Detection and Diagnosis of Psychological Disorders through Decision Rule Set Formation

Author

ANILA UMAR

00000205711

Supervisor

Dr. Usman Qamar

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

SEPTEMBER, 2021

Detection and Diagnosis of Psychological Disorders through Decision Rule Set Formation

Author

ANILA UMAR

00000205711

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Software Engineering

Thesis Supervisor:

Dr. Usman Qamar

Thesis Supervisor's Signature:_____

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

SEPTEMBER, 2021

Declaration

I certify that this research work titled "*Detection and Diagnosis of Psychological Disorders Through Decision Rule Set formation"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

ANILA UMAR

2017-NUST-MS-Software-00000205711

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student
ANILA UMAR
00000205711

Signature of Supervisor
Dr. Usman Qamar

Copyright Statement

Acknowledgements

I am thankful to my Creator Allah The Almighty, who guided me throughout this work at every step and for every new thought which you setup in my mind to improve it. Indeed, I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am exceptionally thankful to my husband who helped and supported me throughout my MS. Without his untiring efforts and support I would never be able to excel in my studies.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor Dr. Usman Qamar for his help throughout my thesis and also for Data Engineering and Web Engineering course which he has taught me. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught.

I would also like to thank Dr. Wasi Haider and Dr. Farhan Hussain my GEC committee for their support and cooperation.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional Husband, my dearest parents and supporting friends whose tremendous support and cooperation led me to this wonderful

accomplishment

Abstract

Mental illness has long term influences on people's relationships, health, education and profession. According to WHO, psychological disorders are continuously growing its burden for the past few years throwing serious impact on people's health, economic and social condition all over the globe. A similar statistics states that in 2011 about 30 million people worldwide were affected by Major Depressive Disorder, 60 Million by Bipolar Affective and 23 Million by Schizophrenia and other Psychosis. Although medical sciences have come with every effective intervention for the diagnosis and treatment of psychological disorders but still people don't seek out the treatment they need. The major reason for this failure is because of the overlapping symptoms of Psychological disorders. It sometimes takes years by the Psychiatrists to accurately diagnose the psychological disorder. The objective of this research paper is to detect and diagnose more accurately the presence of three of the major and most common psychological disorders i.e. Major Depressive Disorder, Bi-Polar Affective Disorder and Schizophrenia through Decision Rule Set Formation (RST).

Key Words: *Decision Rule Set, Psychological Disorders, Data Mining Technique, Detection, Diagnosis..*

Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

1.1 Overview

Mental illness is a health condition that involves a change in mood, thinking or behavior (or a combination of these). Mental illness is associated with grief and / or problems that affect the social, work or family environment. [1] To get an accurate diagnosis of mental illness, it is important to diagnose it properly. It is even harder for psychiatrists to diagnose and diagnose certain mental disorders. Therefore, it is very important to get an accurate diagnosis of mental disorders before starting treatment. Through Decision Rule Set formation effective decision making is possible by the practitioners to diagnose and analyze well the mental health of a patient.

1.2 Background

Statistics show that up to 50% of adults experience mental illness at some point in their lives. Despite the high incidence of psychiatric disorders, only about 20% of people with a mental illness receive professional help [1]. Mental health problems in the United States affect 25% of adults, 18% of teens and 13% of children. These disorders have a greater impact on the economy than cancer, cardiovascular disease, diabetes, and respiratory diseases, but communities and governments spend far less money on mental disorders than in other diseases. [2]

Statistics depicts that mental illness is highly common in Asia Pacific, South Asia, Sub-Saharan and Africa. This is due to lack of awareness about mental health in under developed countries. [3]

