Ripple Down Rule Based Decision Intelligence for Mental Disorder Diagnosis

by

G.M. Arafat Rahman
19101498
Tahmid Nizam Ibnul
19101529
MD. Shamim Mia
19101532
Abid Mahmood Akash
19101533
Avinandan Banerjee
19101541

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
January 2023

Declaration

It is hereby declared that

1. The report submitted is our own original work while completing degree at Brac University.

2. The report does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The report does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Arafat

G.M. Arafat Rahman
19101498

Tahmid Nizam Ibnul
19101529

Shamim

MD. Shamim Mia
19101532

Akash

Abid Mahmood Akash
19101533

Avinandan

Avinandan Banerjee
19101541

Approval

The report titled "Ripple Down Rule Based Decision Intelligence for Mental Disorder Diagnosis" submitted by

1. G.M. Arafat Rahman(19101498)
2. Tahmid Nizam Ibnul(19101529)
3. MD. Shamim Mia(19101532)
4. Abid Mahmood Akash(19101533)
5. Avinandan Banerjee(19101541)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 24, 2023.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Ripple Down Rule (RDR), a rule-based incremental system, enables knowledge acquisition from human experts to knowledge-based systems (KBS). The majority of modern decision intelligence systems rely on machine learning algorithms, despite the fact that most machine learning algorithms have their own limitations, such as a lack of explainability, an inability to provide multiple outputs, and poor performance with imbalanced or unbalanced data. In addition, RDR still needs to be implemented in the mental health field, and most of the current screening tests cannot diagnose multiple mental disorders at a time. Because of these issues, this paper presents an RDR-based approach for diagnosing mental disorders based on data gathered from primary sources. Since RDR is both a knowledge-based system and an inference engine where domain experts provide rules and conclusions, it can correctly explain its conclusion and provide multiple outputs using the Multiple Classification Ripple Down Rule (MCRDR). In addition, a version of the XGBoost classification algorithm called 'XGBoost Binary Classification Block' has been presented to produce multiple outputs. Comparing the experimental outcomes of three classifier models, we find that XGBoost multiclass classification has 49% accuracy, XGB Binary Classification Block has 96% accuracy, and RDR outperforms the other two by accurately predicting all outputs.

Keywords: Ripple Down Rule, Knowledge-based System, Decision Intelligence, Machine Learning, Explainable, Multiple Outputs, Mental Disorder, Inference Engine, Domain Expert, MCRDR, XGBoost Binary Classification Block.

Acknowledgement

We would like to begin by expressing our gratitude to Allah for enabling us to complete our thesis without any significant obstacles.

Our supervisor, Dr. Md. Golam Rabiul Alam, deserves recognition for his unwavering support and guidance throughout our research. He was always willing to help us whenever we needed it.

We also want to thank Dr. Shafika Afroz for her significant contributions to our research, particularly for serving as the lead expert psychiatrist in developing our system, from creating questionnaires and rules to labeling cases.

Finally, we would like to acknowledge the support of our parents, without whom this would not have been possible. Thanks to their support and prayers, we are now on the brink of graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$DSM4$ Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

$DSM5$ Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

GAD Generalized Anxiety Disorder

KBS Knowledge-Based System

$MCRDR$ Multiple Classification Ripple-Down Rules

MDD Major Depressive Disorder

$ModSevere$ Moderately Severe

RDR Ripple Down Rules

$SCRDR$ Single Classification Ripple-Down Rules

$XGBoost$ Extreme Gradient Boosting

Chapter 1

Introduction

Decision Intelligence is a commercialized application of artificial intelligence that helps make decision-making easier [32]. Decision intelligence applies to all business areas. Decision intelligence can be applied to any decision-making situation, including disease diagnosis. In decision intelligence, people from many fields work together to create, model, coordinate, run, monitor, and change decision models, and processes. These disciplines include agent-based systems, decision management, and decision support. Techniques such as predictive analytics are also included. The goal of decision intelligence is not to omit humans from the process of decision-making entirely. It aims to make the decision-making process easier for humans using vast amounts of data and with the help of artificial intelligence.
Knowledge-based systems can be helpful in decision-making. A knowledge-based system (KBS) examines knowledge, data, and other information from various sources to deduce new knowledge. These systems, which have built-in problem-solving abilities, can grasp the context of the data, analyze and process the data, and make intelligent decisions based on the information they have stored.
Ripple Down Rule is a rule-based system mainly used for knowledge acquisition with the help of domain experts. It is a type of knowledge-based system that will be used to differentiate between various mental diseases in our problem. Ripple Down Rules (RDR), according to Compton, is a method for iteratively building systems while they are in use.

1.1 Problem Statement

Supervised learning models for classification problems have grown in popularity in recent years. This strategy, however, comes with its own set of difficulties. Machine learning has produced spectacular results thus far, but a critical question remains: is there a sufficient amount of labeled data to acquire the necessary concepts? Furthermore, providing necessary and relevant data is difficult. Relevant data works best with machine learning algorithms that maintain quality. Besides that, the amount of human labor required to create a machine-learning system is a major concern [29]. The Watson-based medical literature is a methodological approach rather than an implementation. According to IBM's 2012 Global Technology Outlook Report [11], "Managing Uncertain Data at Scale" is a critical problem for learning and analytics relating to the primary issue of data quality. The accuracy of the label or categorization applied to the data is of particular concern.

A. Impure and insufficient data on mental health

Machine learning is used to classify decisions based on preprocessed data. This is the major form of machine-learning algorithm being used. This data should be labeled for supervised learning applications. Researchers frequently use crowd-sourcing for this labeling task. As a result, the integrity of those who label the data becomes a major concern [20]. Information obtained from the general public is frequently untrustworthy due to personal, ethical, or societal norms. This is especially true for data on mental health. According to clinical psychologist Nikki Massey-Hastings, Psy.D., our society places illogical taboos on mental health issues over physical conditions, causing many people to feel embarrassed or ashamed of their symptoms [14]. This leads to the second issue of insufficient data when working on a new topic. Because machine learning performance is dependent on the purity and quantity of data, it fails to perform as expected in this case. As a result, decisions are frequently misguided or misjudged [29]. Large, reliable data plays a huge role in decision intelligence. Vast amounts of reliable data help to make decisions more accurately. Additionally, the problem arises when there is imbalanced or unbalanced data for machine learning methods. Any statistical method is prone to unbalanced dataset [29]. However, there are various ways to handle this problem, such as resampling and feature reduction.

B. Biases in machine learning

Another issue with machine learning is its bias. Humans have prejudices when making decisions [2]. People frequently reflect on data with their own beliefs and opinions. Because machine learning is data-driven, if the data is skewed, so will the machine learning models' decisions. For example, we frequently assign specific tasks to men and others to women. We believe that a baby girl will play with dolls and a baby boy will play with cars, for example. Alternatively, we may believe that nurses are exclusively female. Female candidates' resumes were penalized by Amazon because the AI model used for recruitment was trained on a dataset dominated by male data [24]. This exemplifies gender bias. There are numerous types of bias in machine learning. Because both men and women can suffer from mental illness, symptom patterns may differ according to gender, age, religion, race, ethnicity, and culture. There may still be partiality in the data after it has been cleaned up. Machine learning works well when there is enough labeled or classified data that is free of human bias.

C. Difficulty in explaining a conclusion to a machine learning model

One of the biggest problems with machine learning is that it can be hard to understand what it is doing. If the end users of the system are not well versed in explainable AI algorithms, they are not very useful to them. Popular explainable AI algorithms, LIME and SHAP, are known for being vulnerable to deliberately misleading explanations [28]. At the same time, they are prone to generate unrealistic scenarios and complex parameters [31]. Researchers in machine learning have turned their attention to the "-omics" sciences (genomics, proteomics, metabolomics, and so on) because they use large, complex databases [34]. Machine learning techniques are much more appropriate there. However, for mental disorder diagnosis, it is necessary for the psychiatrist to know exactly why certain outputs are the way they are. Therefore, despite their obvious success in some fields, machine learning methods are hard to understand for not experts.

The principal motivation behind our research is that there is no RDR based system in the mental health domain. Moreover, the majority of existing machine learning algorithms do not generate multiple conclusions for a single case. These existing problems drive us to explore the field of Python MCRDR implementation along with making a system that can classify multiple disorders in a single case. Ripple Down Rule system is efficient when there is insufficient reliable data available for machine learning and data that can be used to train machine learning model are costly to obtain [29]. RDR system can handle this problem, as a domain expert adds rules for every single case, and RDR is not affected by imbalanced or unbalanced dataset problems. Furthermore, we inspect the possibility of finding a way to provide multiple outputs using a machine learning classifier.

1.2 Research Contribution

In this work, we apply a rule based model named RDR to build a decision intelligence system for diagnosing mental disorder. Along with that we propose an XGBoost classifier model named as 'XGB Binary Classification Block' for multi-output classification. In particular, we overcame the shortcomings of most of the machine learning algorithms such as unexpainability of reaching a conclusion and poor performance on small amount of data by applying RDR. To be more specific, the main contributions of this paper are:

- Since no RDR based system exists in medical disease diagnosis, we have developed an RDR Knowledge base system for mental disorder diagnosis. The rules made by experts are much more reliable, explainable and trustworthy. The ruleset also helps to explain RDR system predictions.
- The existing online screening tests determine only the severity level of disorders, but none of them classify the types of disorders. A person can be conditioned to more than one mental disorder. Our system classified 17 types of mental disorders by implementing a MCRDR algorithm.
- We collected primary dataset for our research. The dataset that has been used come from surveys. Alongside that we have collaborated with a psychiatrist to annotate our data and make the questionnaire of the survey.
- The current XGBoost Classifier provides only binary and multiclass classifications. For our purpose, we need multiple outputs as conclusion. Hence, we present XGBoost Binary Classification Block for giving multiple conclusions for a single case.
- MCRDR implementation most commonly existed only in Excel. In contrast, we have developed a Python package to implement MCRDR to diagnose mental disorders.

1.3 Research Organization

A brief overview of each chapter's topics are provided here. The report has been divided into five chapters named as Introduction, Literature Review, Methodology, Implementation & Performance Evaluation and finally Conclusion. Each of these chapters are subdivided into further section's to categorize the contents.

Introduction: This chapter gives a brief summary of decision intelligence, RDR and mental disorder problems in the world. Motivation behind the research is later fulfilled by the contribution of the research provided in this chapter.

Literature Review: In this chapter existing works related to mental disorder diagnosis is provided along with giving a short introduction of DSM-5 [12] book. DSM-5 is widely used to diagnose mental disorders. Another book named 'Ripple-Down Rules: The Alternative to Machine Learning' [29] is also summarized in this section. This will give enough insight into the RDR background. This chapter also includes the working mechanism of Knowledge Based Systems.

Methodology: This chapter starts with the data collection. This section describes how data has been collected for the research as the primary source. Secondly, the preprocessing of the data is depicted in this chapter. Finally, MCRDR model specification is highlighted.

Implementation & Performance Evaluation: This chapter contains performance metrics through which the system will be evaluated followed by the results of the system. In the last section, a practical implementation of the system has been described in a detailed manner.

Conclusion: The research concludes its findings in this section along with pointing out the rooms for future improvement.

Chapter 2

Literature Review

2.1 Background Study

According to this paper [25], decision-making requires the most satisfactory or optimal solution to a decision problem to be identified and evaluated, followed by the selection of the ”best” alternative. It is the goal of Data-Driven Decision-Making (D3M) to make an inference from past and present data relevant to the decision problem, which is fundamentally distinct from traditional model-driven decision-making. Portfolio management, for example, aims to optimize the rate of return by managing a portfolio of securities. Data mining and machine learning techniques will be used to infer future investment strategies from historical data if we address this challenge via D3M.

The first category, Programmable Data-Driven Decision-Making (P-D3M) is a two-step process for making decisions. First, a decision model (sometimes called a programming or optimization model) is derived from data mining or statistical learning and a model that may be used to give decision assistance, such as a multi-objective decision model and a multi-level one. The second step is to create algorithms for solving the resulting decision model in order to discover the optimal option.

Non Programmable Data-Driven Decision-Making (NP-D3M), in contrast to P-D3M, can be used in cases when the derivation of a decision model is computationally infeasible or excessively expensive. In order to conduct data analysis, NP-D3M makes use of a variety of machine learning strategies. This is accomplished by characterizing the decision problem and locating connections between the problem variables (input, internal, and output variables) without having explicit knowledge of the physical behavior of the decision model. Rules, preference, and reinforcement learning are what drive the learning module, which then directly generates a decision solution.

This article [27] provides a brief overview of the major trends, studies, methodologies, and models in fuzzy and linguistic decision-making that have emerged over the past 50 years. Fuzzy and linguistic approaches to decision-making enable the solution of real-world, complex decision problems in which humans exhibit imprecision, vagueness and/or use natural language to evaluate decision alternatives, criteria, etc. This article has three purposes. First, the principal fuzzy set theory and computing with words-based representation paradigms of decision information are reviewed, along with their varying expressive complexity and richness. Second, three fundamental decision-making frameworks are examined: 1) decision making based

on multiple criterias , 2) group consensus-driven decision making, and 3) multi-person multi-criteria decision making. Third, the article talks about new complex decision-making frameworks that have come up in recent years. In these frameworks, decisions are made based on the ”wisdom of the crowd.” The challenges that come with these frameworks are discussed, as well as key guidelines that are needed for future research in the field.
The major purpose of this study is to offer an overview of the existing situation and prospective future research possibilities in this sector. In order to achieve this objective, the following research obstacles are discussed: 1) making wiser decisions more effectively; 2) using realistic and common sense approaches to represent expert assessments; 3) incentivizing more impact-driven and practical developments; and 4) stimulating future developments toward ’crowd decision-making’ phenomena, i.e. events guided by the intelligence and wisdom of crowds and aided by recent socio-technical developments. They conclude by outlining other outstanding issues in the area, which are, of course, related to the preceding areas.
Large-Scale Social Network With Opinion Dynamics: Extensive study has been conducted on the methods and theory of opinion dynamics in large-scale social networks, but there are still gaps in the practical implementation of decision-making difficulties, since individuals’ opinions are readily affected by others in this era of social networks.
Detecting and Influencing Opinions: With the help of advertising, companies may persuade customers to buy their products. However, customers’ preferences and opinions may fluctuate. To accomplish the intended impact, the most suitable decision making procedures and publicity tactics must be adopted in order to uncover the preferences of clients via online and offline data mining. Using case studies from the actual world, these strategies must be verified.
Complex Linguistic Determination: Most online evaluations use linguistic information in complicated situations such dynamic social networks, heterogeneous choice environments, and large-scale decision environments. In linguistic decision-making, the concept that words have varied meanings for different people is referred to as personalizing computing with words semantics (CW). As online review approaches have been examined extensively, they need validation in complicated decision-making environments based on individual semantics.
The article [9] discusses the use of multiple-attribute group decision-making problems, including various sources of ambiguous attribute preference information. In the issues, precise attribute values are unknown, but value ranges may be deduced, and the preferences of decision-makers are conveyed in three unique ways: 1) Interval utility values; 2) Interval fuzzy preference relationships; and 3) Interval multiplicative preference relationships.
They have constructed models to extract attribute weight data from each preference format type. Then, they merged all of these models into one to estimate the attribute weights and total attribute values of alternatives, which may reflect the consensus of all decision-makers. Not only can the integrated model prevent the loss or distortion of original preference information throughout the information integration process, but it can also bring the group opinion as close as possible to the perspective of each individual decision-maker. It is vital to note that the ranking of the provided alternatives may be easily computed using the overall attribute values of the integrated model. Moreover, they have developed models to address multiple-attribute

group decision-making problems where attribute values are expressed as real numbers and preferences as binary strings. The information provided by decision-makers over attributes takes three distinct exact preference formats: utility values, fuzzy preference relations, and multiplicative preference relations.
As found in [12], people with depressive disorders experience various emotions, including sorrow and irritability. They also experience changes in their mental and physical health that can substantially impact their daily lives. There are various types of depressive disorders. They are disruptive mood dysregulation disorder, major depressive disorders (including major depressive episode), persistent depressive disorder, premenstrual dysphoric disorder, substance/medication-induced depressive disorder, depressive disorder due to another medical condition, and other specified depressive disorders, and unspecified depressive disorder. These different depressions differ among them in issues of duration, timing, or presumed etiology. Children up to 12 years of age may have bipolar disorder, which is a presentation of children with persistent irritability and frequent episodes of extreme behavioral dyscontrol. This disorder falls into the category of depressive disorder and also a new diagnosis that is a disruptive mood dysregulation disorder.
Major depressive disorder is characterized by discrete episodes of at least two weeks' duration. This disorder has a clear-cut change in affect, cognition, neurovegetative functions, and interepisode remissions. In most cases, major depressive disorder is a recurrent episodic disorder. However, diagnosis of a single episode is possible.
Although bereavement may induce great suffering, it is not a major depressive disorder because it generally does not induce an episode of major depressive disorder. However, when they do occur together, the symptoms tend to be more severe. Those with other vulnerabilities to depressive disorders are more likely to experience bereavement-related major depressive episodes. In order to identify persistent depressive disorder, the mood disturbance must persist for at least two years in adults or one year in children.
A premenstrual dysphoric condition is a form of depression that manifests after ovulation and subsides within a few days of menstruation. Moreover, a depression-like phenomenon can occur when substance abuse happens in large amounts and in the cases of several medical conditions. This substance abuse can be due to some prescribed medications.
In this book [29], a user considers a case as data when drawing conclusions from the data. There are three different Ripple Down Rules that have been mentioned. The first one is the Single Classification RDR (SCRDR). This classification can only give a single conclusion for a case. The second one is Multiple Classification (MCRDR) which can provide more than one conclusion for a case. The final type of classification is the General Ripple Down Rule (GRDR). Repeat inference with facts asserted and retracted and subsequent inference based on changes to the available facts until there are no more changes is an essential feature of a GRDR.
Conventional knowledge acquisition methods can be separated from Ripple Down Rules(RDR) in four ways. The four key features are:

1. To deal with a specific case, rules are added.
2. The knowledge base defines in which order the rules will be evaluated. For order, the inference engine does not have any effect.
3. One can only add rules and can not change or remove the rules.

