," 2 2022.

[2] N. B. M. Association, "Diagnostic and Statistical Manual of Mental Disorders," *DSM-5*, 5 2013.

[3] E. F. Maassen, B. J. Regeer, E. J. Regeer, J. F. G. Bunders, and R. Kupka, "The challenges of living with bipolar disorder: a qualitative study of the implications for health care and research," *International Journal of Bipolar Disorders*, vol. 6, 11 2018.

[4] S. Team, S. Team, and SingleCare, "Bipolar disorder statistics 2023," *The Checkup*, 1 2023.

[5] M. Screening, "How common is bipolar disorder? - MHA Screening," 10 2022.

[6] U. Ösby, L. Brandt, N. Correia, A. Ekbom, and P. Sparén, "Excess Mortality in Bipolar and Unipolar Disorder in Sweden," *Archives of General Psychiatry*, vol. 58, p. 844, 9 2001.

[7] J.-H. Hsu, I.-C. Chien, and C.-H. Lin, "Increased risk of ischemic heart disease in patients with bipolar disorder: A population-based study," *Journal of Affective Disorders*, vol. 281, pp. 721–726, 2 2021.

[8] "Janssen Pharmaceutica, L. P. (2000). The economic burden of bipolar disease. J Clin Psychiatry, 61(13), 38-41.."

[9] "Medical Practice Project. (1979). A State-of-the-Science Report for the Office of the Assistant Secretary for the US Department of Health, Education and Welfare.."

[10] "Bipolar disorder - Symptoms and causes - Mayo Clinic," 12 2022.

[11] P. López, F. Mosquera, J. de Leon, M. Gutiérrez, J. Ezcurra, F. Ramírez, and A. González-Pinto, "Suicide attempts in bipolar patients," *Journal of clinical psychiatry*, vol. 62, no. 12, pp. 963–966, 2001.

[12] J. F. Scott and M. Leboyer, "Consequences of delayed diagnosis of bipolar disorders," *L'Encéphale*, vol. 37, pp. S173–S175, 12 2011.

[13] Sandy, "Early detection of bipolar disorder - NeuRA Library," 2 2022.

[14] L. Culpepper, "The Diagnosis and Treatment of Bipolar Disorder," *The primary care companion for CNS disorders*, 6 2014.

[15] J. D. Lish, S. Dime-Meenan, P. C. Whybrow, R. A. Price, and R. Hirschfeld, "The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members," *Journal of Affective Disorders*, vol. 31, pp. 281–294, 8 1994.

[16] "Attention-deficit/hyperactivity disorder (ADHD) in children - Symptoms and causes - Mayo Clinic," 6 2019.

[17] "Depression (major depressive disorder) - Symptoms and causes - Mayo Clinic," 10 2022.

[18] "Substance Use and Co-Occurring Mental Disorders."

[19] "Anxiety disorders - symptoms and causes - mayo clinic."

[20] R. M. Post, L. L. Altshuler, M. A. Frye, T. Suppes, P. E. Keck, S. L. McElroy, G. S. Leverich, D. A. Luckenbaugh, M. Rowe, S. Pizzarello, R. Kupka, H. Grunze, and W. A. Nolen, "Complexity of Pharmacologic Treatment Required for Sustained Improvement in Outpatients With Bipolar Disorder," *The Journal of Clinical Psychiatry*, vol. 71, pp. 1176–1186, 9 2010.

[21] E. A. Youngstrom, L. E. Arnold, and T. W. Frazier, "Bipolar and ADHD comorbidity: Both artifact and outgrowth of shared mechanisms.," *Clinical Psychology-science and Practice*, vol. 17, pp. 350–359, 12 2010.

[22] B. M. Stiles, A. F. Fish, R. Vandermause, and A. M. Malik, "The Compelling and Persistent Problem of Bipolar Disorder Disguised as Major Depression Disorder: An Integrative Review," *Journal of the American Psychiatric Nurses Association*, vol. 24, pp. 415–425, 6 2018.

[23] C. L. Bowden, "Strategies to Reduce Misdiagnosis of Bipolar Depression," *Psychiatric Services*, vol. 52, pp. 51–55, 1 2001.

[24] "Misdiagnosis of bipolar disorder in children and adolescents: A c...: Ingenta Connect."

[25] J. Tomasik, S. K. Han, G. Barton-Owen, D.-M. Mirea, N. A. Martin-Key, N. Rustogi, S. Lago, T. Olmert, J. D. Cooper, S. Ozcan, P. Eljasz, G. Thomas, R. Tuytten, T. Metcalfe, T. S. Schei, L. P. Farrag, L. V. Friend, E. Bell, D. Cowell, and S. Bahn, "A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data," *Translational Psychiatry*, vol. 11, 1 2021.