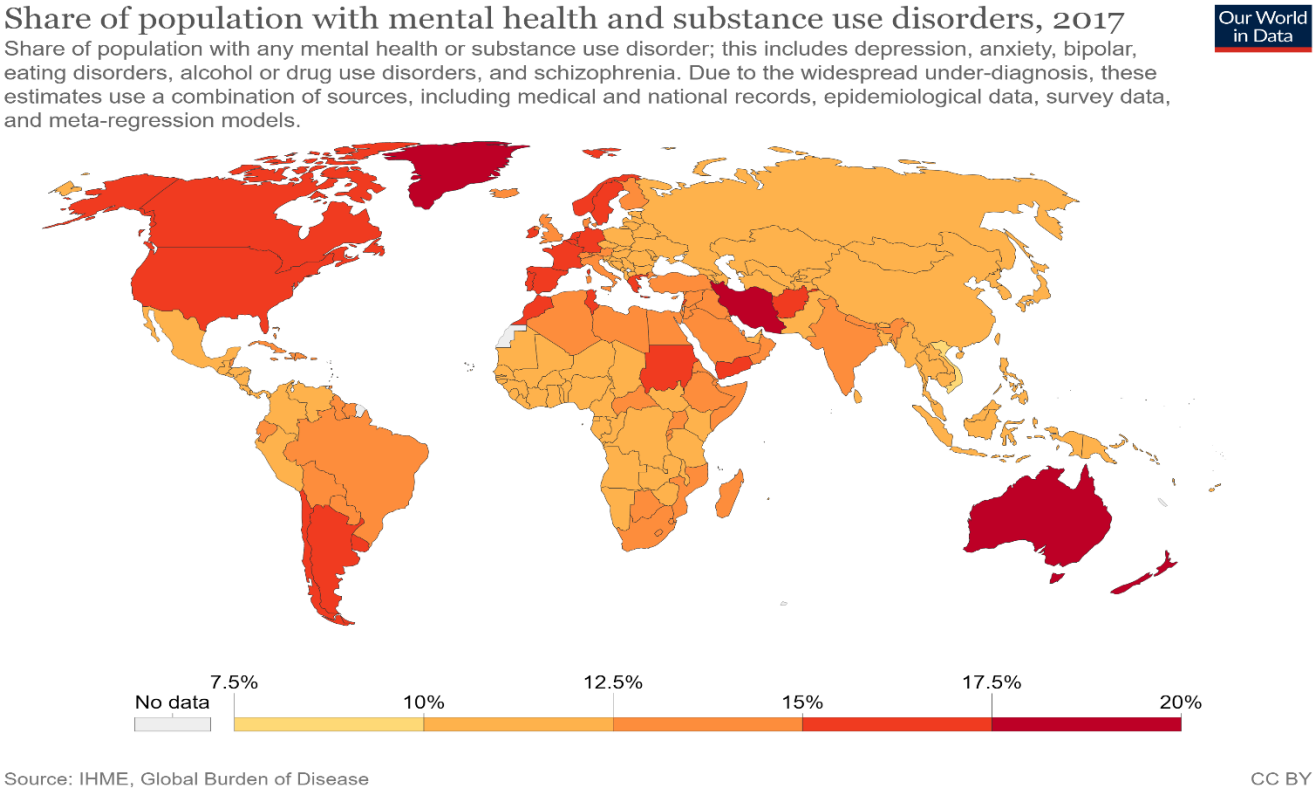

Fig 1.2: Share of population with Mental Health [3]

In Pakistan, Mental health is the most neglected field where 10- 16% of the population, more than 14 million, suffers from mild to moderate psychiatric illness, majority of which are women. Pakistan has only one psychiatrist for every 10,000 persons suffering from any of the mental disorders, while one child psychiatrist for four million children, who are estimated to be suffering from mental health issues. Only four major psychiatric hospitals exist for the population of 180 million and it is one the major factors behind increase in number of patients with mental disorders.[2]

In Pakistan, majority of the psychiatric patients go to traditional faith healers and religious healers who believe that mental illness is caused by supernatural forces such as spirit possession or testing by God**.** "All this is due to acute shortage of mental health professionals and relatively low levels of awareness about mental disorders". There is also no political will and no proper mental health policy in Pakistan. All this adversely affects the integration of care delivered by government health care professionals for patients with mental illnesses**.** Major

mental disorders in Pakistan are depression (6%), schizophrenia **(**1.5%) and epilepsy (1-2%). [2]

Current approaches to the assessment and monitoring of psychiatric conditions rely primarily on internet reports from affected individuals or their caregivers. These reports are often subjective and include patients' retrospective recall biases, cognitive limitations and social stigma. There is an urgency to objectively diagnose, monitor over time, and provide evidence-based interventions for individuals with mental illnesses, particularly those who are unable to access traditional psychological or psychiatric services due to geographical, financial, or practical barriers.