4. Rule actions do not rebut facts, but instead, affirm them.

The addition of a rule to a case is determined by certain criteria. They are as follows:
A) Case-Driven Knowledge Acquisition: Classification in RDR typically involves drawing a conclusion for a specific case. In the RDR classification problem, classification is made between cases by some distinguishing features. The users are asked which features made them give a particular conclusion for a particular case. They will never be asked to give a rule. A user may choose to ascribe many, possibly interdependent conclusions to a case, however only one conclusion is questioned at a time, along with the features that contribute them to reach that conclusion. To do this, a previous case with a different conclusion might have the same features. If this happens, the user is prompted to specify one or more distinguishing characteristics for each scenario. If this is also not possible, then the user should then be convinced that the conclusion for the preceding case should be altered. The main reason for this is that two different cases will have some distinguishing features that make them different from each other. If two cases do not have any distinguishing features, they are identical. Rules are added for the cornerstone cases and they serve as the validation standard for the rule.
B) Order of Cases Processed: A rule can be added to a case in any order, and adding more general rules first is not required.We can construct RDR systems while using them by doing this. After adding a rule to a case, the output is monitored. A new rule is added if the case does not give a correct conclusion. This can lead to rule redundancy as well. Users who create rules that are too exact and close to the actual situation can result in a lot of unnecessary rules being added. The addition of a too general rule is another issue. It is pointless to include a rule that is too broad in scope because over-generalization will lead to errors.To minimize the errors, new rules are added. Adding new rules will automatically shorten the scope of an over-generalized rule.
C) Linked Production Rules: Expert or knowledge-based systems consist of two components. The first one is the production rule, and the second one is the inference engine. A production rule [33] consists of two components, one is the conditional part, and the other is the action list. For instance:

$$If\ \ [condition]\ \ then\ \ [action - list]$$

A knowledge-based system is made of such rules. Rules are reviewed in random order and are independent of one another. An extremely primitive inference engine might just analyze rules in the order they are listed in a file. But the inference engine takes the help to determine the next evaluated rule by a conflict resolution strategy. The conflict resolution approach determines the next rule to be reviewed based on the following conditions:

1. It prefers a rule that has more conditions. Here conditions mean features.

2. It prefers rule conditions that utilize the most recently proclaimed fact.

3. As with the primitive inference engine, it will select the first rule in the knowledge base that satisfies the case.

4. Alternatively, it can pick the last rule that is most recently added to the knowledge base.

5. Rules may also be allocated additional information, such as a salience score. If two rules can be executed, the rule with the greater salience is selected.

6. Choose a candidate rule at random if conflict resolution strategies do not result in a single candidate rule.

Although a conflict resolution strategy gives a better approach to rule selection, it often leads to a problem. The problem is when a new rule is changed or added; it is impossible to predict how the new rule will affect the other rules.When other rules change, a case may no longer be judged based on the same criteria, which complicates the case-differentiation technique outlined above. RDR, in contrast, specifies the next rule to be evaluated. RDR can determine the next rule that will be evaluated because almost every knowledge bases based on RDR are n-ary (especially binary) trees and RDR rules are linked production rules. For instance,

$$\begin{aligned} If \;\; [condition] \;\; then \;\; & [case\, action - list], \\ & [inference \;\; action] \\ Else \qquad\qquad\qquad & [inference \;\; action] \end{aligned}$$

Here case action-list is a normal inference engine that will assert one or more conclusions or facts or it can specify some other actions. RDR also specifies what to do when a rule fires by giving an inference action, and it also specifies what to do when a rule fails to fire. The inference action merely indicates the next rule to be evaluated. This is accomplished by linking a new rule to the preceding rule. This link information can not be changed. Moreover, a depth-first approach is used in RDR, evaluating earlier rules and their refinements before evaluating newer rules and their refinements.

D) Adding Rules: To avoid complications in RDR, the body of a rule is impossible to change, delete, or add. This is why linked production rules are required. In contrary, a conclusion of a rule is modifiable, as the conclusion's inference is not affected by this; only the inference path's assigned label is changed. Wrong conclusion can be corrected by adding a new rule. This new rule is called a refinement rule for the previous rule with an incorrect conclusion. The refinement rule provides conclusion to supersede the conclusion of the parent rule.Typically, despite having the possibility of being identical to any previous rule, refinement rule limits the parent rule's reach; this is equivalent to changing the rule's conclusion. A refinement rule can be used as a 'stopping' rule that will give no conclusion if the refinement rule fires.

E) Assertions and Retractions: Standard knowledge-based systems permit rules to retract and assert (assign) facts. On the other hand, only the assertion of facts is permitted by RDR, except in the exceptional circumstance where the conclusion of the refined rule is replaced by a refinement rule. A fact that would demand retraction is one that has been falsely asserted, hence retraction is not required. Instead of using inference to control conclusions that should not have been expressed, the RDR technique employs knowledge acquisition to prevent the inaccurate conclusion from being reached in the first place. How this operates in practice has been demonstrated later.

F) Formulae in Conclusions: In the preceding explanation, we assumed that conclusions were factual claims; but, like other rule-based methods, a piece of code, a formula or essentially anything that is available to be used by and sent to outside world can be a rule action. Bekmann [6][7] took the help of RDR to create the formulas for fitness function and the mutation operator for a genetic algorithm program which resulted in the application of different formulas in different situations with the advancement of the generations. Misra selected the processes or the parameters of the process that should be employed in complex multi-process activities like image processing by creating a system based on RDR [10]. Drake and Beydoun have also proposed RDR based on predicate logic [3].

2.2 Related Works

The article [16] says that Global abnormalities in tractography-based graph metrics in Major Depressive Disorder (MDD) have not been reported to date. In order to distinguish between healthy and depressed individuals, they employed a machine learning approach based on "support vector machines." Graph metrics were also looked at to see how relevant they were in making this distinction. People with MDD can be distinguished from healthy controls using a combination of whole brain graph metrics that can be optimally combined utilizing results from SVMs in the current study, as demonstrated. Researchers showed that classification accuracy and feature weights were higher when small-worldness was taken into account. Finally, their findings show that MDD sufferers have abnormally low levels of regional connectivity. A degree centrality graph analysis of local connections was used to compare regional connectedness in MDD to healthy controls, and three brain regions were identified as being differentiating for the MDD group vs. healthy controls: the right inferior parietal cortex, right pars orbitalis, and left pre-rostral cingulate. The pars orbitalis and rostral anterior cingulate were more connected in MDD than the parietal area, which showed diminished connectivity. This study has four drawbacks that should be noted: (1) Half of the MDD participants in their study had comorbid anxiety disorders; (2) three MDD individuals were using psychotropic drugs; (3) all participants were female; and (4) their sample size (N = 32) was small. As a result, aside from reproducing current findings in a larger cohorts of depressed and healthy persons in the future, it will be essential to look for possible influence of anxiety comorbidities, pharmacological agents, genders and age on measures of network connection in depression.
According to [1], The currently accepted diagnoses of personality disorders are not supported by measurement in a significant way and have not been independently verified. They studied the personality features of 130 psychiatric patients using a standardized interview schedule for assessing disordered personality. Of these patients, half were clinically determined to have personality disorders. They considered 24 symptoms to classify a patient. they have used International Classification of Disease (ICD) proposed by World Health Organization to classify those patients. However, they came out to find only 37 patients had personality disorder but in the original dataset half (65) of the patients were having personality disorder. To analyze the data, they subjected the ratings of both the normal and personality disordered groups separately and collectively, to a factor analysis that utilized the Varimax rotation in order to determine whether or not the same variables loaded in both sets of evaluations. In addition, a cluster analysis was performed on all

of the collected data in order to find the most appropriate classification of groups, particularly with regard to the personality disorders. This was a test of the ICD classification of personality disorders, which is a categorization that has never been validated by independent sources. Cluster analysis was performed using both hierarchical and non-hierarchical approaches simultaneously. The findings of the factor analysis indicate that the fundamental structure of variables is the same in individuals who have a primary personality disorder as well as in individuals who do not have this disorder. This finding lends support to the notion that personality disorders are located at the extreme end of a multidimensional continuum.
In this paper [26] a variety of machine learning studies on various anxiety disorders are reviewed. Sribala [17] used the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) standard questionnaire to construct a neural network-based model for predicting Generalized anxiety disorder (GAD). The results were 96.43 percent with sensitivity analysis and 90.32 percent without it. Hussain et al. [19] measured the severity of depression using data from the Beck Depression Inventory (BDI) exam. They utilized a random forest tree, and a balanced dataset produced the most accurate results. Using functional MRI data, Leuken et al. [15] predicted the comorbidity status of patients with panic disorder and agoraphobia. The prediction was made using an ensemble tree classifier (Random Under Sampling Boost algorithm). The accuracy of the comorbidity status was 0.79, with a sensitivity of 0.73 and a specificity of 0.85. Individual response to cognitive behavioral therapy (CBT) in panic disorder with agoraphobia was predicted by Sundermann et al. [23] using multivariate pattern analysis with soft margin support vector machines. Authors have attained a classification accuracy of approximately 51%. Reece et al. [22] related childhood tragedy to anxiety disorder in their lives and Twitter tweets. The study utilized the Random Forest algorithm with 234,000 posts, 63 subjects with anxiety, and 111 participants with good health. The study gathered predictive variables from participant tweets that quantified the effect, language style, and context (n = 279,511) as well as models with supervised learning algorithms that utilized these data. The Random Forest Model accurately identified an anxious symptom for clinical diagnosis with 89% precision.

Chapter 3

Methodology

As per the usual plan to build any system, ours starts with data acquisition. The primary goal has been to get information about the symptoms of the differential diagnosis of mental health. This symptom data is primarily provided by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [12]. Following that, we pre-process the data to fit our usage better. For instance, we categorized the several age requirements to 4 discrete age groups, for 0-5, 6-11, 12-17, and above 18. More instances of pre-processing can be seen in the later sections.

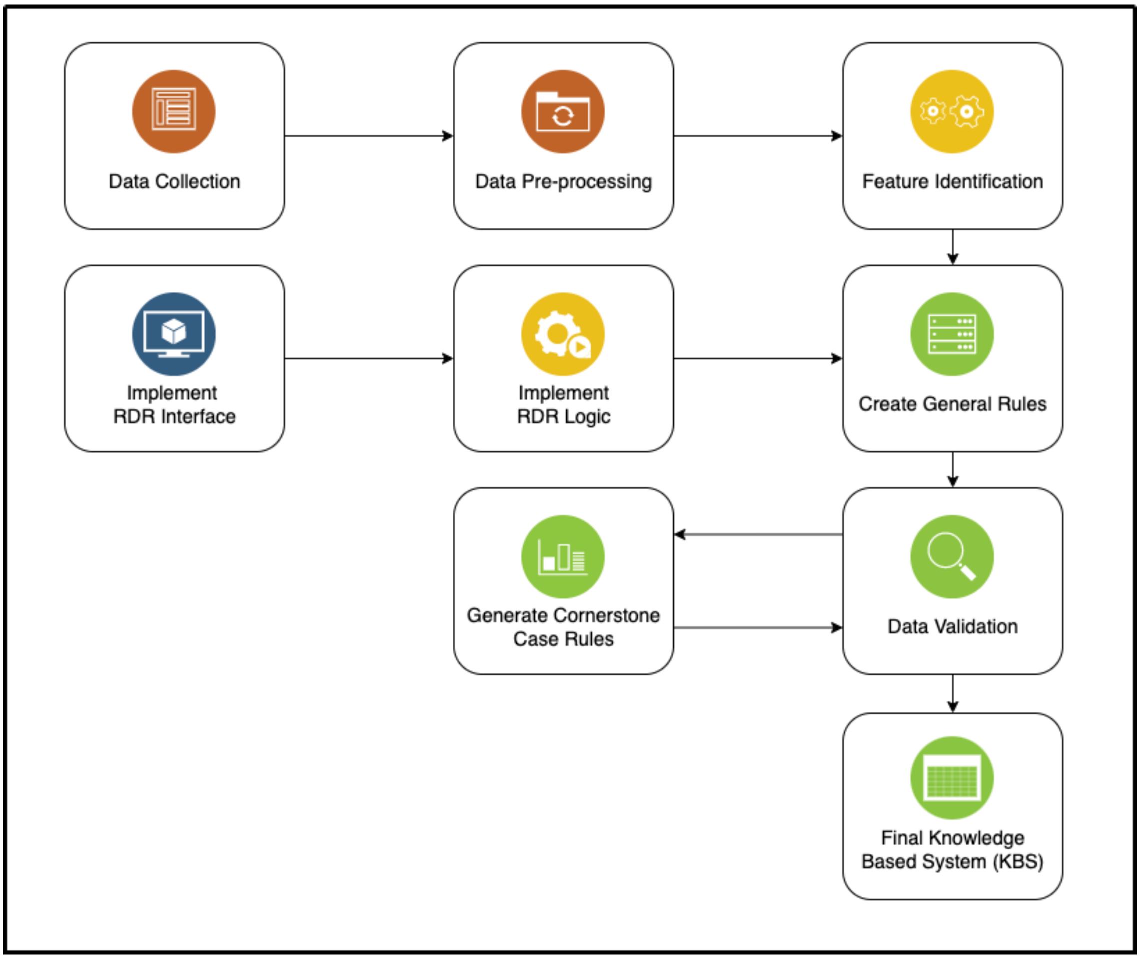


Figure 3.1: Core Architecture of the System

Afterward, we choose the most defining features from the symptoms in our system. For this, we study the common and unique symptoms. For example, agoraphobic patients express fear for open or close spaces and patients displaying depression symptoms with a history of any other medical condition or substance abuse should be classified as Depressive Disorder due to Another Medical Condition and Substance/Medication-Induced Depressive Disorder accordingly. Several steps of the feature identification process has the potential to benefit in creating rules. Thus, after implementing a good interface and logic for the RDR System, we can start writing the general naive rules. In a industrial case, this task would be assigned to domain experts [29]. Utilizing their professional expertise, such domain experts decide which features are necessary and sufficient for a particular case. In this research, we along with a psychiatrist would be playing the roles of said domain experts and craft the generalized rules. From which, as depicted in figure 3.1 upon progressive validation and cornerstone case rule setting, we refine the knowledge base.

3.1 Data Collection

The data collection for the studies involved a self report measure where an anonymous survey questionnaire was distributed among accessible respondents. The survey contained questions in two sets, namely part A, and part B. Part A contained a total of 17 questions from the PHQ-9 and the GAD-7 in appendix A.1 and part B contained 16 additional questions from practicing psychiatrist and researcher, Dr. Shafika Afroz in appendix A.2, that were created from the DSM-5 [12] to aid the system with respect to her expert opinion.

Patient Health Questionnaire - 9 (PHQ-9):
The PHQ (Patient Health Questionnaire) is a self-report questionnaire that is used to screen for and assess the severity of common mental health disorders, including depression, anxiety, and substance abuse [4]. It consists of several questionnaires that cover different aspects of mental health and well-being, including mood, sleep, energy, and concentration. The PHQ-9 is a specific version of the PHQ that is used to screen for and assess the severity of depression. It consists of nine questions that are based on the diagnostic criteria for depression in the DSM-4 (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) [5]. The questions ask about the frequency of symptoms of depression over the past two weeks, such as feeling down, depressed, or hopeless, and losing interest in activities that the person usually enjoys. Each question is scored on a scale of 0 to 3, with higher scores indicating more severe symptoms. The total score can range from 0 to 27, with higher scores indicating more severe depression. Table 3.1 depicts the severity levels.

PHQ-9 has been found to be a reliable and valid measure of depression severity in a variety of populations [4]. It has also been found to have good internal consistency, meaning that the questions are all measuring the same underlying concept (i.e., depression). In addition, PHQ-9 has been found to be correlated with other well-established measures of depression, such as the Beck Depression Inventory, and it has been shown to be sensitive to changes in depression severity over time [13]. However, it is important to note that the PHQ-9 is intended to be used as a screen-

Table 3.1: Scores in PHQ-9

Score Range	Severity
0 to 4	None to minimal
5 to 9	Mild depression
10 to 14	Moderate depression
15 to 19	Moderately severe depression
20 to 27	Severe depression

ing tool to identify people who may be experiencing symptoms of depression and who may benefit from further evaluation by a mental health professional such as the expert in our system.

Generalized Anxiety Disorder - 7 Questionnaire:
The GAD-7 (Generalized Anxiety Disorder-7) is another self-report questionnaire that is used to screen for and assess the severity of generalized anxiety disorder (GAD) and other anxiety disorders using sumscores (i.e. summing the scores of each item) [8]. It consists of 7 questions that ask about the frequency of common symptoms of anxiety over the past two weeks, such as feeling excessively worried, having difficulty controlling worry, and feeling tense or on edge. Similarly to the PHQ-9, each question is scored on a scale of 0 to 3, meaning the total score can range from 0 to 21, with higher scores indicating more severe anxiety. Specifically, the severity is determined according to table 3.2.

Table 3.2: Scores in GAD-7

Score Range	Severity
0 to 4	No to Low risk
5 to 9	Mild anxiety
10 to 14	Moderate anxiety
15 to 21	Severe anxiety

In terms of internal consistency, reliability and validity for assessing anxiety, GAD-7 has shown promising results across a wide range of samples and settings [8]. The questionnaire has also been found to have acceptable sensitivity and specificity values at 0.77 and 0.82 respectively when used as a general tool to identify other anxiety disorders (Panic Disorder, Social Anxiety, and PTSD) [21].

Additional specific questions from psychologist:
Although both PHQ-9 and GAD-7 has remarkable reliability for identifying Major Depressive Disorder (MDD) and Generalized Anxiety Disorder (GAD) with the severity, in order to reach a conclusive result on differential diagnosis of the different sub-categories of the disorders, additional data is needed. Moreover, we require more information regarding patient history, behaviors and demographic. For this we consult with Dr. Shafika Afroz, to also ask the questions from part B of the questionnaire found in appendix A.2.

The purpose of these questions is to gather more information about the respondents and their specific demographics. This information will be used to further classify the respondents into different disease categories. The questionnaire was distributed to the general population in Dhaka, Bangladesh, with the aim of reaching as many people as possible. The use of the PHQ-9 and GAD-7 questionnaires has been proven to be effective in similar regions, particularly in rural areas of India [30]. Therefore, these questionnaires added with the new questions were chosen for this study.