[26] T. Olmert, J. D. Cooper, S. K. Han, G. Barton-Owen, L. P. Farrag, E. Bell, L. V. Friend, S. Ozcan, N. Rustogi, R. L. Preece, P. Eljasz, J. Tomasik, D. Cowell, and S. Bahn, "A Combined Digital and Biomarker Diagnostic Aid for Mood Disorders (the Delta Trial): Protocol for an Observational Study," *JMIR Research Protocols*, vol. 9, p. e18453, 8 2020.

[27] G. L. Faedda, G. Serra, C. Marangoni, P. Salvatore, G. Sani, G. H. Vázquez, L. Tondo, P. Girardi, R. J. Baldessarini, and A. Koukopoulos, "Clinical risk factors for bipolar disorders: A systematic review of prospective studies," *Journal of Affective Disorders*, vol. 168, pp. 314–321, 10 2014.

[28] L. De Oliveira, L. C. L. Portugal, M. G. Pereira, H. W. Chase, M. A. Bertocci, R. Stiffler, T. Greenberg, G. Bebko, J. Lockovich, H. Aslam, J. Mourao-Miranda, and M. L. Phillips, "Predicting Bipolar Disorder Risk Factors in Distressed Young Adults From Patterns of Brain Activation to Reward: A Machine Learning Approach," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 4, pp. 726–733, 8 2019.

[29] G. S. Malhi, D. M. Bargh, C. Coulston, P. Das, and M. Berk, "Predicting bipolar disorder on the basis of phenomenology: implications for prevention and early intervention," *Bipolar Disorders*, vol. 16, pp. 455–470, 8 2014.

[30] D. V. Skjelstad, U. F. Malt, and A. Holte, "Symptoms and signs of the initial prodrome of bipolar disorderA systematic review," *Journal of Affective Disorders*, vol. 126, pp. 1–13, 10 2010.

[31] F. Cassidy, K. Forest, E. Murry, and B. J. Carroll, "A Factor Analysis of the Signs and Symptoms of Mania," *Archives of General Psychiatry*, vol. 55, p. 27, 1 1998.

[32] A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," 1967.

[33] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," *Stroke and vascular neurology*, vol. 2, pp. 230–243, 12 2017.

[34] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, pp. 44–56, 1 2019.

[35] *Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.*

[36] Z. Keita, "Classification in Machine Learning: An Introduction," 9 2022.

[37] "What is Logistic regression? — IBM."

[38] W. contributors, "Support vector machine," *Wikipedia*, 3 2023.

[39] "What is the k-nearest neighbors algorithm? — IBM."

[40] W. contributors, "Naive Bayes classifier," *Wikipedia*, 3 2023.

[41] W. contributors, "Artificial neural network," *Wikipedia*, 4 2023.

[42] H. Singh, "Understanding Gradient Boosting Machines - Towards Data Science," 11 2018.

[43] W. contributors, "Deep learning," *Wikipedia*, 5 2023.

[44] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, p. 1157–1182, mar 2003.

[45] S. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, 11 2006.

[46] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, pp. 427–437, 7 2009.

[47] "The Depression Dataset," 2 2021.

[48] "Actigraphy in the evaluation of sleep-Uptodate."

[49] Canadian Agency for Drugs and Technologies in Health, "Validity of outcome measures."

[50] L. C. Quilty, H. M. Colhoun, J.-P. Rolland, F. De Fruyt, F. Rouillon, and R. M. Bagby, "The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression," *International Journal of Methods in Psychiatric Research*, vol. 22, pp. 175–184, 9 2013.

[51] M. M. Müller, H. Himmerich, B. Kienzle, and A. Szegedi, "Differentiating moderate and severe depression using the Montgomery–Åsberg depression rating scale (MADRS)," *Journal of Affective Disorders*, vol. 77, pp. 255–260, 12 2003.

[52] J. Torous, P. Staples, I. Barnett, L. Sandoval, M. S. Keshavan, and J.-P. Onnela, "Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia," *npj digital medicine*, vol. 1, 4 2018.

[53] B. S. Fernandes, L. M. Williams, J. Steiner, M. Leboyer, A. F. Carvalho, and M. Berk, "The new field of 'precision psychiatry'," *BMC Medicine*, vol. 15, 4 2017.

[54] T. Kato, "Neurobiological basis of bipolar disorder: Mitochondrial dysfunction hypothesis and beyond," *Schizophrenia Research*, vol. 187, pp. 62–66, 9 2017.

[55] G. Bedi, F. Carrillo, G. A. Cecchi, D. F. Slezak, M. Sigman, N. B. Mota, S. Ribeiro, D. C. Javitt, M. Copelli, and C. Corcoran, "Automated analysis of free speech predicts psychosis onset in high-risk youths," *npj schizophrenia*, vol. 1, 8 2015.