1.3 Motivation

People from diverse cultural backgrounds may express mental health conditions differently. "In Pakistan, majority of the psychiatric patients go to traditional faith healers and religious healers who believe that mental illness is caused by supernatural forces such as spirit possession or testing by God". [2] Though Mental illness can be treated but people often get the wrong treatment because of misdiagnosis by the psychiatrists. "Certain psychological disorders have so overlapping symptoms that its almost impossible for the most trained psychiatrists to make the correct diagnosis". A mental health diagnosis is the first step on the path to treatment. If that first foray is a misstep and someone receives an incorrect diagnosis, a few consequences can be devastating. A misdiagnosis will result in the patient becoming confused and potentially distraught when the course of treatment recommended isn't working. If a mental health condition gets misdiagnosed or never diagnosed at all, the patient is likely to keep getting worse condition is misdiagnosed or not found at all, the patient may continue to get worse.

1.4 Problem Statement

"An approach for the formation of Decision Rule set for the Detection and Diagnosis of Psychological Disorders".

Through decision Rule set formation our aim is to generate rules to detect and diagnose more accurately the presence of three of the major and most common psychological disorders i.e. Major Depressive Disorder, Bi-Polar Affective Disorder and Schizophrenia

and to facilitate doctors and patients by providing them with computer aided diagnosis systems.

1.5 Objective

The prime objective of this thesis is to:

- Formulate a dataset that can accurately target the correct symptoms of the psychological patients.
- Generate rules that can help in detection and diagnosis of the psychological disorders.
- Detect and Diagnose Psychological disorders to assist the psychiatrists in medical domain.

1.6 Scope

Various data mining techniques are in practice for the diagnosis of psychological disorders. But it is for successful treatment of a mental illness to get the disease correctly diagnosed. Previous approaches in this field mostly focuses on ML techniques and often use text-based data, such as datasets from social media adopt mainstream techniques such as DT, ANN, KNN and SVM as methods for machine learning experiments. Moreover Datasets of only HDD is implemented in most cases. So it is very much needed to have a dataset that is based on real time data of patients and train it on a comparatively better technique used so far i.e. Decision Rules. Through this accuracy can also be improved further.

1.7 Research Contributions

The contribution of our work includes:

- Formulation of dataset
- Generation of rules
- Detection and diagnosis of three major psychological disorders
- Improve the accuracy of techniques used so far

1.8 National Needs

The ability to analyze and process data is growing exponentially in today's world. The pace of change requires organizations to be able to respond quickly to changes in customer needs and environmental conditions. The project is therefore trying to reduce the current gap between engineering research, modern technology and industry in Pakistan.

Moreover, in Pakistan, mental health is a neglected field where 10-16% of the population, over 14 million, suffer from moderate to moderate mental illness, most of them women. [2] This work will help psychiatrists diagnose and diagnose complications that are difficult to detect by routine clinical examinations.

1.9 Applications

Some of the useful applications of this work includes:

- Provide ease for psychiatrists in diagnosis.
- Consumes less time in treatment.
- Help the psychiatrists to drive the treatment in right direction**.**

1.10 Thesis Structure

The overall thesis structure is as follows:

- **Chapter 2: Literature Review.** Chapter 2 discusses data mining approaches that are used so far in the detection of psychological disorders. Also the advantages and limitations of their findings.
- **Chapter 3: Proposed Methodology.** In this chapter, we will discuss the steps that are involved in the methodology applied.
- **Chapter 4: Results and Analysis.** In this chapter, we will discuss the final evaluation of the applied methodology and also analyze the results obtained.
- **Chapter 5: Conclusion and Future Work.** This chapter concludes the whole thesis and some future work is also given.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we will explain the steps which we follow to do the literature review and the related work. In section 2.1, sequence of research is mentioned. In section 2.2, related work is explained and in last section 2.3, analysis on literature review is done.

2.1 Research Sequence

In this section, we will explain the steps which we follow to do the literature review for our thesis.