3.2 Data Preprocessing

Before we proceed to analysis and implementation on the system, the collected data has to be pre-processed. This involves cleaning, transforming and preparing the data. After the collection of data from the survey respondents, we proceeded by first labeling them with the help of Dr. Shafika Afroz. Due to the nature of mental disorder in the end, the total target/label are found as the following table 3.3.

Table 3.3: Summary of the Labels/Targets

Disorder Name	**Count**
Mild MDD	2
Moderate MDD	8
Moderately Severe MDD	10
Severe MDD	4
Mild GAD	4
Moderate GAD	7
Severe GAD	23
Persistent Depressive Disorder**	14
Selective Mutism	2
Specific Phobia	13
Social Anxiety Disorder	12
Agoraphobia	24
Separation Anxiety Disorder**	22
Anxiety Disorder due to Another Medical Condition**	15
Depressive Disorder due to Another Medical Condition	6
Substance/Medication Induced Anxiety Disorder**	9
Substance/Medication Induced Depressive Disorder	8
No Disorder	33
Total Respondent	**112**

In the labels, MDD refers to Major Depressive Disorder and GAD refers to Generalized Anxiety Disorder. The label 'Mod Severe MDD' refers to moderately severe major depressive disorder. As suggested by our collaborator, Dr. Shafika Afroz, the diseases marked with two asterisks (**) are to be confirmed upon further consultancy by a clinical psychologist.

Our problem is that of a multi classification of data. Many of the data collected was not numeric and categorical. Therefore, we then proceed to appropriate means to encode each attribute/feature. Following are the techniques used for such.

Sum-scores: PHQ-9 and GAD-7 are standardized to be measured by summing up the scores in each question. PHQ-9 has 9 questions and GAD-7 has 7 questions, that after evaluating the sum-score, depicts the severity of depression and anxiety respectively. These 16 questions of part A of the questionnaire in appendix A.1, have these 4 options and are encoded as the table 3.4. As the response to each question is evaluated to 0 to 3, their categorical encoding is easily done. Afterwards the scores are added up to form the PHQ and GAD features name.

Table 3.4: Scores for each question in PHQ-9 and GAD-7

Response	**Score**
Not at all	0 points
Several days	1 point
More than half the days	2 points
Nearly every day	3 points

Both the PHQ-9 and GAD-7 tests include a question that asks about the level of difficulty the patient experiences in their daily life as a result of the problems addressed in the preceding questions. This question is the question numbered 17 in part A of our questionnaire found in appendix A.1 and is also categorically encoded and given the feature name GEN. While this question does not contribute to the PHQ-9 of GAD-7 score, psychologists may use this information for diagnostic purposes [5]. The categories are as the table 3.5 portrays.

Table 3.5: Score rating for difficulty in daily life

Response	**Score**
Not difficult at all	0
Somewhat difficult	1
Very difficult	2
Extremely difficult	3

Other Feature Encoding: The responses to the questions in part B in appendix A.2 were also in need of encoding the categorical values by label encoding and binary encoding, with the exception of the free responses in questions 13, 14 and 15. The exact techniques and feature labels are as the table 3.6.

Table 3.6: Encoding Details

Question Number	**Technique Used**	**Feature Name**
Part A		
PHQ-9	Sum-scores	PHQ
GAD-7	Sum-scores	GAD
17	Label Encoding	GEN
Part B		
1	Binary Encoding	F1
2	Binary Encoding	F2
3	Label Encoding	F3
4	Binary Encoding	F4
5	Binary Encoding	F5
6	Binary Encoding	F6
7	Label Encoding	F7
8	Binary Encoding	F8
9	Binary Encoding	F9
10	Binary Encoding	F10
11	Label Encoding	F11
12	Label Encoding	F12
13	Binary Encoding	F12
14	Binary Encoding	F12
15	Binary Encoding	F12
16	Label Encoding	F16

For the free answered questions 13, 14, 15. After the target was chosen, affirmative answers were marked as 1, and negative responses were marked as 0, making these values also binary encoded. As the MCRDR system does not require encoding the target labels, we then proceeded to the system and building rules for this dataset.In total, 112 respondents' data was processed carefully to implement in the final expert system.

A brief summary of the dataset collected is shown in the table 3.7 and the table 3.8.

Table 3.7: Summary of the Dataset (Part A)

1. Little interest or pleasure in doing things			
Not at all (0)	14	Several days (1)	60
More than half the days (2)	13	Nearly every day (3)	25
2. Feeling down, depressed, or hopeless			
Not at all (0)	10	Several days (1)	48
More than half the days (2)	22	Nearly every day (3)	32
3. Trouble falling or staying asleep, or sleeping too much			
Not at all (0)	26	Several days (1)	29
More than half the days (2)	17	Nearly every day (3)	40
4. Feeling tired or having little energy			

Not at all (0)	14	Several days (1)	40
More than half the days (2)	19	Nearly every day (3)	39
5. Poor appetite or overeating			
Not at all (0)	33	Several days (1)	29
More than half the days (2)	22	Nearly every day (3)	28
6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down			
Not at all (0)	16	Several days (1)	32
More than half the days (2)	18	Nearly every day (3)	46
7. Difficulty concentrating on things, such as reading the newspaper or watching television			
Not at all (0)	34	Several days (1)	33
More than half the days (2)	14	Nearly every day (3)	31
8. Moving or speaking so slowly that other people could have noticed Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual			
Not at all (0)	46	Several days (1)	29
More than half the days (2)	15	Nearly every day (3)	22
9. Thoughts that you would be better off dead, or of hurting yourself			
Not at all (0)	55	Several days (1)	21
More than half the days (2)	12	Nearly every day (3)	24
10. Feeling nervous, anxious, or on edge			
Not at all (0)	12	Several days (1)	39
More than half the days (2)	23	Nearly every day (3)	38
11. Not being able to stop or control worrying			
Not at all (0)	18	Several days (1)	29
More than half the days (2)	18	Nearly every day (3)	47
12. Worrying too much about different things			
Not at all (0)	10	Several days (1)	31
More than half the days (2)	21	Nearly every day (3)	50
13. Trouble relaxing			
Not at all (0)	26	Several days (1)	40
More than half the days (2)	20	Nearly every day (3)	26
14. Being so restless that it is hard to sit still			
Not at all (0)	43	Several days (1)	32
More than half the days (2)	16	Nearly every day (3)	21
15. Becoming easily annoyed or irritable			
Not at all (0)	24	Several days (1)	35
More than half the days (2)	22	Nearly every day (3)	31
16. Feeling afraid, as if something awful might happen			
Not at all (0)	22	Several days (1)	36
More than half the days (2)	14	Nearly every day (3)	40
17. If you checked off any problems, how difficult have these problems made it for you at work, home, or with other people?			
Not difficult at all (0)	6	Somewhat difficult (1)	68
Very difficult (2)	24	Extremely difficult (3)	14

Table 3.8: Summary of the Dataset (Part B)

1. Do you happen to experience temper outbursts in response to relatively milder situations multiple times through the week?			
No (0)	46	Yes (1)	66
2. Would you say that you are persistently irritable or angry most of the day, nearly every day?			
No (0)	69	Yes (1)	43
3. Do you have excessive and persistent worries about specific situations?			
Social situation (0)	16	Losing a major attachment figure (1)	29
Open space (2)	8	Closed space (3)	41
Any (4)	16	None of the above (5)	2
4. Does the specific situation always provoke anxiety?			
No (0)	35	Yes (1)	77
5. Is the worry out of proportion to the actual threat?			
No (0)	50	Yes (1)	62
6. Do you avoid the specific situation of worry or fear?			
No (0)	29	Yes (1)	83
7. Do you have physical symptoms for worry or fear?			
Negative (0)	23	Affirmative (1)	89
8. Do you have trouble speaking or unable to speak in specific situations in-spite of speaking in other situations?			
No (0)	49	Yes (1)	63
9. Is the worry and related symptoms affecting your day to day activities?			
No (0)	31	Yes (1)	81
10. Do you have excessive anxiety and worry (apprehensive expectation) occurring most of the day?			
No (0)	59	Yes (1)	53
11. Which one of the following age groups apply to you?			
12 to 17 (2)	3	18+ (3)	109
12. What gender do you identify as?			
Female (0)	55	Male (1)	56
Other (2)	1		
13. Have you ever been diagnosed with any mental disorder?			
Negative (0)	101	Affirmative (1)	11
14. Do you have any medical conditions present that may induce symptoms of mental disorder?			
Negative (0)	96	Affirmative (1)	16
15. Do you have any history of substance usage or medication that may induce symptoms of mental disorder?			
Negative (0)	101	Affirmative (1)	11
16. How long has it been since you experienced at least one of the symptoms you answered yes to?			
I don't have any of the symptoms (0)	24	Less than 2 weeks (1)	8

More than 2 weeks but less than 1 month (2)	3	More than 1 month but less than 6 months (3)	17
More than 6 months but less than 2 years (4)	32	More than 2 years (5)	28

For the targets, although the RDR does not require any encoding, using targets as conclusions directly, we still use binary encoding for each disorder to have the dataset usable by machine learning models. Thus the target was copied to produce a second set of data and the targets spread into 17 separate columns for each of the disorders we are capable of classifying. In a separate copy of the dataset, to aid multi-class classification techniques, the targets were separated to individual data-points. To explain, every case would be repeated for each of it's target, while keeping the features same. In this third set, there was 216 data-points in total.

3.3 Model Specification

We are comparing the performance of three models. These models are RDR model, Extreme Gradient Boosting (XGBoost) multiclass classifier model and a preferred XGB Binary Classification Block for multiple output classification. For our system, we are using RDR model while XGBoost multiclass classifier and XGB Binary Classification Block are being used to give a comparative result of RDR model.

3.3.1 RDR Model

Initially, the classification started with SCRDR which provides a single conclusion for each case. However, the system has drawbacks considering the domain of mental disorders. As a person can be identified with multiple disorders, the system that can provide only one conclusion is not sufficient enough. According to the psychiatrist, we had to adapt to a model that can give multiple conclusions for a case. Following the suggestions, we had shifted our model based on MCRDR which replaced SCRDR as core classification model. MCRDR is known as Multiple Classification Ripple-Down Rules. In MCRDR, there are two types of rules. They are parent rules that give conclusions and stopping rules that prevent the parent rules from giving wrong conclusions. If a case is not given its required conclusion, a parent rule is added. On the other hand, if a case is misclassified, a stopping rule is added against the rule that got fired for the case. Based on these two principles, rules are added. While evaluating a case, all the parent rules get checked which allows the model to give multiple classifications.

The RDR model can be divided into two parts. The one used for the development phase and the one used for the testing phase. As RDR is an incremental system, it can be developed while using it for classifying purposes. This idea is used for dividing the system in two phases for our research purpose. In the development phase, the system will be used to classify and by monitoring those classifications, new rules will be added if necessary. On the other hand, no rules can be added in the testing phase. For this reason, all the features for adding rules will not be allowed in our testing phase. However, for real life purposes, the training phase model will be used as it can both classify and add rules based on necessity.

Development Phase Model:

Development phase includes the model of the testing phase which only runs the existing rules for a case. Along with this, this phase includes some more features to add rules. Figure 3.2 represents the development phase model of MCRDR. So, the development phase has two basic functionalities and they are:

1. Evaluation and Stopping Rule Addition
2. Parent Rule Addition

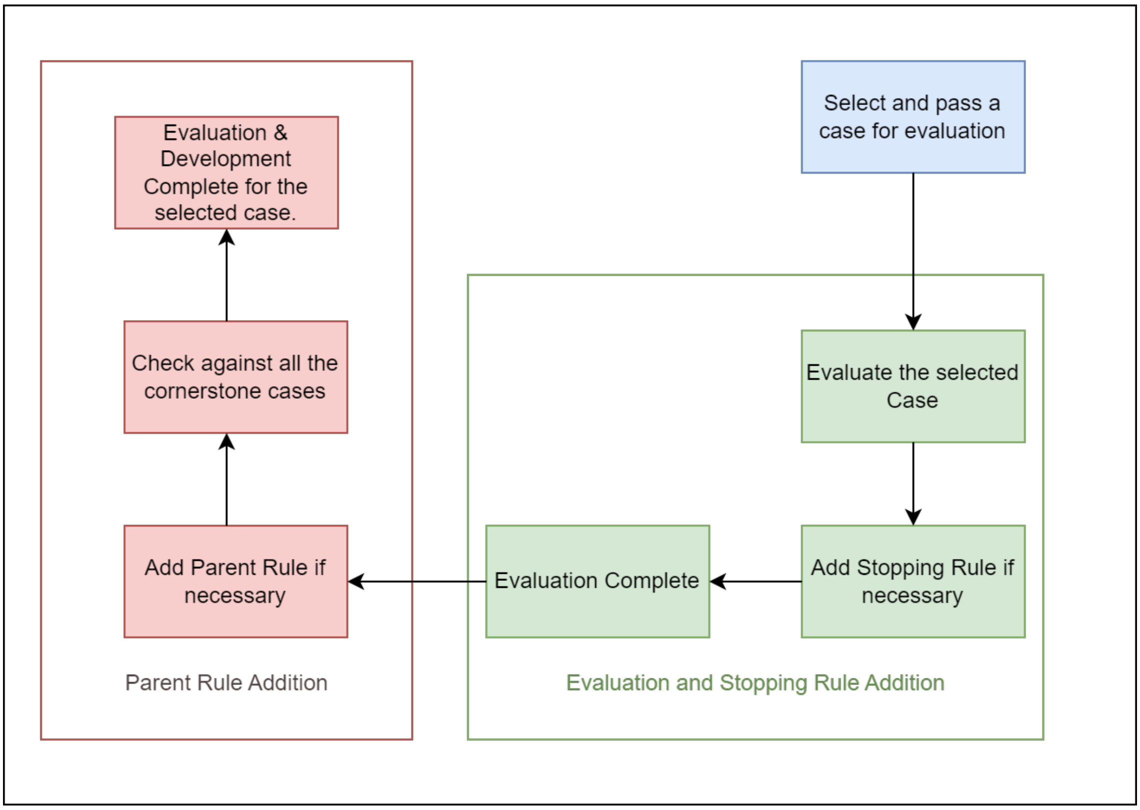


Figure 3.2: Development Architecture of MCRDR

For the first functionality, a case is selected for evaluation. First of all, all the parent rules are checked for the case. If any of the parent rules gets fired, then the system checks whether there is any stopping rule for that parent rule or not. If stopping rule exists for that parent rule, the system runs all of those stopping rules for the case. If any of the stopping rules fires, the conclusion of parent rule is not given and the system moves to the next available parent rule to check for the case. However, if there is no more stopping rule and none of them gets fired, the system checks whether the given conclusion of the parent rule is correct or not. If the conclusion is correct for the case, the system provides the conclusion and checks for the next available parent rule. In contrast, if the conclusion is wrong, a new stopping rule is added for that parent rule. First the features are selected that clearly aligns with the given case but doesn't align with the cornerstone case of the parent rule. Thus, it excludes the parent rule's cornerstone case not to get fired by this stopping rule. Lastly, in the conclusion segment of the rule addition feature, no conclusion is given since it is a stopping rule. Then, the system again goes to run the next available parent rule for the case. Figure 3.3 depicts the flow of activities for evaluation and stopping rule addition functionality of development phase.

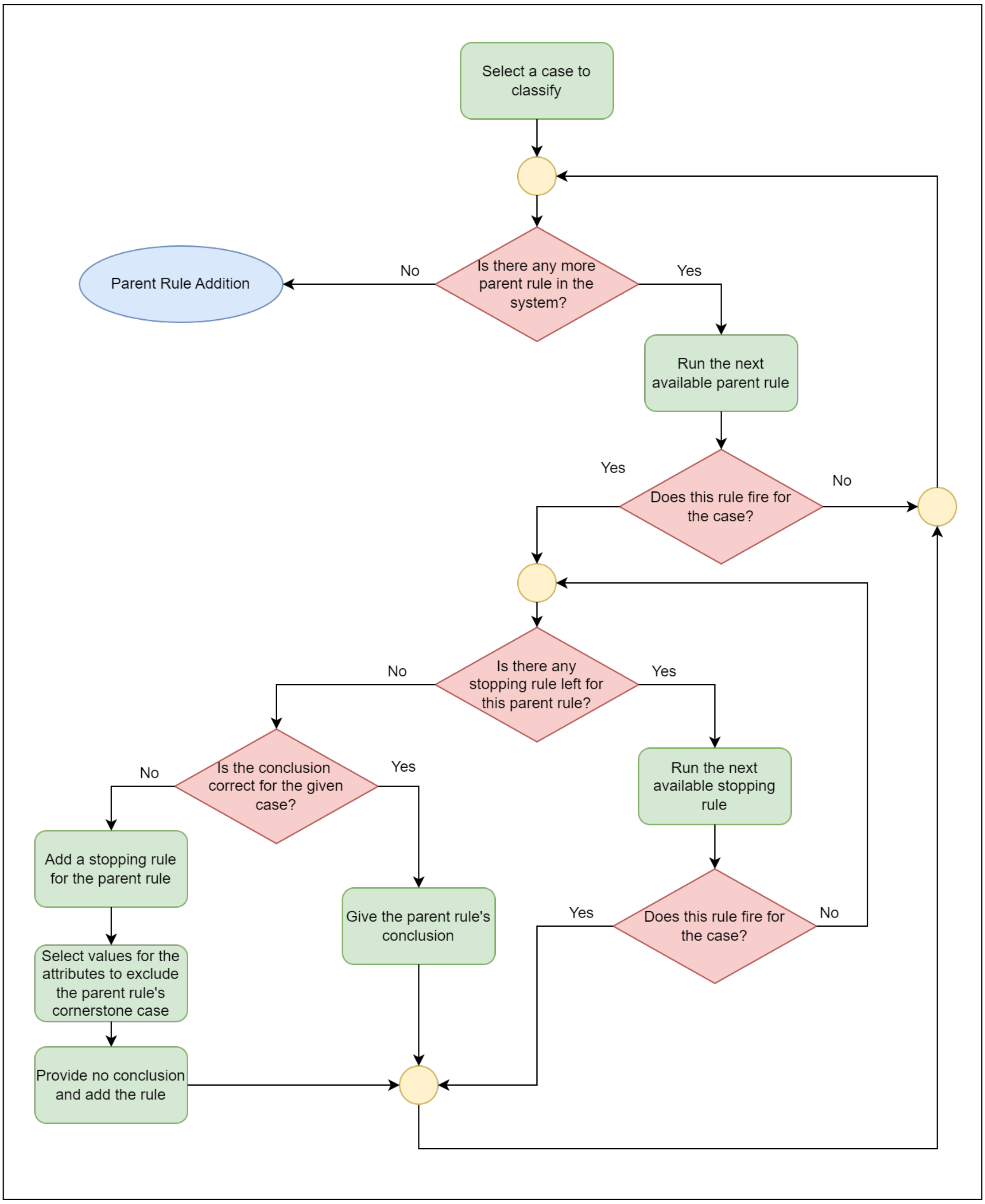


Figure 3.3: Flowchart for Evaluation and Stopping Rule Addition of Development Phase

After completing the evaluation, the system checks whether all the required conclusions are provided or not. If any of the conclusions are not given, the system jumps to adding a new rule for the case. All the values for the features are given along with providing a conclusion. While adding the rule, all the cornerstone cases are checked whether they fire the new rule or not. If none of them fire the new rule, the rule gets added and the case is added to the cornerstone cases list. However, if any of the previous cornerstone cases fire the rule, then it is checked if the given conclusion is correct for that case or not. If the conclusion is correct, the system goes to check the rule against the next cornerstone case. Contrarily, if the given conclusion is not correct, then that case needs to be excluded from firing this rule. This is done by setting different values to the features of rule addition that clearly don't align with that cornerstone case. Then, the system goes to check the next available cornerstone case whether that fires the on making rule or not. When all the cornerstone cases are checked, the initial case for which the rule had to be added goes to the cornerstone cases list. This is done since there is a new parent rule that is made because of this case. This process continues, checks all the cases, adds parent rules for the cases where necessary and also prevents parent rules from giving conclusion by adding a stopping rule if that's required. Figure 3.4 represents the flow of activities for parent rule addition functionality of development phase.

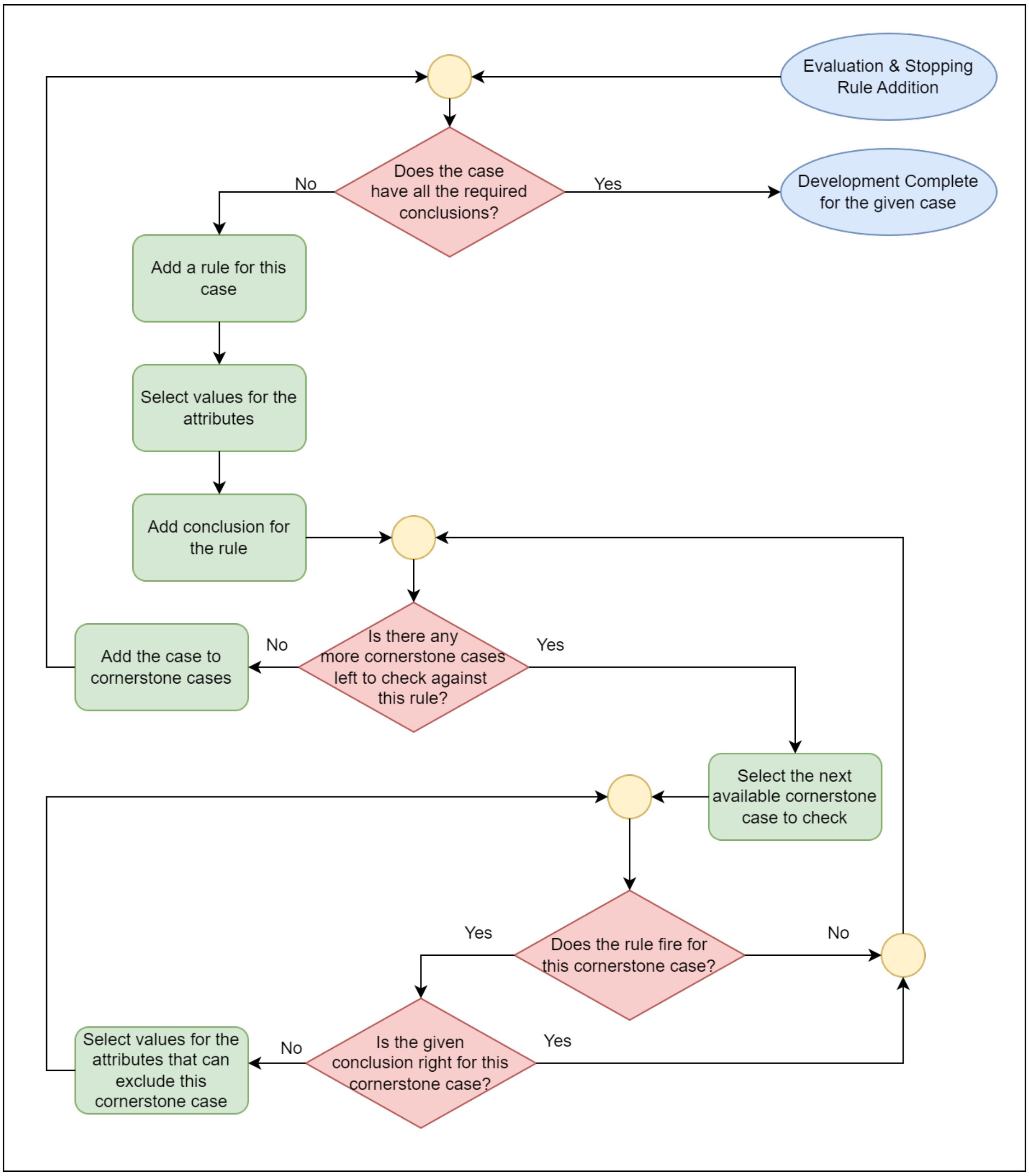


Figure 3.4: Flowchart for Parent Rule Addition of Development Phase

Combining and following these two parts, we are able to set 21 rules consulting with the psychiatrist. Among these 21 rules, 18 rules are parent rule whereas 3 rules are stopping rules. Upon crafting the rules, the system automatically generates and updates a decision tree. Table 3.9 shows the decision tree made by the MCRDR system. Each rule represents a node of the the tree where the first rule is the root. As described above, parent rules gives conclusion if none of their stopping rule matches with the case. Following this term, the decision tree made by our system is shown below in the table. The rules number surrounded by parenthesis are stopping rules and the rests are parent rules.

Table 3.9: Decision Tree made by the system

Rule	Go to if true	Go to if false
R1	(R21)	R2
(R21)	R2	R2 [Selective Mutism]
R2	R3 [Social Anxiety Disorder]	R3
R3	R4 [Substance/Medication Induced Depressive Disorder]	R4
R4	(R12)	R6
(R12)	R6	(R5)
(R5)	R6	R6 [Agoraphobia]
R6	R7 [Persistent Depressive Disorder**]	R7
R7	R8 [Substance/Medication Induced Anxiety Disorder**]	R8
R8	R9 [Severe GAD]	R9
R9	R10 [Separation Anxiety Disorder**]	R10
R10	R11 [Mod Severe MDD]	R11
R11	R13 [Moderate GAD]	R13
R13	R14 [Specific Phobia]	R14
R14	R15 [Anxiety Disorder due to Another Medical Condition**]	R15
R15	R16 [Moderate MDD]	R16
R16	R17 [Depressive Disorder due to Another Medical Condition]	R17
R17	R18 [Mild GAD]	R18
R18	R19 [Severe MDD]	R19
R19	R20 [Specific Phobia]	R20
R20	Exit [Mild MDD]	Exit

Testing Phase Model:

The testing phase model is almost like the evaluation part of the development model. To check the results of the MCRDR system, all kinds of rule addition is prevented in this phase. An input is provided to the MCRDR model which gives output based on the case. A single MCRDR model is capable of providing multiple output for a case. Figure 3.5 depicts the MCRDR testing phase model.

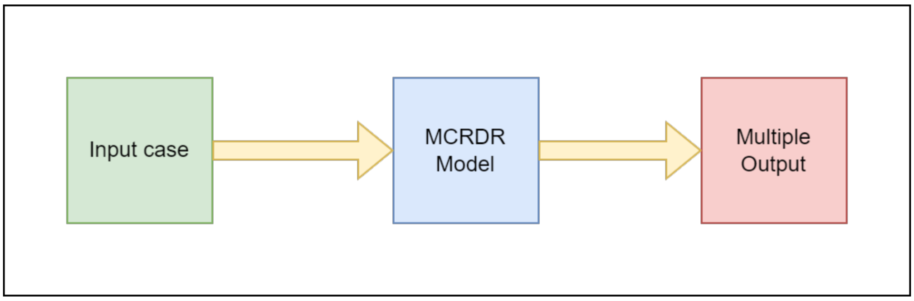


Figure 3.5: MCRDR testing phase model

At the beginning a case is selected to be evaluated. Then, all the parent rules are checked for this case. If any of the parent rules gets fired, the system checks if the parent rule has any stopping rule or not. If there is no stopping rule, the conclusion of the parent rule is given. Contrarily, if there are stopping rules for the parent rule, all of them are checked for the case one by one. If any of the stopping rules gets fired, the system stops checking the rest of the stopping rules for that parent rule since this means that the conclusion of the parent rule should not be given. As a result of this, the conclusion of the parent rule is not given and the system runs the next available parent rule. However, if none of the stopping rules gets fired, the system gives the conclusion of the parent rule. For example, we can see from our decision tree shown at Table 3.9 that Rule No 1 has one stopping rule and that is Rule No 21. If any case matches with Rule No 1, the system will check whether the case matches with Rule No 21 or not. If the case matches with the Rule No 21, the system will proceed to next parent rule which is in our case Rule No 2 without giving any conclusion. On the other hand, if the selected case doesn't match with Rule No 21, the system will proceed to next parent rule (Rule No 2) but this time it will provide conclusion 'Selective Mutism' since none of the stopping rule fired. Figure 3.6 depicts the testing phase flowchart of MCRDR.

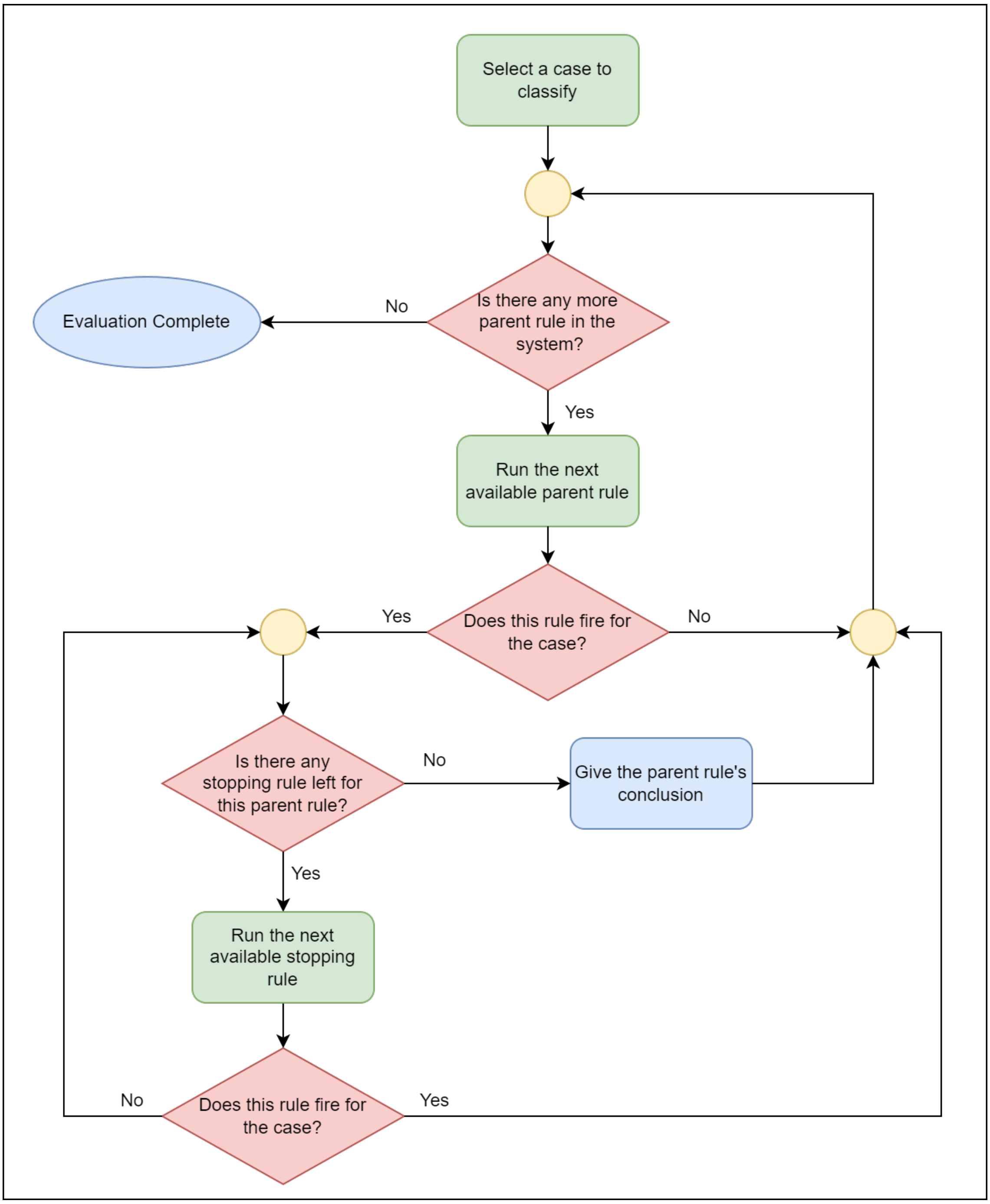


Figure 3.6: Flowchart for Testing Phase

3.3.2 XGBoost Multiclass Classifier Model

XGBoost is a scalable machine learning system for tree boosting. The most crucial aspect of XGBoost's success is its scalability in all situations. The system operates more than 10 times quicker than current popular solutions on a single machine and scales to billions of samples in distributed or memory-constrained environments. XGBoost's scalability is the result of several significant system and algorithmic enhancements [18].

Unlike many other algorithms, XGBoost is an ensemble learning method, which

uses the findings of several models, known as base learners, to create a prediction. XGBoost's trees differ slightly from conventional decision trees. They are known as CART trees (Classification and Regression trees), and instead of a single decision in each ”leaf” node, they hold real-value scores that indicate whether an instance belongs to a group. After the tree reaches its maximum depth, a judgment may be taken by translating the scores into categories using a specified threshold.
In our case, we are using an XGBoost multiclass classifier to our preprocessed dataset that has been prepared for multiclass classification. The model is trained by the cases made from first 89 instances. This model fails to provide multiple output. Through this model, we are showing that the traditional XGBoost model provides only one output which is a shortcoming of the model. An input case will be provided to the XGBoost multiclass classifier model and the model will provide only one conclusion out of 18 conclusions (17 disorders and 1 output if there is no disorder). Figure 3.7 shows the XGBoost multiclass classifier model.

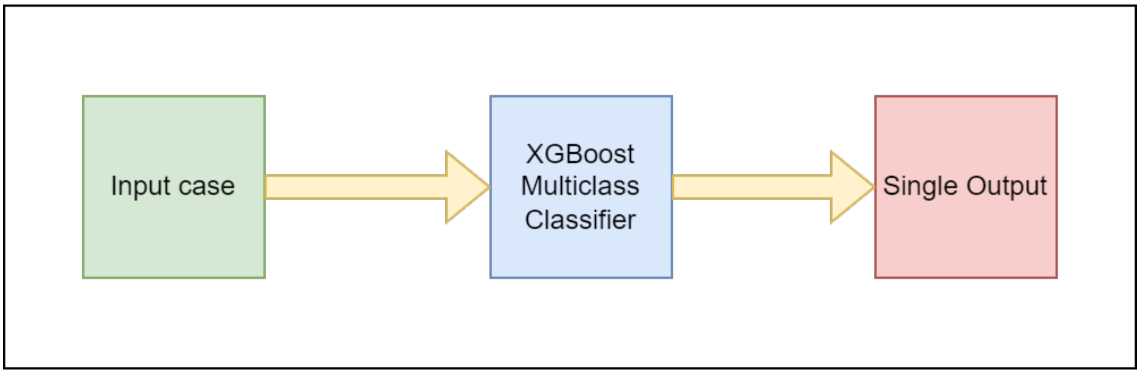


Figure 3.7: XGBoost multiclass classifier model

3.3.3 XGB Binary Classification Block Model

In the previous model specification, it is clear that XGBoost multiclass classifier is not able to provide multiple output for a single case. Everytime the model will provide only a single output for a case even though the case belongs to multiple classes. To overcome this proplem, we present an XGB Binary Classification Block model that works with the help of multiple XGB Binary Classifier models. There are different XGBoost binary classifier for each class in the XGB Binary Classification Block. Like, we are working on 17 disorders and if we consider each disorder as a class, XGB Binary Classification Block will have 17 XGBoost binary classifier for each disorder. These individual binary classifiers are trained to classify only one disorder. As shown in Figure 3.8, for the disorder 'Mild MDD' there is an XGBoost binary classifier, same goes for the other disorders. XGB Binary Classification Block later combines these outputs provided by 17 XGBoost binary classifiers to provide multiple output.

For exapmle, if a case is given as input, the case will be passed to all the bianry classifiers made for different disorders. Each of these classifiers will provide a single output of either positive(1) or negative(0). If the classifier for 'Severe MDD' provides 1, the case is diagnosed with 'Severe MDD'. Contrarily, if the classifier for 'Severe MDD' provides 0, the case is not diagnosed with 'Severe MDD'. XGB Binary

Classification Block provides 17 predictions from these 17 classifiers. Later, these predictions are combined in a single output with multiple conclusions. Thus, the model is capable of providing multiple outputs.