2.1.1 Research Questions

Following are the research questions for our thesis.

- What are the Data Mining approaches used so far for the detection of mental disorders?
- What is the source of data that is used by the techniques?
- How many Mental disorders are taken under consideration?
- What is the accuracy of the approach used?
- Which is the best data mining approach used so far in the medical domain?

2.1.2 Review Period

Our review period is from 2010 to 2020, because we want to stick with recent quality research. For that purpose, we reject old studies.

2.1.3 Objective of literature Review

Following are the objectives of our literature review:

- Our main objective is to survey the literature of the research questions described above.

- To find gaps in the current literature and to perform critical analysis on the literature.
- To identify Controversy present in the literature.
- Identify future research areas.
- To present our study and research in an organized way.

2.1.4 Keywords

Following are the keywords for our literature review:

- Decision Rule set formation
- Psychological disorders
- Detection and diagnosis
- Data set formation
- Data Mining

2.1.5 Inclusion/Exclusion Criteria

Following are the inclusion and exclusion criteria of our literature review:

- Paper properly related to keywords are selected
- Review period described in section 2.1.2 is taken into consideration
- Famous Databases are targeted i.e. Science Direct , IEEE Xplore etc.
- Authentic and valid researches are included
- Irrelevant researches are excluded

2.1.6 PRISMA 2009 Categorization

Following figure shows the search process in detail.

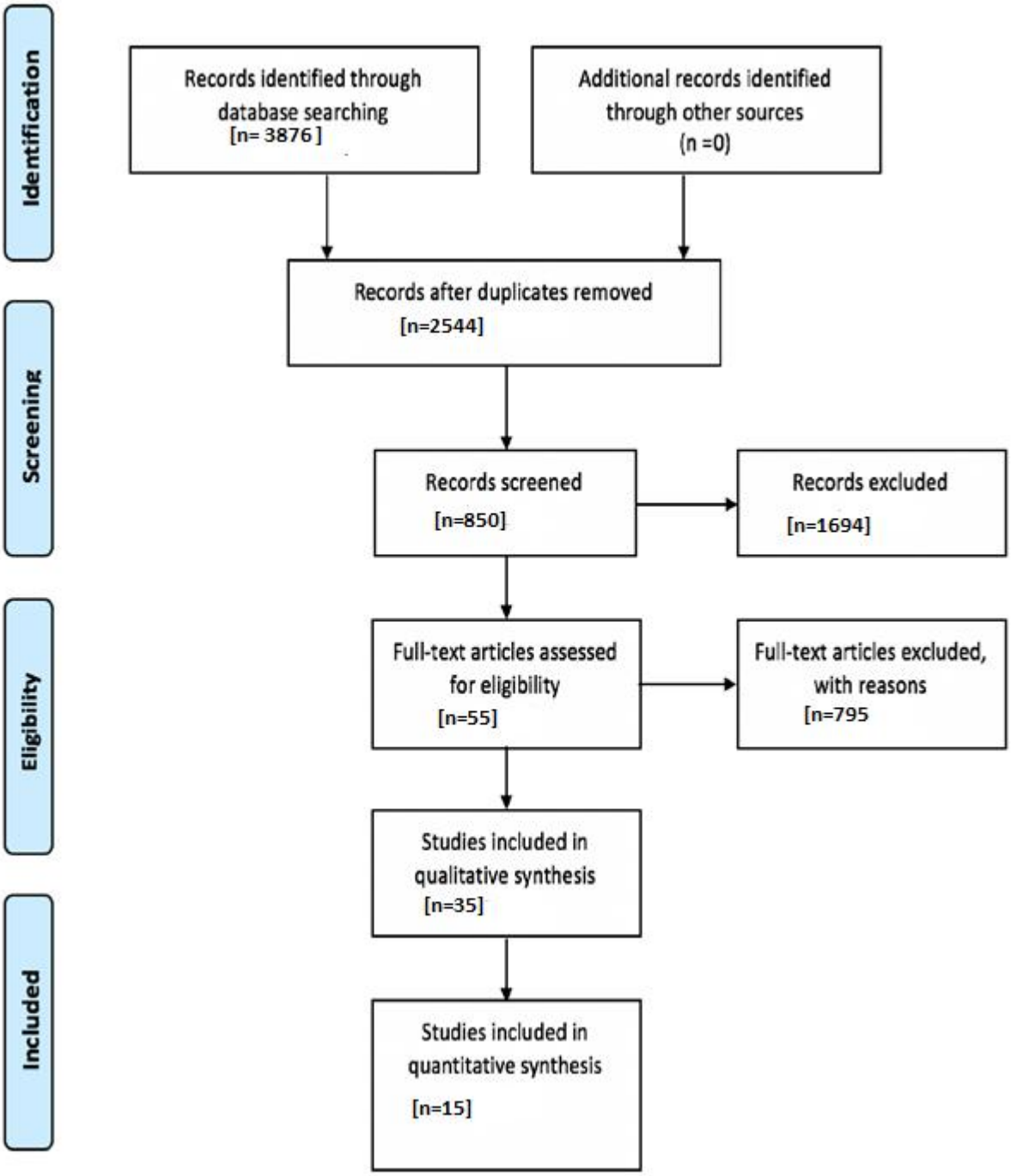

Figure 2.1-1: Prisma 2009 Categorization

2.2 Related Work

2.2.1 Data Mining Techniques in Mental Health

Several previous studies have used learning algorithms to detect symptoms of mental disorders in data sets containing behavioral interventions, in the most common form: Support Vector Machine algorithm, Naive Bayes method and Random Forest technique. In their systematic review of ML strategies in the field of mental health, Shatte et al. [24] identified regression and decision trees (DT) as other common methods used. Along with KNN, these methods comprise typical representative techniques in machine learning experiments.

Choudhury et al. [11] has developed a model with the potential to train people posting on Twitter and to develop an indicator of social stress to identify levels of depression in a sample of the population. The study used SVM classification with RBF kernel to identify depressive conditions. Five-fold verification was used to confirm the operation of the separator, with the results showing an average accuracy of 73% and a high accuracy of 82% [11]. The stress index had a strong correlation with national stress statistics [11]. Importantly, this approach established the need to add a social and external environment to MDD testing.

Tsugawa et al. [12] developed a SVM supervised learning model to use features from online activities to predict the current state of user stress. The features used to predict depression have been removed from users' work history. In this way, 69% accuracy can be achieved by predicting depressed users by the proposed separator [12]. Reliable status (critical rating) for users was made for CES-D and BDI test scales for all participants. Limitations of this study found that longer observation times for data collection could reduce accuracy.

In combining the Random Forest (RF) process with SVM, Fatima et al. [15] have been able to discriminate against oppressive posts and communities on non-oppressive posts and communities in an online social network. Live Journal empowers users to provide pre-defined “mood tags” to the user’s post, which features features to measure levels of frustration among users who create posts and participate in communities. The study used the Random Forest algorithm and the SVM text class classifier to find the highest margin between depressed, depressed and non-depressed classes. In the experiment, RF performed better compared to the conventional SVM method, as the proposed model achieved approximately 90% and 95% accuracy in the separation of depressive posts and depressed communities, respectively [15].

Hassan et al. [14] used the majority vote to divide and further predict the reduction from three different divisions: the SVM division, the Naive Bayes (NB) division, and the Maximum Entropy division (ME). Research has shown how to find the scale of each depression by looking at and expressing emotions as features from the text on various social media platforms [14]. SVM performance accuracy is 91%, 83% and 80%, respectively, for NB and ME dividers. In another study to test data from China's social networking sites Weibo, Peng, Hu and Dang [13] used a model based on multiple SVM types into three categories, feature microblog text, user profile and user behavior [18]. The multi-kernel SVM method had a very low error rate of 16.5% for identifying depressed people [13]. This research has shown that the ensamble method can obtain better predicting performance using more learning algorithms than traditional learning algorithms [13].