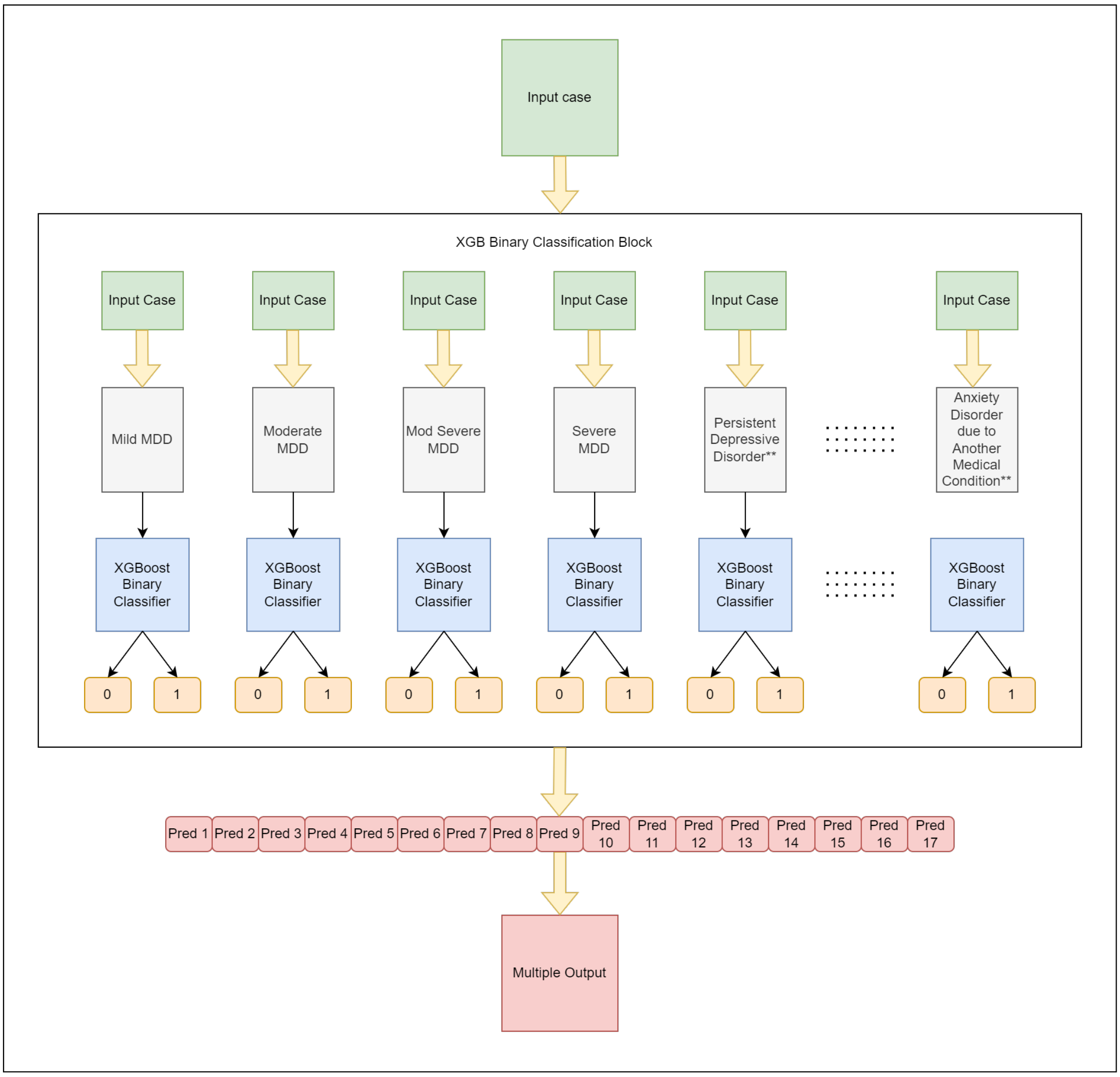


Figure 3.8: XGB Binary Classification Block model

Chapter 4

Implementation & Performance Evaluation

In this chapter, we are going to describe the performance metrics that are used to evaluate our system. Secondly, the results from the system based on the performance metrics and finally we are going to depict the implementation of our system in real world.

4.1 Performance Metrics

After carrying out the customary activities, putting a model into action and obtaining some output in the form of a probability or a class, the following step is to determine how effective the model is based on some metric by employing test datasets. Since we are developing a system that can diagnose a person with some disorders, we can use the performance metrics used for classification problems. In our case, we are going to check our system's performance by evaluating accuracy, precision, recall and F1 score. Another performance metric, confusion matrix will be used to calculate some of these metrics. Each of these metrics are elaborately described below along with how they are calculated.

Confusion Matrix: A confusion matrix includes information on the model's actual and expected classifications. The amount of accurate and inaccurate predictions is summed using count values and split down by class. Then, the count values are shown in a matrix manner. To create a confusion matrix, the predicted and actual class labels are compared, and the number of accurate and wrong guesses for each class is tallied. The counts are then arranged in a matrix, with the actual classes serving as rows and the anticipated classes as columns. The diagonal components of the matrix reflect accurate forecasts, whereas the other elements represent inaccurate guesses. Confusion matrix is a helpful technique for assessing the performance of a model, particularly when the data is unbalanced. In a binary classification issue (Positive and Negative), for instance, the matrix will appear as the table 4.1.

Table 4.1: Confusion Matrix Parameters

	Predicted (0)	**Predicted (1)**
Actual (0)	True Negative (TN)	False Positive (FP)
Actual (1)	False Negative (FN)	True Positive (TP)

Specifically, the terms are defined as:

- True Negative: Number of observations labeled as Negative and predicted as Negative
- True Positive: Number of observations labeled as Positive and predicted as Positive
- False Negative: Number of observations labeled as Positive but predicted as Negative
- False Positive: Number of observations labeled as Negative but predicted as Positive

Accuracy: Accuracy is the proportion of correct predictions made by a model relative to the total number of predictions made. It is determined by dividing the number of right guesses by the number of total predictions. Accuracy is a straightforward statistic that offers an overall sense of how well a model is doing. However, accuracy might be deceptive if the data are unbalanced, i.e., if one class is much more frequent than the other. In such situations, a model that always predicts the majority class can attain great accuracy, but it will not be practical. In our case, our data are imbalanced and because of this, we are going to use other metrics along with accuracy.

Accuracy is calculated as the equation 4.1:

$$Accuracy = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total number of predictions}} \tag{4.1}$$

True Positives are the number of observations correctly classed as positive (belonging to the positive class), while True Negatives are the number of observations correctly classified as negative (belonging to the negative class).

Precision: Precision is the ratio of accurate positive predictions produced by a model to the total number of positive predictions made. It measures the model's ability to avoid producing false positive predictions. A precision number suggests a low rate of false positives for the model. When the cost of false positives is large, precision is especially advantageous. When working with unbalanced datasets, precision alone might be deceiving. In our situation, we will utilize it in conjunction with recall and F1-Score.

Precision is calculated as the equation 4.2.

$$Precision = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{4.2}$$

True Positives are the number of observations correctly classed as positive (belonging to the positive class), while False Positives are the number of observations incorrectly classified as Positive (belonging to the negative class).

Recall: Recall, also known as sensitivity or true positive rate, quantifies the proportion of accurate positive predictions generated by a model relative to the total number of actual positive observations. It measures the ability of the model to accurately recognize all positive observations. A high recall value suggests that the model has a low false-negative rate. When the cost of false negatives is significant, recall is particularly valuable. When working with unbalanced datasets, recall alone might be deceptive. For this reason, we also employ F1-Score.

Recall is calculated as the equation 4.3:

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{4.3}$$

True Positives are the number of observations correctly classed as positive (belonging to the positive class), while False Negatives are the number of observations incorrectly classified as Negative (belonging to the positive class).

F1-Score: F1-Score is a harmonic mean of precision and recall. It goes from 0 to 1, with a greater number signifying superior performance. It is used to balance the trade-off between accuracy and recall. When the cost of false positives and false negatives is substantial, the F1-Score is an excellent optimization statistic. Due to having an uneven class distribution (Negative cases are higher than positive cases), we are going to use F1-Score as our main performance metric.

F1-Score is calculated as the equation 4.4.

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{4.4}$$

4.2 Results and Discussion

In this section, we discuss the outcomes of the models and evaluate them based on the performance metrics. After splitting the dataset, we got 23 data points for testing and we will be evaluating our models by passing on these data points.

Prior to proceeding to the evaluation of the models, we also check the consistency of the dataset collected without the labels. For this, we make use of Cronbach alpha values for the questions relating to PHQ-9, GAD-7 and the additional demographic and other questions. Table 4.2 represents the internal consistency results for these three question sets.

Table 4.2: Internal Consistency

Type	**Cronbach Alpha**
Questions from PHQ-9	0.83
Questions from GAD-7	0.88
Other Questions	0.47

RDR Model Results: MCRDR can provide multiple outputs for a single case. A case can have 1 prediction for each disorder. Like, if the conclusion for a case us 'Mild MDD, Agoraphobia', the case is dianosed with 'Mild MDD' and 'Agoraphobia'. So, the values predicted for 'Mild MDD' and Agoraphobia' will be 1 and for all other disorders the value will be 0. This means for a single case, there are 17 outputs that we need to consider. So, we will be getting $23 \times 17 = 391$ predictions in total for our 23 cases. After applying the MCRDR model in the test dataset, all the conclusions given by the systems matched with the labels provided by the psychiatrist. For this reason, all the predictions made by MCRDR matches with all the target values of the test dataset. Figure 4.1 shows the confusion matrix for the MCRDR model.

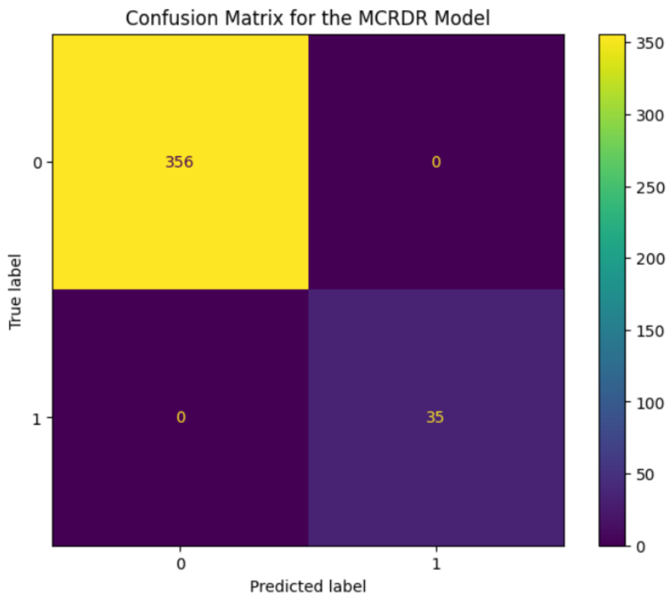


Figure 4.1: Confusion matrix from MCRDR predictions

From the confusion matrix, it is visible that the MCRDR system successfully classifies all the test cases. Among 356 negative cases for different disorders, the system predicts all of them to be negative. Same goes for the 35 positive cases. For this reason, there are no False Negative and False Positive case in the confusion matrix. So, the accuracy will be as shown in equation 4.5.

$$\begin{aligned} Accuracy &= \frac{\text{Total correct predictions for 17 disorders of all 23 cases}}{\text{Total target instances}} \times 100 \\ &= \frac{391}{391} \times 100 = 100\% \end{aligned} \tag{4.5}$$

Table 4.3: Classification report from MCRDR predicitons

	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	1.00	1.00	356
1	1.00	1.00	1.00	35
Accuracy			**1.00**	391
Macro Avg	1.00	1.00	1.00	391
Weighted Avg	**1.00**	**1.00**	**1.00**	391

Table 4.3 shows the classification report of MCRDR model. It is no wonder that the classification report shows all the precision, recall, f1-score and accuracy values are 1.00 for both positive and negative cases, since the system could successfully classify all the cases. As the precision value is very high, our system can avoid producing false positive predictions. Similarly, the system can accurately recognize

all the positive observations since the recall value is high.

XGBoost Multiclass Classifier Model Results: There are multiple outputs in our target label. We had to clone the cases that have multiple outputs to make the dataset suitable for multiclass classification. If a case has 3 outputs, that case has 3 instance in the dataset each time having different conclusions from those 3 outputs. In this way, our 89 training cases have become 173 instances and 23 testing cases has become 43 instances. It is worth mentioning that these 43 instances are clone of those 23 testing cases. Figure 4.2 shows the confusion matrix of the predictions when we tested the model with those 43 instances.

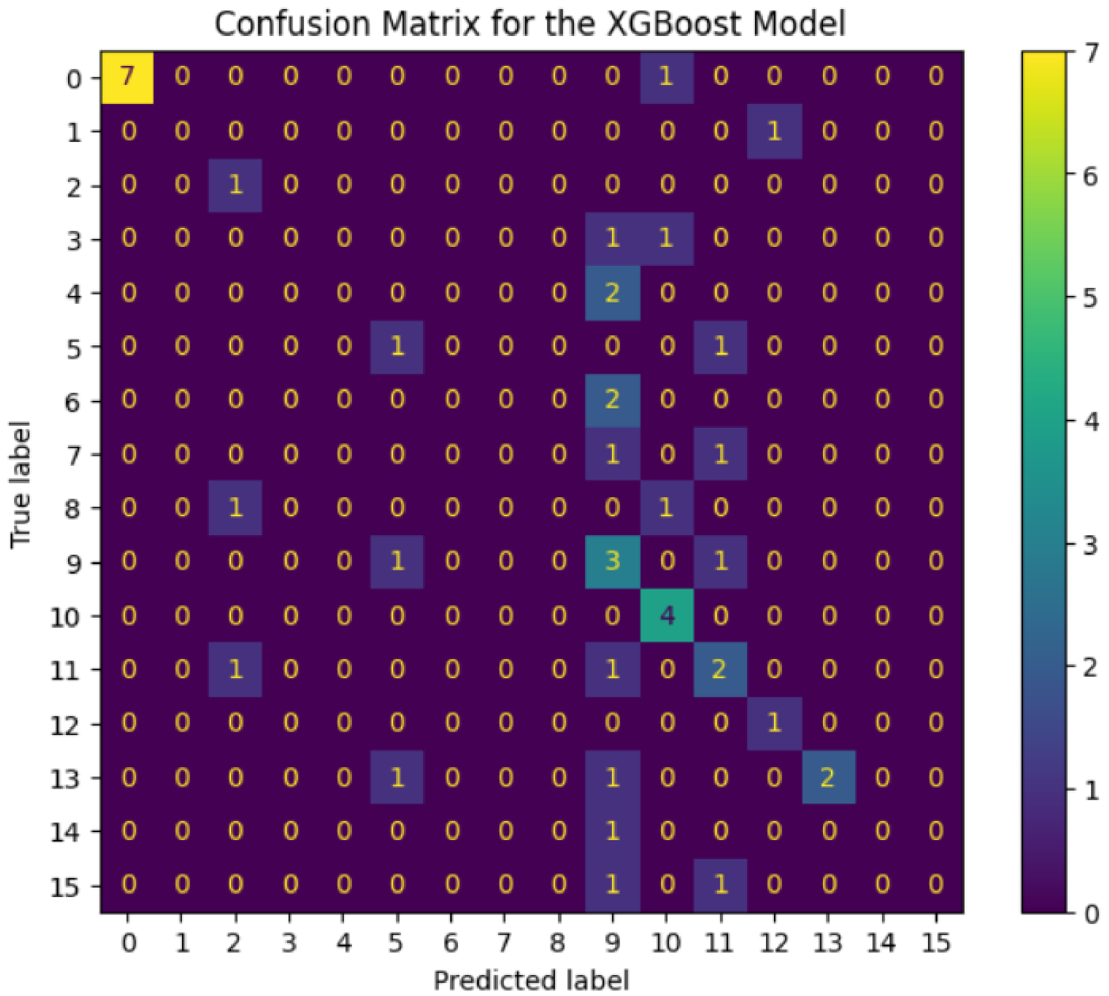


Figure 4.2: Confusion matrix from XGBoost multiclass classifier model predictions

The confusion matrix shows that out of 43 instances, the model can correctly classify only 21 instances correctly. If we calculate the accuracy for this case, the accuracy will be as the equation 4.6.

$$Accuracy = \frac{21}{43} \times 100 = 49\% \tag{4.6}$$

Table 4.4: Classification report from XGBoost multiclass classifier predicitons

	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.88	0.93	8
1	0.00	0.00	0.00	1
2	0.33	1.00	0.50	1
3	0.00	0.00	0.00	2
4	0.00	0.00	0.00	2
5	0.33	0.50	0.40	2
6	0.00	0.00	0.00	2
7	0.00	0.00	0.00	2
9	0.00	0.00	0.00	2
10	0.23	0.60	0.33	5
11	0.57	1.00	0.73	4
12	0.33	0.50	0.40	4
14	0.50	1.00	0.67	1
15	1.00	0.50	0.67	4
16	0.00	0.00	0.00	1
17	0.00	0.00	0.00	2
Average			**0.49**	43
Macro Avg	0.27	0.37	0.29	43
Weighted Avg	**0.42**	**0.49**	**0.43**	43

Table 4.4 shows the classification report of XGBoost multiclass classifier model. The classification report suggests that the accuracy of the model is only 49%. One of the reason for this is because, if a case has three instances, 3 of them will have different targets in the target label of test dataset. Since the model uses same decision tree to produce only one output, for each of those cases, the model provides the same output. This is the main disadvantage of this model.

XGB Binary Classification Block Model Results: As discussed earlier, this model provides 17 predictions for each case and those predictions are later combined to provide multiple outputs. Figure 4.3 demonstrates the correct predictions as well as the accuracy of these 17 binary classifiers for 23 examples. Here, it is evident that the model is able to classify certain disorders without making inaccurate predictions. However, the accuracy of the binary classifiers used to diagnose Social Anxiety Disorder, Moderate GAD, and Severe MDD is 87%, 91%, and 91%, respectively. Since the majority of XGB Binary Classification Block classifiers perform adequately, the average mean will work fine.

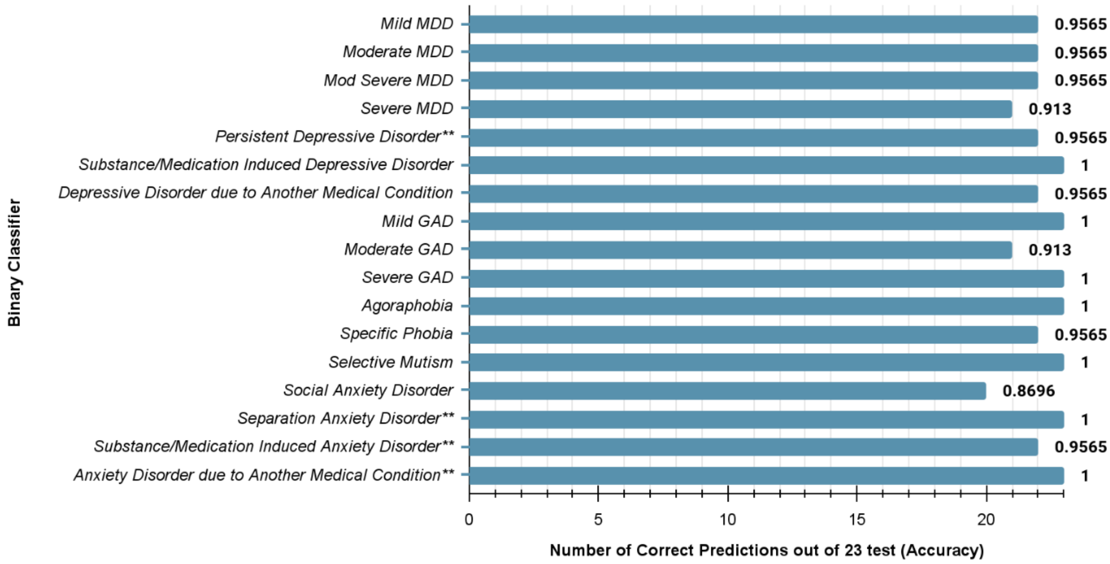


Figure 4.3: Accuracy of each binary classifier model after evaluating on 23 cases

The average accuracy of this model will be calculated as the same way as we did for MCRDR model. So, 23 cases will have in total 23 $\times 17 = 391$ predictions. Figure 4.4 shows the confusion matrix after applying XGB Binary Classification Block.

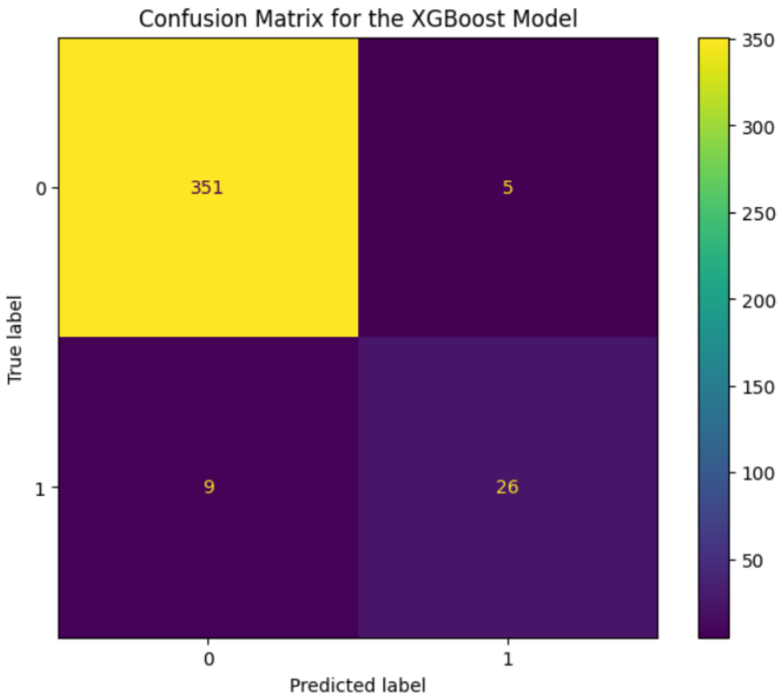


Figure 4.4: Confusion matrix from XGBoost Classifier predictions

The confusion matrix shows that among 391 predictions, a total of 377 predictions are correct. Among these 377 predictions, 351 cases are True Negatives whereas 26 cases are True Positives. However, there are 9 False Negative cases and 5 False Positive cases. This adds up to the number of misclassication which is 14 in total. However, it is worth mentioning that among these 391 predictions, 23 of them are for 'Mild MDD', 23 of them are for 'Moderate MDD' and so on. If we calculate the accuracy for this case, the accuracy will be as equation 4.7.

$$
\begin{aligned}
Accuracy &= \frac{\text{Total correct predictions for 17 disorders of all 23 cases}}{\text{Total target instances}} \times 100 \\
&= \frac{377}{391} \times 100 = 96\%
\end{aligned} \tag{4.7}
$$

Table 4.5: Classification report from XGB Binary Classification Block predicitons

	Precision	**Recall**	**F1-Score**	**Support**
0	0.97	0.99	0.98	356
1	0.84	0.74	0.79	35
Accuracy			**0.96**	391
Macro Avg	0.91	0.86	0.88	391
Weighted Avg	**0.96**	**0.96**	**0.96**	391

Table 4.5 depicts the classification report of XGB Binary Classification Block model. According to the confusion matrix, the classification report shows an accuracy of 96% for the XGBoost model. However, it is also noteworthy that the precision, recall, f1-score for classifying positive cases is significantly lower than compared to the precision, recall, f1-score for classifying negative cases. This happened since we have a lower number of positive cases in our data. Out of 391 cases, positive cases have support of only 35 cases compared to negative cases having support of 356 cases. As the precision value for positive cases is 0.84, the model might produce 1 false positive case among 6 predictions on average. Similarly, the recall value of 0.74 suggests that among 4 actual positive cases, XGBoost classifier can predict 3 of them correctly.

Table 4.6: Accuracy comparison among the models after cross validation

Model Name	**Accuracy on Shuffle 1**	**Accuracy on Shuffle 2**	**Accuracy on Shuffle 3**
MCRDR	100%	100%	100%
XGBoost Multiclass Classifier Model	49%	49%	45%
XGB Binary Classification Block Model	96%	96%	95%

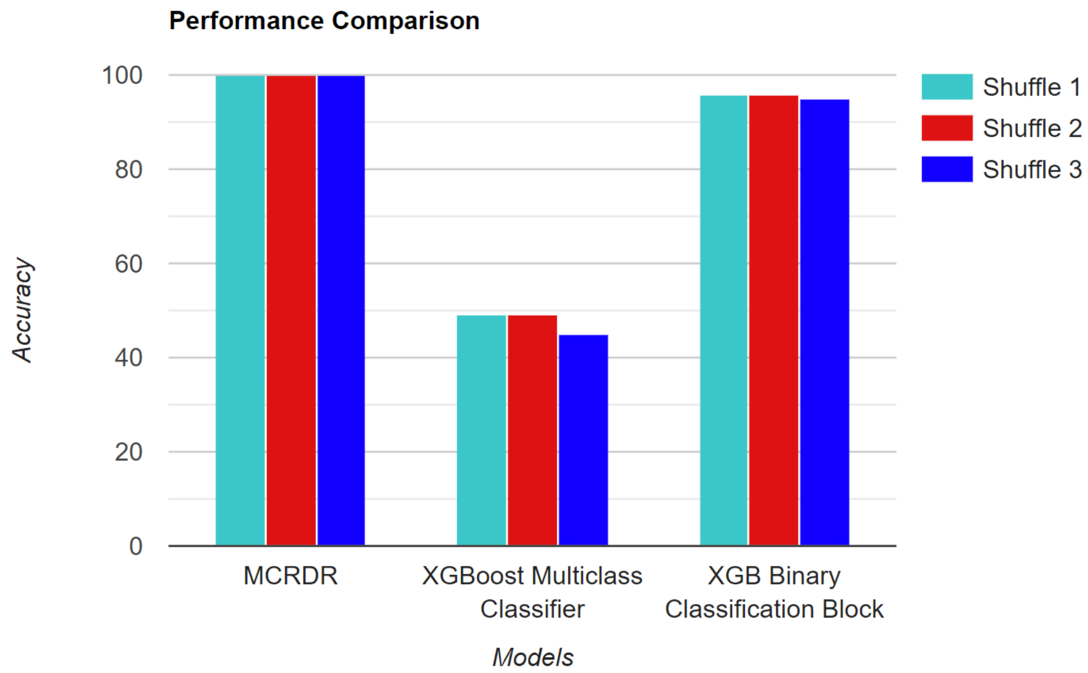


Figure 4.5: Accuracy comparison among the models after cross validation

Table 4.6 and figure 4.5 represents the comparison between the accuracies of the models. Since the data size is limited, we have randomized the sequence of the data on the dataset. Thus, we made two more shuffles of datasets by randomizing the sequence and cross validated the models based on these two shuffles. If we compare the accuracies of the models on all 3 shuffles, it is visible that MCRDR edges both of the XGBoost classifiers on all 3 occasions. MCRDR model has accuracy of 100% on all 3 shuffles, where XGBoost Multiclass Classifier Model has accuracy of 49%, 49%, 45% respectively and XGB Binary Classification Block Model has accuracy of 96%, 96% and 95%. The performance of the multiclass classifier is not so good and the reason is that this classifier is not able to provide multiple outputs for a case. However, the accuracies of the MCRDR model and the suggested XGB Binary Classification Block model are quite good. This might seem that they have close accuracies but when we consider the precision and recall value of both models in classifying positive cases on shuffle 1, MCRDR gives a significant amount of better performance. From table 4.3, it is visible that MCRDR has 1.0 for both precision and recall value of the positive cases. Contrarily, table 4.5 shows that XGB Binary Classification Block has 0.84 and 0.74 as precision and recall value respectively for positive cases. This happens because a machine learning algorithm's performance is not up to the mark if there is a small amount of data for a class. On the other hand, RDR can perform simultaneously good for all the classes even if the amount of data is not that big for a class. However, it is worth mentioning that the presented XGB Binary Classification Block can provide multiple outputs for a single case like MCRDR.

After evaluating the results, it can be stated that the outcome of applying MCRDR to the test dataset is quite extraordinary. One of the primary reasons for this is because rules have been established consulting with an experienced psychiatrist. As a result, the rules are extremely accurate and may correctly identify a person with the disorders.

Another metric used by Compton and Kang in their book [29] to check how many cases the system can correctly classify without adding the next rule. This can be done for both development and testing phase. Since the number of rules will be lower at the beginning, intially more rules need to be added for the cases. Eventually, the system starts to evaluate cases without producing any error when a sufficient amount of rules has been added. The graph for rule no vs the case no for which the rule has been added shows a significant detail about the system. The number cases that has been correctly evaluated by the system during our development phase has been shown in the figure 4.6.

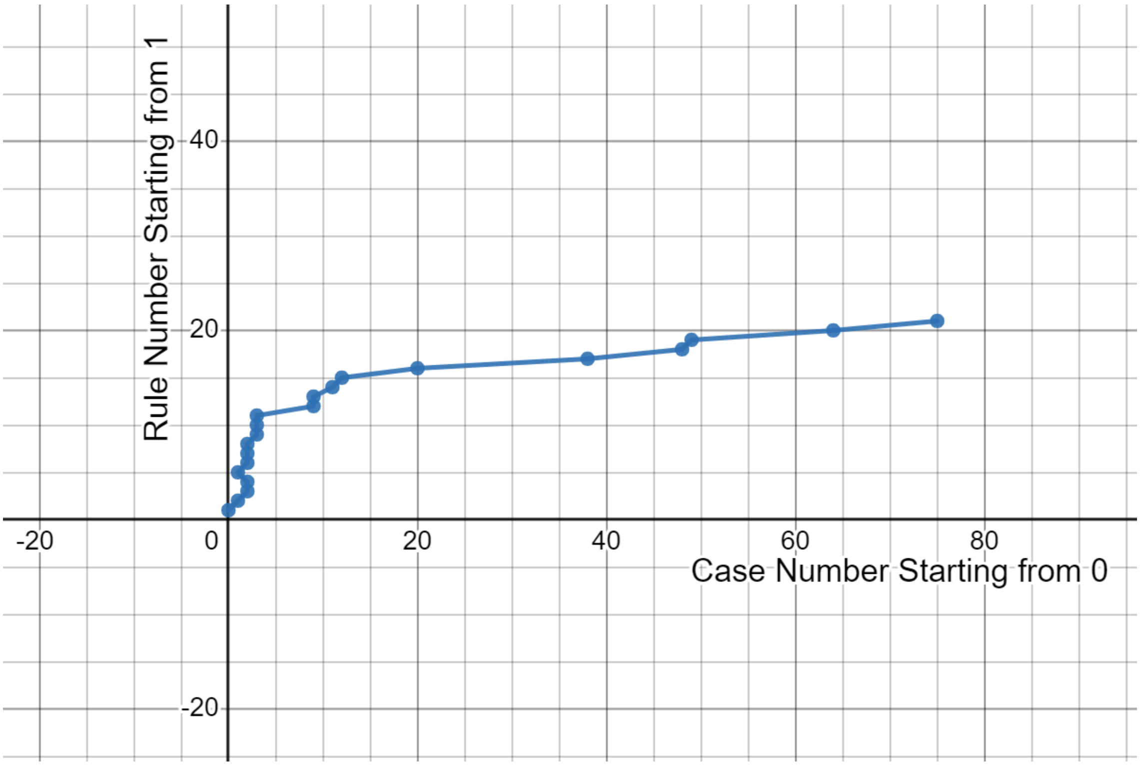


Figure 4.6: Rule vs Case Graph

From the graph in figure 4.6, it is visible that we needed 16 rules to evaluate first 21 cases (0 to 20) in our development phase. However, we had to add only 5 rules while evaluating rest of the cases (21 to 88) in total 68 cases. As discussed above, initially the cases might produce more error since the number of rules is low. It is also visible in the graph. For case number 2, we had to add 5 rules. Based on this graph, it can be stated that once the system is developed with enough data, the expert will need to add only a few amount of rules after evaluating large amount of cases as it is visible for the cases 21 to 88.

4.3 Implementation

To implement an MCRDR based system for diagnosing Mental Disorders, a web-interface was built in such fashion that it could be deployed in any machine capable of running python code. The interface has a total of five(5) primary pages among which only one is accessible to the public users. The other pages require an expert to sign in using provided credentials by the system admins. These pages are used to develop, evaluate and monitor the performance of the MCRDR system. The description of the pages are given below:

Homepage: Figure 4.7 shows the homepage of the system. This is the landing page of the web-interface and consists of a series of questions that was made in the supervision of an expert to evaluate the mental condition of any users. Any user can use this page to answer the questions and after submitting the form they will be provided a diagnosis of a possible mental disorder that they may potentially have. This prediction will come from the same system following the rules set by experts.

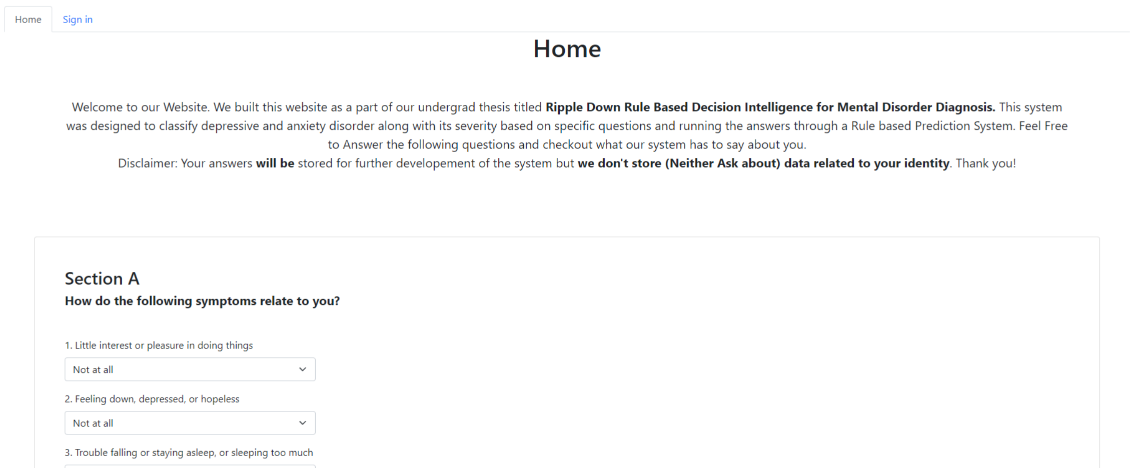
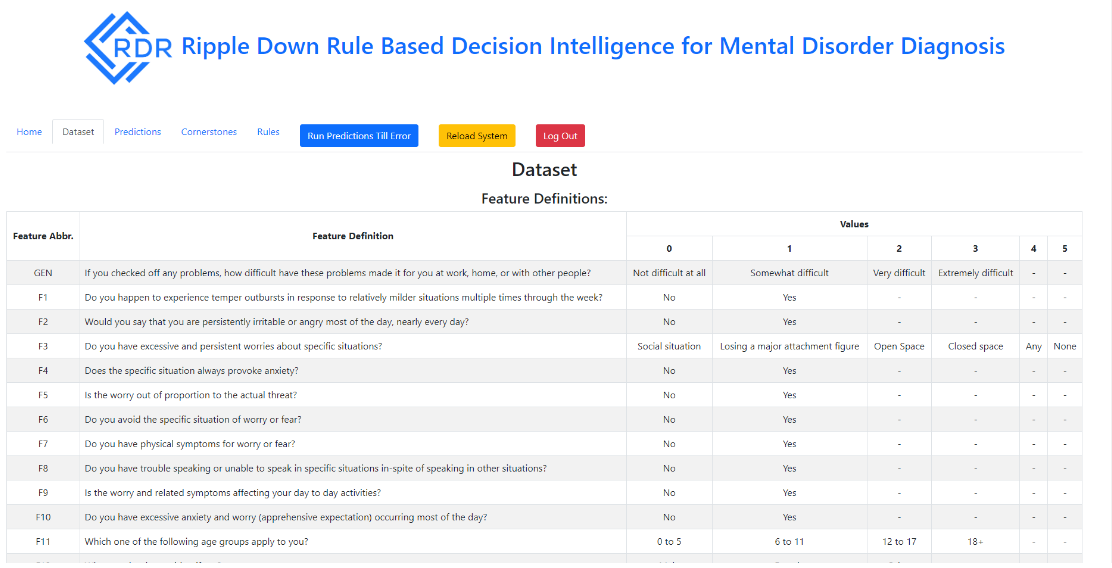


Figure 4.7: Homepage of the System

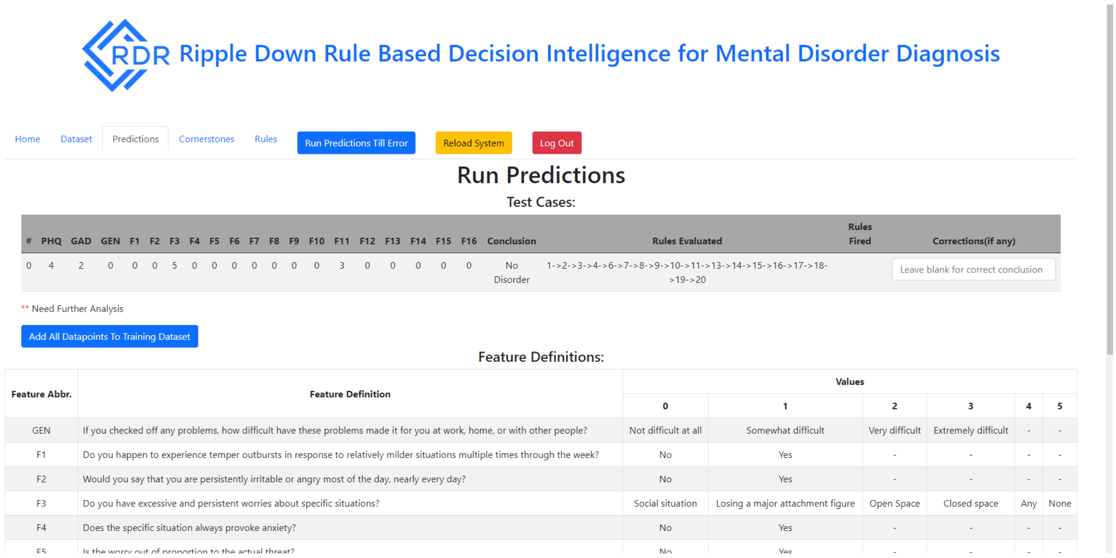
Dataset Page: Figure 4.8 presents the Dataset page of the system. This page requires login. This page consists of the primary training dataset with any additional data points added by an expert using the 'prediction page'. An expert may add any amount of primary rules or stopping rules using this page. They can also run predictions using existing rules and view which rules are evaluated or fired for a particular data point.