Shivangi Jain, Mohit Gangwar et. Al used different classification and performed using J48 (C4.5), Random forest (RF) and Random Tree (RT) approach. The classification with precision, recall, ROC curve and F-measure is taken in as computation parameter. An analysis shows that the Random tree based approach find efficient result while comparing with J48 and Random forest algorithm.[22]

2.2.2 Decision Rules approach in Medical Diagnosis

Most of the research activities in this context are based on hospital data. In their study, Paetz and Brause [4] demonstrated the results of data-driven law-abiding patient-by-data data by using an algorithm that works well for standard patterns, and by measuring the effectiveness of generated rules in terms of frequency and confidence, introducing the best rules. Brossette et at [5] analyzed the problem of identifying interesting patterns in hospital infection control and public health monitoring data using organizational rules. Ohsaki et al. [6] has established a law enforcement support system to obtain interesting rules from data set in chronic hepatitis C tests. Ordonez et al. [7] focuses on finding the rules of integration into a set of real data to predict the absence or presence of colds by introducing a greedy algorithm. Ordonez et al. [8] find the rules of association in medical data to predict heart disease. Their study introduced an advanced algorithm for finding problematic meeting rules and introduced a test section that summarized several of the found rules. Another study by Nahar et al.[9] also aimed to find the cause of heart disease through association rules, and analyzed data available on sick and healthy men and women. Ordonez [7] pointed out that the biggest problem with ARM is the set of medical data with a large size of applied law where most of them were insignificant resulting in slow searches and difficult translations by field experts. Therefore, in his study, search problems were found to find only the rules relating to medical relationships to make the search more efficient and faster. According to the experimental setting, Ordonez [7] used vascular data and found that the rules of the organization of healthy and diseased arteries.

Table 2.2-1: Related Work

Paper	**TECHNIQUE AND TECHNOLOGY**	**Improvements**	**Limitations**
Choudhury et al. [11]	SVM	-Accuracy 72% -10 fold cross validation	-Data from Twitter -Only for HDD
Tsugawa et al. [12]	SVM	- Accuracy: 69%	-Only Depression Patients -Data of Depressed Tweets
Fatima et al. [15]	SVM combined with Random Forest	- Accuracy: 90%	- Data is based on: Pre defined moods - Source: Online Live Journal
Hassan et al. [14]	SVM, NB and Max Entropy	Accuracy: 80%	Data from social media platforms
Weibo, Peng, Hu and Dang [13]	Multiple learning algorithms	Accuracy: 84%	Data not reliable
ShivangiJain, Mohit Gangwar et. Al [22]	J48 (C4.5), Random forest (RF) and Random Tree (RT)	Accuracy: 88%	Applied on small dataset

Y. Ghafoor, Ping Haung et al [20]	Association analysis and Frequent pattern tree	Accuracy: 67%	Only one disorder taken
Mohit Gangwar, R. B. Mishra, et al [19]	ANN and heuristic method	Accuracy: 86%	Applied on small dataset
AhYoung Kim, Eun Hye Jang et. al. [18]	SVM, DT, is used	Accuracy: 74%	Applied on HDD only

2.3 Analysis

After having the results of the literature review following are the major gaps.

- Mostly focuses on ML techniques and often use text-based data, such as datasets from social media.
- Adopt mainstream techniques such as DT, ANN, KNN and SVM as methods for machine learning experiments.
- Datasets of HDD is implemented in most cases
- Accuracy can be improved further by applying more techniques

In our proposed approach described in chapter 3, we have tackled both the above-mentioned gaps.

CHAPTER 3: PROPOSED SOLUTION

In this chapter, we will discuss in detail all the steps that are involved methodology applied for the detection and diagnosis of Psychological disorders through decision rule set formation. Fig shows the flow of process in detail.