Feature Abbr.	Feature Definition	Values 0	1	2	3	4	5
GEN	If you checked off any problems, how difficult have these problems made it for you at work, home, or with other people?	Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult	-	-
F1	Do you happen to experience temper outbursts in response to relatively milder situations multiple times through the week?	No	Yes	-	-	-	-
F2	Would you say that you are persistently irritable or angry most of the day, nearly every day?	No	Yes	-	-	-	-
F3	Do you have excessive and persistent worries about specific situations?	Social situation	Losing a major attachment figure	Open Space	Closed space	Any	None
F4	Does the specific situation always provoke anxiety?	No	Yes	-	-	-	-
F5	Is the worry out of proportion to the actual threat?	No	Yes	-	-	-	-
F6	Do you avoid the specific situation of worry or fear?	No	Yes	-	-	-	-
F7	Do you have physical symptoms for worry or fear?	No	Yes	-	-	-	-
F8	Do you have trouble speaking or unable to speak in specific situations in-spite of speaking in other situations?	No	Yes	-	-	-	-
F9	Is the worry and related symptoms affecting your day to day activities?	No	Yes	-	-	-	-
F10	Do you have excessive anxiety and worry (apprehensive expectation) occurring most of the day?	No	Yes	-	-	-	-
F11	Which one of the following age groups apply to you?	0 to 5	6 to 11	12 to 17	18+	-	-

Figure 4.8: Dataset page of the System

Prediction Page: Figure 4.9 depicts the Implementation page of the system. This page requires login. This page will display all the new data points that the system has collected when someone fills up the form in the home page along with its predictions. The Expert can Input any corrections if he decides and add all the new data points to the primary development dataset page to add new rule(s) if required.

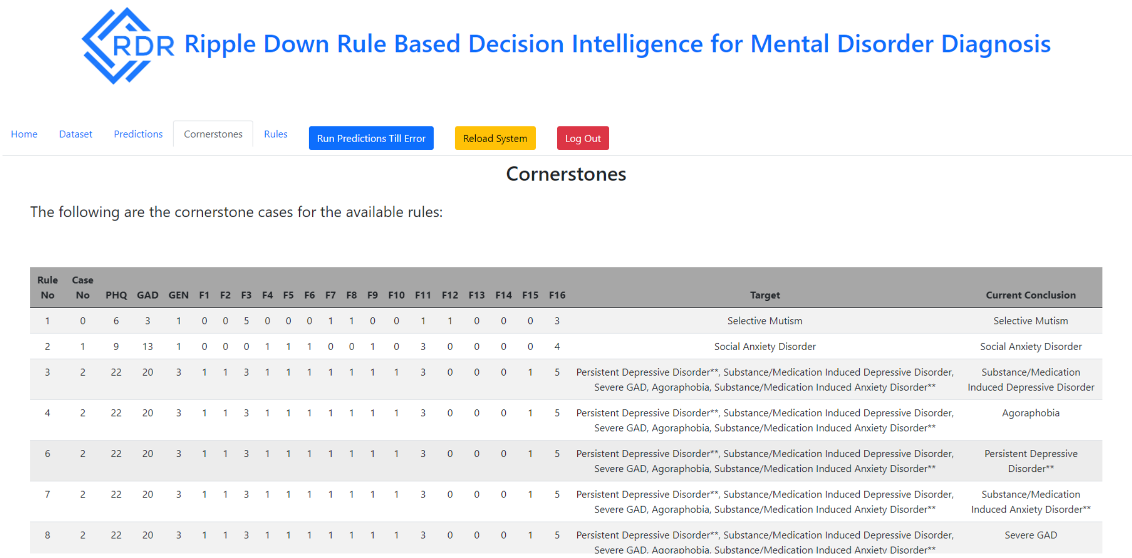


#	PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Conclusion	Rules Evaluated	Rules Fired	Corrections(if any)
0	4	2	0	0	0	5	0	0	0	0	0	0	0	3	0	0	0	0	0	No Disorder	1->2->3->4->6->7->8->9->10->11->13->14->15->16->17->18->19->20		Leave blank for correct conclusion

Feature Abbr.	Feature Definition	Values 0	1	2	3	4	5
GEN	If you checked off any problems, how difficult have these problems made it for you at work, home, or with other people?	Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult	-	-
F1	Do you happen to experience temper outbursts in response to relatively milder situations multiple times through the week?	No	Yes	-	-	-	-
F2	Would you say that you are persistently irritable or angry most of the day, nearly every day?	No	Yes	-	-	-	-
F3	Do you have excessive and persistent worries about specific situations?	Social situation	Losing a major attachment figure	Open Space	Closed space	Any	None
F4	Does the specific situation always provoke anxiety?	No	Yes	-	-	-	-

Figure 4.9: Prediction page of the System

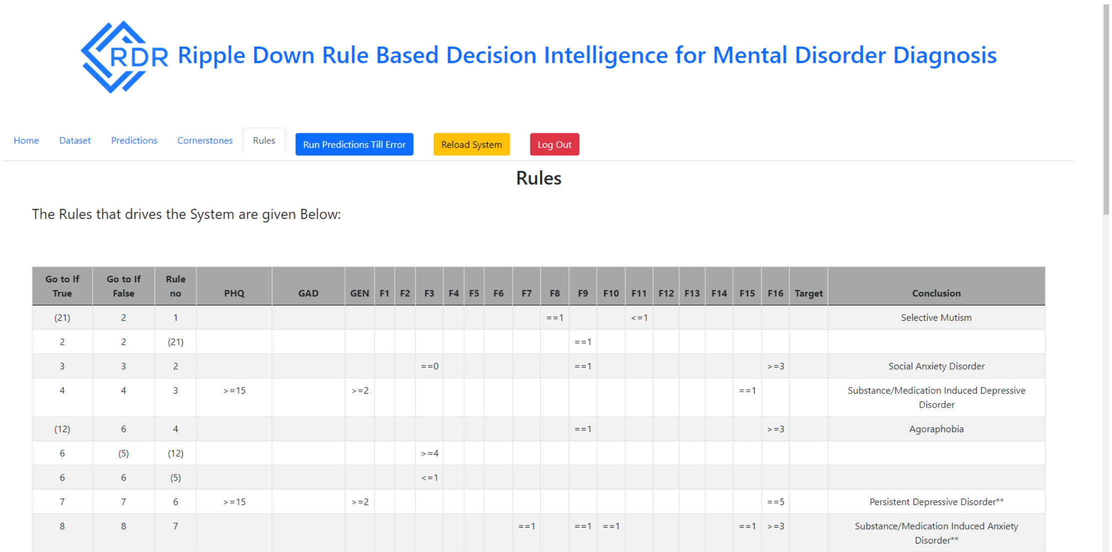
Cornerstone Page: Figure 4.10 presents the Cornerstone cases page of the system. This page requires login. It shows the cornerstone cases for the added rules. The case number from the dataset and the rule no for which this case has been used as a cornerstone case is shown in this page.



Rule No	Case No	PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Target	Current Conclusion
1	0	6	3	1	0	0	5	0	0	0	1	1	0	0	1	1	0	0	0	3	Selective Mutism	Selective Mutism
2	1	9	13	1	0	0	0	1	1	1	0	0	1	0	3	0	0	0	0	4	Social Anxiety Disorder	Social Anxiety Disorder
3	2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**	Substance/Medication Induced Depressive Disorder
4	2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**	Agoraphobia
6	2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**	Persistent Depressive Disorder**
7	2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**	Substance/Medication Induced Anxiety Disorder**
8	2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**	Severe GAD

Figure 4.10: Cornerstone page of the System

Rules Page: Figure 4.11 shows the Rules page of the system. This page requires login. It shows existing rules of the MCRDR system. This page also shows the order they will be evaluated by showing which rule to evaluate in case a particular rule is true or false.



Go to If True	Go to If False	Rule no	PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Target	Conclusion
(21)	2	1											==1			<=1							Selective Mutism
2	2	(21)												==1									
3	3	2						==0						==1							>=3		Social Anxiety Disorder
4	4	3	>=15		>=2															==1			Substance/Medication Induced Depressive Disorder
(12)	6	4												==1							>=3		Agoraphobia
6	(5)	(12)						>=4															
6	6	(5)						<=1															
7	7	6	>=15		>=2																==5		Persistent Depressive Disorder**
8	8	7										==1		==1	==1					==1	>=3		Substance/Medication Induced Anxiety Disorder**

Figure 4.11: Rules page of the System

Buttons: The system also has two action buttons that are accessible by any logged in expert . They are "Run Predictions Till Error" button and "Reload System" button. "Run Predictions Till Error" button runs a prediction algorithm for each case of the development dataset till the end of the dataset or till an error in predictions occurs. On the other hand "Reload System" button reloads all the datasets and resets any variables and re- initializes the knowledge base(without deleting existing rules). Figure 4.12 depicts the buttons used in the system.

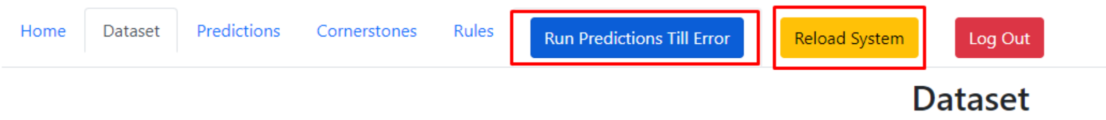


Figure 4.12: Buttons available for expert in the System

Architecture: The MCRDR System uses Python as its development language. The core class of this system is the 'KnowledgeBase' class that supports the prediction for multiple classes on a single case. The Class has the following main methods accessible to its user. Which are:

- **KnowledgeBase.get_kb(features=[](optional)) :** The class uses a Singleton architecture so only one object is used at any number of operations. This method provides that object. The Features parameter is required to be passed only for the first time as a list of features of the dataset to initialize the class object.
- **KnowledgeBase.add_rule(rule_object: Rule) :** This method adds any parent or stopping rules. This method takes a Django Database Custom Model Rule Object . The KnowledgeBase uses a combination of Regular List and Linked-List structure to store its Rules. This method implements that structure.
- **KnowledgeBase.eval_case(case: list, all_rules=None):** this method evaluates a single case based on the stored rules or any rule object passed as an optional parameter. This method takes the help of a helper function to check if a rule satisfies any case. And returns a tuple of conclusions,rules evaluated and rules fired which all are python lists except conclusion which is boolean if no rules are fired.

The web interface uses the Django python framework and also uses pandas and numpy libraries to organize datasets. It also uses the standard json python library to store and maintain high level data structures to the database.

Encoding categorical data to numerical values: The system automatically encodes categorical values from the input form in the home page to generate the PHQ and GAD Scores and to generate numeric values for training and testing. To do this the answers of Question 1-9 from section A is given a weight of 0:Not at all to 3:Nearly Everyday (options that are given). Then weights corresponding to the selected answers are summed to find the PHQ scores. Similarly from the answers of Question 10-16 GAD score is generated. The GEN value refers to the Last question and is weighted 0-3 in the given options.In section B however, the values are simply encoded 1: Yes or 0:No. This concludes the encoding part.
Usages: Any user can use the home page to answer the questions and get an evaluation/prediction. Any System Expert will need to login to access the development facilities of the system. At the beginning of the development the expert is greeted with the dataset page. That holds the primary dataset of this system.

F15	Do you have any history of substance usage or medication that may induce symptoms of mental disorder?	No	Yes	-	-	-	-
F16	How long has it been since you experienced at least one of the symptoms you answered yes to?	No Symptom	<2W	<1M	<6M	<2Y	>2Y

Cases:

#	PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Target
0	6	3	1	0	0	5	0	0	0	1	1	0	0	1	1	0	0	0	3	Selective Mutism
1	9	13	1	0	0	0	1	1	1	0	0	1	0	3	0	0	0	0	4	Social Anxiety Disorder
2	22	20	3	1	1	3	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**
3	16	13	2	0	0	1	1	0	1	1	1	1	1	3	0	1	0	0	4	Mod Severe MDD, Moderate GAD, Separation Anxiety Disorder**
4	16	13	1	1	0	3	1	0	1	1	1	1	1	3	1	0	0	0	4	Agoraphobia
5	24	21	3	0	1	3	1	1	1	1	1	1	1	3	1	1	0	0	5	Persistent Depressive Disorder**, Severe GAD, Agoraphobia
6	2	0	1	0	0	3	0	0	0	0	0	0	0	3	1	0	0	0	0	No Disorder
7	19	19	3	1	1	2	1	1	0	1	0	1	1	3	1	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Agoraphobia, Substance/Medication Induced Anxiety Disorder**
8	9	7	1	0	0	3	0	1	1	1	1	1	0	3	1	0	0	0	5	Agoraphobia
9	7	6	1	1	0	4	0	0	1	1	1	1	0	3	1	0	0	0	4	Specific Phobia
10	10	6	1	0	1	2	0	1	1	1	0	0	0	3	0	0	0	0	4	No Disorder
11	13	18	1	0	1	0	1	1	1	1	1	1	1	3	0	0	1	0	4	Social Anxiety Disorder, Anxiety Disorder due to Another Medical Condition**
12	11	19	3	1	0	0	1	1	0	1	1	1	0	3	1	0	0	0	5	Moderate MDD, Severe GAD, Social Anxiety Disorder
13	10	7	1	0	0	0	1	0	1	0	1	1	0	3	1	0	0	0	4	Social Anxiety Disorder
14	12	12	1	0	1	1	1	0	1	1	1	1	1	3	0	0	1	1	4	Separation Anxiety Disorder**, Anxiety Disorder due to Another Medical Condition**, Substance/Medication Induced Anxiety Disorder**
15	0	0	0	1	1	2	1	1	1	0	0	0	0	3	2	0	0	1	0	No Disorder
16	27	21	3	1	1	0	1	1	1	1	1	1	1	3	0	0	0	1	5	Persistent Depressive Disorder**, Substance/Medication Induced Depressive Disorder, Severe GAD, Social Anxiety Disorder,

Figure 4.13: Cases to generate rules by expert in the System

The Expert can click on any row of the dataset and an interface to add rules will appear to add a rule for that case(as cornerstone case). Figure 4.13 shows the cases by clicking on which expert can add rule. This interface will show the relevant case that was selected to add rules referring to the column name of that case (for example: PHQ, F1 etc.) The rules can be added as any relation to that column in the relevant column field. Figure 4.14 presents the rule addition interface of the system. The rules can be written as a comparison operator (for example: ==0 or >=4) or as a range (for example 14<= PHQ <=19). Notice in case of writing a range the column name is also written inside the condition unlike in case of comparison operator.

F14	Do you have any medical conditions present that may induce symptoms of mental disorder?	No	Yes	-	-	-	-
F15	Do you have any history of substance usage or medication that may induce symptoms of mental disorder?	No	Yes	-	-	-	-
F16	How long has it been since you experienced at least one of the symptoms you answered yes to?	No Symptom	<2W	<1M	<6M	<2Y	>2Y

Rule Check:

#	PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Target	Conclusion
current case	2	0	1	0	0	3	0	0	0	0	0	0	0	3	1	0	0	0	0	No Disorder	No Disorder
cornerstone	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Run Case

Final Conclusion:

Add a new rule for this case:

PHQ	GAD	GEN	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	Target
conditic	conditio	conditic	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	conditi	condition

Conclusion

New Conclusion or N/A for Stopping Rule

Close Add Rule

Figure 4.14: Rule addition page of the System

After writing necessary conditions and conclusions for that rule, a rule can be added using the add rule button. Clicking this button will check the rule conditions with existing cornerstone cases and if the rule does not fire for any existing cases then the rule will be added to the database of the system. In case a cornerstone does fire for the newly added rule the expert is given the choices to update the conclusion of the matching cornerstone, add the new rule as a separate rule, or edit some of the conditions. Figure 4.15 shows the interface that appears a cornerstone case matches with the rule.

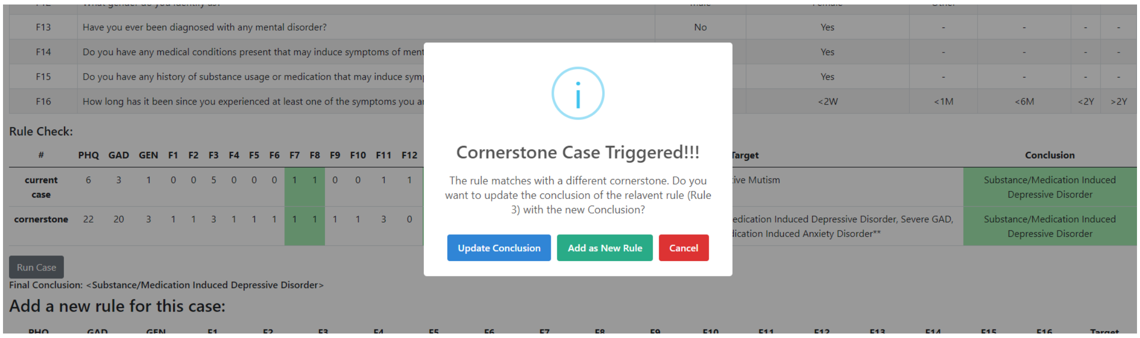


Figure 4.15: Cornerstone case triggered pop-up message of the System

After adding a few rules the expert can click the 'Run Predictions Till Error' button. This will run all the cases through the rules and give conclusions according to the added cases. If no rule is fired the system gives a default conclusion of 'No disorder'. The predictions stop when the conclusions column of the page does not match the target column.

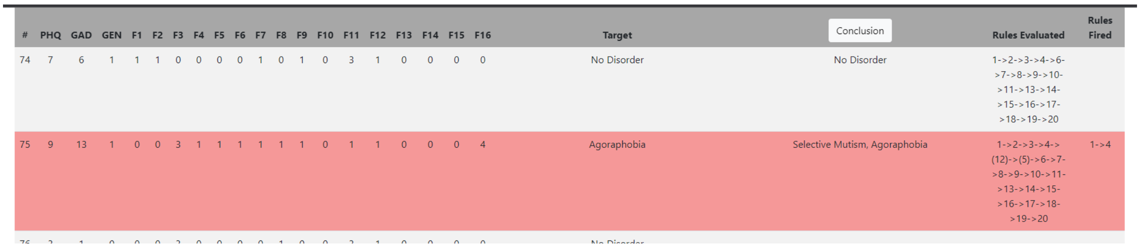


Figure 4.16: Unmatched target and Conclusion for the case

Figure 4.16 shows that there are unmatched conclusion for a case by highlighting the case. This can happen when any of the multiple target classes is not present in the conclusion or if any wrong conclusion is present. An expert can write a new rule taking that case as a cornerstone case which can be either a stopping rule for a wrong conclusion or an additional parent rule for missing conclusion.

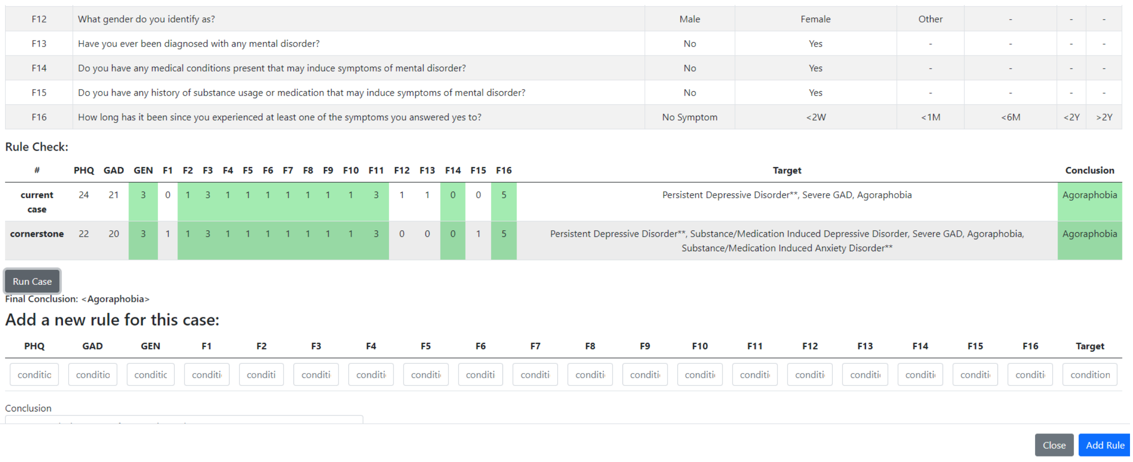


Figure 4.17: Stopping rule addition

Figure 4.17 shows the interface to add stopping rule. In case of adding any stopping rule the expert has to use the same interface as before and double click on that

case. After that the expert must click on the ‘run case’ button to run each rule fired and the conclusion it provides. Upon getting the wrong conclusion for whose rule the new stopping rule needs to be added the expert may add any conditions and the special keyword ‘N/A’ to add any stopping rule for that parent Rule that gave the wrong conclusion in the first place. Clicking the ‘add rule’ button will add the stopping rule to the parent rule. As the system as a whole gives multiple conclusions to any case, the run case cycles through all the rules and conclusions so experts can add stopping rules for any particular conclusion they require. After adding enough rules the Dataset page will not show any more errors and the expert can move on to the testing phase.
Figure 4.18 depicts how a set of test cases can be evaluated by uploading a csv file. The testing can be performed from the run prediction page. Normally this page will have the available predictions given by the system when a user uses the form at ‘Home page’. But if no such predictions are available, a File input button will appear on the predictions page.

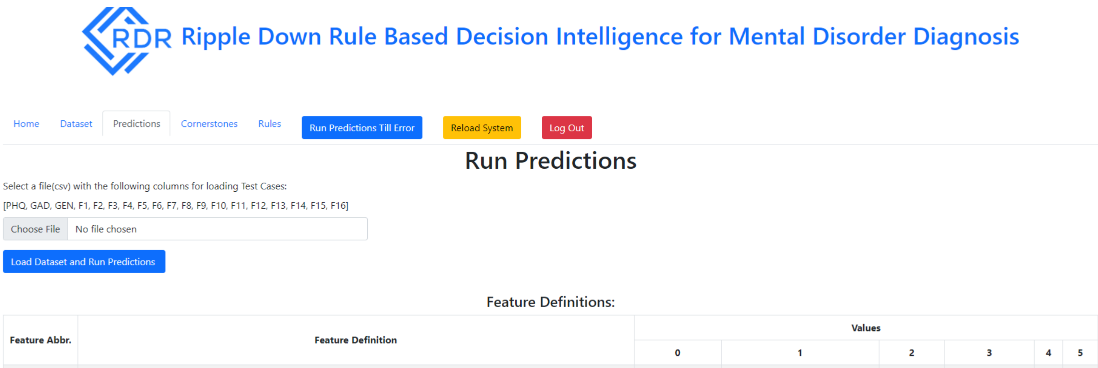


Figure 4.18: Running a set of cases to get conclusion by selecting csv file

An expert can select test data points formatted on a csv file from their system using this feature. And the system will provide predictions. The expert then can manually check if the predictions are correct. If they are not correct then the expert can input the correct conclusions and add those cases to the development dataset and from the dataset page they can add additional rules for those cases with wrong conclusions.

Chapter 5

Conclusion

In this paper, we present Ripple Down Rules (RDR) - to build a knowledge base system, targeted to perform differential diagnosis of mental diseases. With a goal to overcome the shortcomings of supervised learning systems bound by the expense of reliable data on mental health, we propose a RDR system, having easy to explain outputs, and the validation of a domain expert crafting the rules. RDR also allows for not only generalizing the mass, but also deal with corner-cases. The overall case by case approach of RDR, supports its long-term use case also. We take the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [12], a widely referenced benchmark for mental disease diagnosis, as our primary base. The proposed knowledge base is not only one of the core solutions to diagnosis of different mental disorders at the same time using such an approach but also an important ground for further research into the related fields. To conclude, we are driven by the necessity of development in mental healthcare with the growing number of population with some variety of both diagnosed and undiagnosed mental disorders. Our vision is that this research will make the diagnosis process much more simpler, inexpensive and generally more accessible.

5.1 Future Works

There are opportunities for enhancement, notwithstanding the system's positive output. Two categories of mental disorders are now being diagnosed using a set of 33 questions. We can extend the system to classify more disorders. In such case, there will be more questions. Secondly, adding new rules requires checking all the cornerstone cases against the new rule; by using association rule mining, we can spare the experts from this extra work. An expert can see how many potential cornerstone cases trigger this new rule with the use of association rule mining. The rules will be more precise as a result of this. Thirdly, we also want to venture the ability to utilize supervised machine learning solutions simultaneously with the RDR system. RDR will be used first to eliminate some disorders like differential diagnosis process and a machine learning algorithm could be applied to diagnose the case among the disorders that have not been eliminated for the given case. If sufficient data is gathered for all disorders, this approach may be viable.

Bibliography

[1] P. Tyrer and J. Alexander, "Classification of personality disorder," *The British journal of psychiatry : the journal of mental science*, vol. 135, pp. 163–7, Sep. 1979. DOI: 10.1192/bjp.135.2.163.

[2] D. Kahneman, S. P. Slovic, P. Slovic, and A. Tversky, *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[3] B. Drake and G. Beydoun, "Predicate logic-based incremental knowledge acquisition," in *Proceedings of the sixth Pacific International Knowledge Acquisition Workshop*, 2000, pp. 71–88.

[4] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: Validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[5] K. Kroenke and R. L. Spitzer, *The phq-9: A new depression diagnostic and severity measure*, 2002.

[6] J. Bekmann and A. Hoffmann, "Heureaka–a new approach for adapting gas to the problem domain," in *Pacific Rim International Conference on Artificial Intelligence*, Springer, 2004, pp. 361–372.

[7] J. Bekmann and A. Hoffmann, "Improved knowledge acquisition for high-performance heuristic search," in *IJCAI*, 2005, pp. 41–46.

[8] R. L. Spitzer, K. Kroenke, J. B. Williams, and B. Löwe, "A brief measure for assessing generalized anxiety disorder: The gad-7," *Archives of internal medicine*, vol. 166, no. 10, pp. 1092–1097, 2006.

[9] Z. Xu, "Multiple-attribute group decision making with different formats of preference information on attributes," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 6, pp. 1500–1511, 2007.

[10] A. Misra, A. Sowmya, and P. Compton, "Incremental system engineering using process networks," in *Pacific Rim Knowledge Acquisition Workshop*, Springer, 2010, pp. 150–164.

[11] I. Research, *Global technology outlook: Ibm*, 2012.

[12] A. P. Association, *Diagnostic and statistical manual of mental disorders: DSM-5 (Vol. 5)*. Washington DC: American Psychiatric Association, 2013.

[13] S. Kung, R. D. Alarcon, M. D. Williams, K. A. Poppe, M. J. Moore, and M. A. Frye, "Comparing the beck depression inventory-ii (bdi-ii) and patient health questionnaire (phq-9) depression measures in an integrated mood disorders practice," *Journal of affective disorders*, vol. 145, no. 3, pp. 341–343, 2013.

[14] M. Tartakovsky, *What prevents people from seeking mental health treatment*, 2013.

[15] U. Lueken, B. Straube, Y. Yang, *et al.*, "Separating depressive comorbidity from panic disorder: A combined functional magnetic resonance imaging and machine learning approach," *Journal of affective disorders*, vol. 184, pp. 182–192, 2015.

[16] M. D. Sacchet, G. Prasad, L. C. Foland-Ross, P. M. Thompson, and I. H. Gotlib, "Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory," *Frontiers in psychiatry*, vol. 6, p. 21, 2015.

[17] M. Sribala and M. Sribala, "An approach of artificial neural networks for prediction of generalized anxiety disorder," *International Journal of Research in Computer Applications and Robotics*, vol. 3, no. 3, pp. 118–124, 2015.

[18] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[19] W. Husain, L. K. Xin, N. Jothi, *et al.*, "Predicting generalized anxiety disorder among women using random forest approach," in *2016 3rd international conference on Computer and information sciences (ICCOINS)*, IEEE, 2016, pp. 37–42.

[20] G. Li, J. Wang, Y. Zheng, and M. J. Franklin, "Crowdsourced data management: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2296–2319, 2016.

[21] F. Plummer, L. Manea, D. Trepel, and D. McMillan, "Screening for anxiety disorders with the gad-7 and gad-2: A systematic review and diagnostic metaanalysis," *General hospital psychiatry*, vol. 39, pp. 24–31, 2016.

[22] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[23] B. Sundermann, J. Bode, U. Lueken, *et al.*, "Support vector machine analysis of functional magnetic resonance imaging of interoception does not reliably predict individual outcomes of cognitive behavioral therapy in panic disorder with agoraphobia," *Frontiers in psychiatry*, vol. 8, p. 99, 2017.

[24] J. Dastin, *Amazon scraps secret ai recruiting tool that showed bias against women*, Oct. 2018. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

[25] J. Lu, Z. Yan, J. Han, and G. Zhang, "Data-driven decision-making (d 3 m): Framework, methodology, and directions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 4, pp. 286–296, 2019.

[26] M. Arif, A. Basri, G. Melibari, *et al.*, "Classification of anxiety disorders using machine learning methods: A literature review," *Insights Biomed Res*, vol. 4, no. 1, pp. 95–110, 2020.

[27] E. Herrera-Viedma, I. Palomares, C.-C. Li, *et al.*, "Revisiting fuzzy and linguistic decision making: Scenarios and challenges for making wiser decisions in a better way," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 191–208, 2020.

[28] H. Lakkaraju and O. Bastani, "" how do i fool you?" manipulating user trust via misleading black box explanations," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 79–85.

[29] P. Compton and B. H. Kang, *Ripple-Down Rules: The Alternative to Machine Learning*. Chapman and Hall/CRC, 2021.

[30] J. De Man, P. Absetz, T. Sathish, *et al.*, "Are the phq-9 and gad-7 suitable for use in india? a psychometric analysis," *Frontiers in psychology*, vol. 12, p. 676 398, 2021.

[31] Majella, *Xai problems part 3 of 3: Technical problems with common explainer algorithms*, Feb. 2021. [Online]. Available: https://elula.ai/technical-problems-with-common-explainer-algorithms/.

[32] *What is decision intelligence?: Defining decision intelligence (di)*, Aug. 2022. [Online]. Available: https://peak.ai/what-is-decision-intelligence/.

[33] OMG. "Production rule representation (prr)." (), [Online]. Available: https://www.omg.org/spec/PRR/1.0. (accessed: 24.05.2022).

[34] M. Stewart. "The limitations of machine learning." (), [Online]. Available: https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6. (accessed: 24.05.2022).

Appendix A

Questionnaire for data collection

A.1 Part A:

The following 17 questions were collected from the PHQ-9 and GAD-7 questionnaires. The last question, is found in both PHQ-9 and GAD-7 sets.

1. Little interest or pleasure in doing things
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
2. Feeling down, depressed, or hopeless
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
3. Trouble falling or staying asleep, or sleeping too much
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
4. Feeling tired or having little energy
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
5. Poor appetite or overeating
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
7. Difficulty concentrating on things, such as reading the newspaper or watching television
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day
8. Moving or speaking so slowly that other people could have noticed Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual
   - Not at all;
   - Several days;
   - More than half the days;
   - Nearly every day

9. Thoughts that you would be better off dead, or of hurting yourself
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

10. Feeling nervous, anxious, or on edge
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

11. Not being able to stop or control worrying
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

12. Worrying too much about different things
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

13. Trouble relaxing
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

14. Being so restless that it is hard to sit still
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

15. Becoming easily annoyed or irritable
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

16. Feeling afraid, as if something awful might happen
    - Not at all;
    - Several days;
    - More than half the days;
    - Nearly every day

17. If you checked off any problems, how difficult have these problems made it for you at work, home, or with other people?
    - Not difficult at all;
    - Somewhat difficult;
    - Very difficult;
    - Extremely difficult

A.2 Part B:

The following 16 questions were given by Dr. Afroz to collect additional information for anxiety related diseases, demographic information and patient history:

1. Do you happen to experience temper outbursts in response to milder situations multiple times through the week?
    - Yes;
    - No

2. Would you say that you are persistently irritable or angry most of the day, nearly every day?
    - Yes;
    - No

3. Do you have excessive and persistent worries about specific situations?
   - Social situation;
   - Losing a major attachment figure;
   - Open space;
   - Closed space;
   - Any;
   - None of the above

4. Does the specific situation always provoke anxiety?
   - Yes;
   - No

5. Is the worry out of proportion to the actual threat?
   - Yes;
   - No

6. Do you avoid the specific situation of worry or fear?
   - Yes;
   - No

7. Do you have physical symptoms for worry or fear?
   - Restlessness;
   - Irritability;
   - Muscle tension;
   - Sleep disturbance;
   - Tremor;
   - Palpitation;
   - None;

8. Do you have trouble speaking or unable to speak in specific situations in-spite of speaking in other situations?
   - Yes;
   - No

9. Is the worry and related symptoms affecting your day to day activities?
   - Yes;
   - No

10. Do you have excessive anxiety and worry (apprehensive expectation) occurring most of the day?
    - Yes;
    - No

11. Which one of the following age groups applies to you?
    - 0 to 5;
    - 6 to 11;
    - 12 to 17;
    - 18+

12. What gender do you identify as?
    - Male;
    - Female;
    - Other

13. Have you ever been diagnosed with any mental disorder?
    - Free Response

14. Do you have any medical conditions present that may induce symptoms of mental disorder?
    - Free Response

15. Do you have any history of substance usage or medication that may induce symptoms of mental disorder?
    - Free Response

16. How long has it been since you experienced at least one of the symptoms you answered yes to?
    - I don't have any of the symptoms;
    - Less than 2 weeks;
    - More than 2 weeks but less than 1 month;
    - More than 1 month but less than 6 months;
    - More than 6 months but less than 2 years;
    - More than 2 years