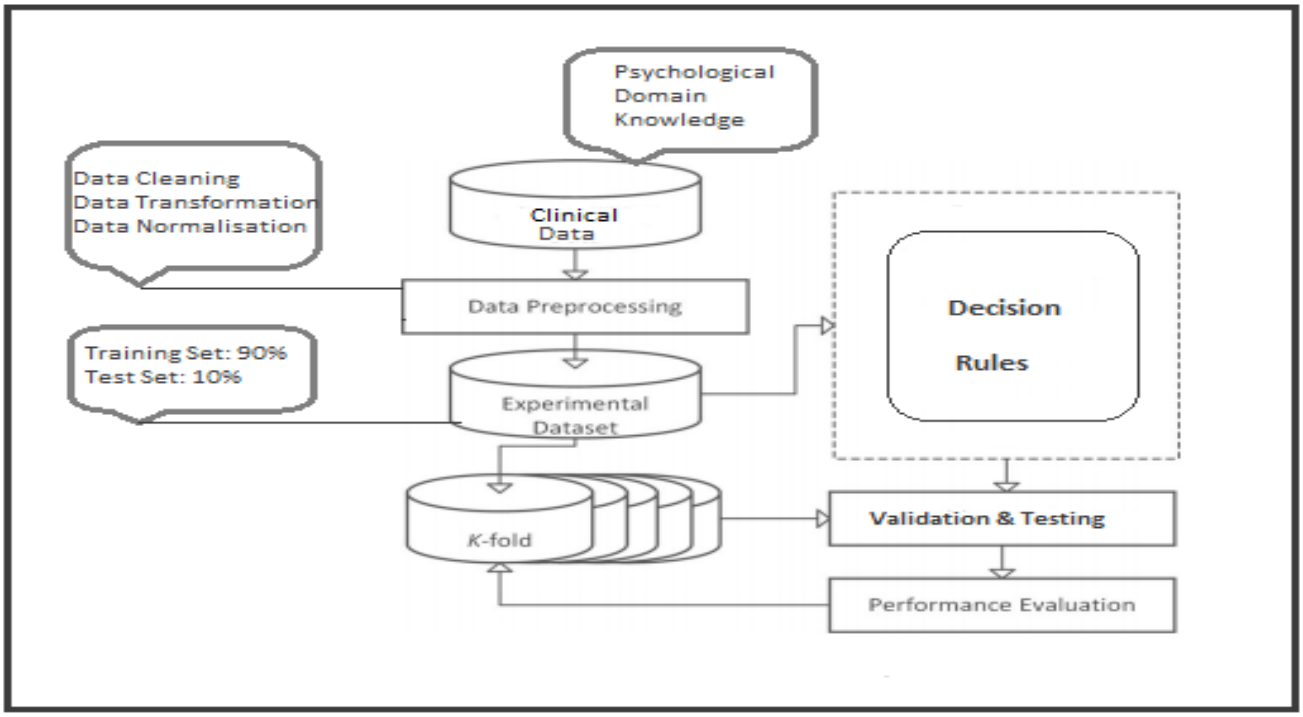

Fig 3.1: Proposed Methodology

3.1 Clinical Data Collection:

Collection of data for the respective research is a complex task in itself because of targeted audience of psychological Patients and the fact that any redundancy in the data may lead to incorrect and misleading symptoms of the patients eventually leading to the drastic results. So our main aim of this research was to collect data based on reality and accurately taken facts and figures from the patients under the supervision of highly qualified and professional psychiatrists.

3.1.1 Source of Data:

The data is collected from **Arms Forces Institute of Mental Health [AFIMH]**, Rwp. Which is one of the renowned institute of mental health working in Pakistan. The hospital is divided into three units i.e. Male ward, Female ward and Intensive care unit (PICU). The data collection is made from all the three units from the records provided for each patient by the psychiatrists. Record of about 100 patients were taken under consideration.

3.1.2 Basis of Data Collection:

To diagnose the disease correctly it is very important to get right factors and symptoms of the targeted patients. Therefore deep knowledge of psychological domain is used to make the data collection process successful. For this purpose the data was collected under the supervision of Head of the Psychological department, AFIMH, Rwp (Brig. Dr. Sohail and Maj. Dr. Shabnam). After careful study and discussion three psychological disorders were taken in consideration for the experimentation purpose i.e. Schizophrenia, Bipolar disease and Hyper Depressive Disorder (HDD).

The three disorders selected are given consideration on the basis of level of difficulty that is being faced by the physicians in the diagnosis of various mental disorders due to overlapping symptoms. Data from each patient is selected separately from the patients record book according to the format specified by the psychiatrists.

3.2 Data Pre processing

Data processing is an important task. It is a data mining method that converts raw data into an understandable, usable and efficient format. Data has a better view. This idea will become clearer and understandable after performing data processing. The following steps are followed by a preliminary consideration of our raw clinical data.

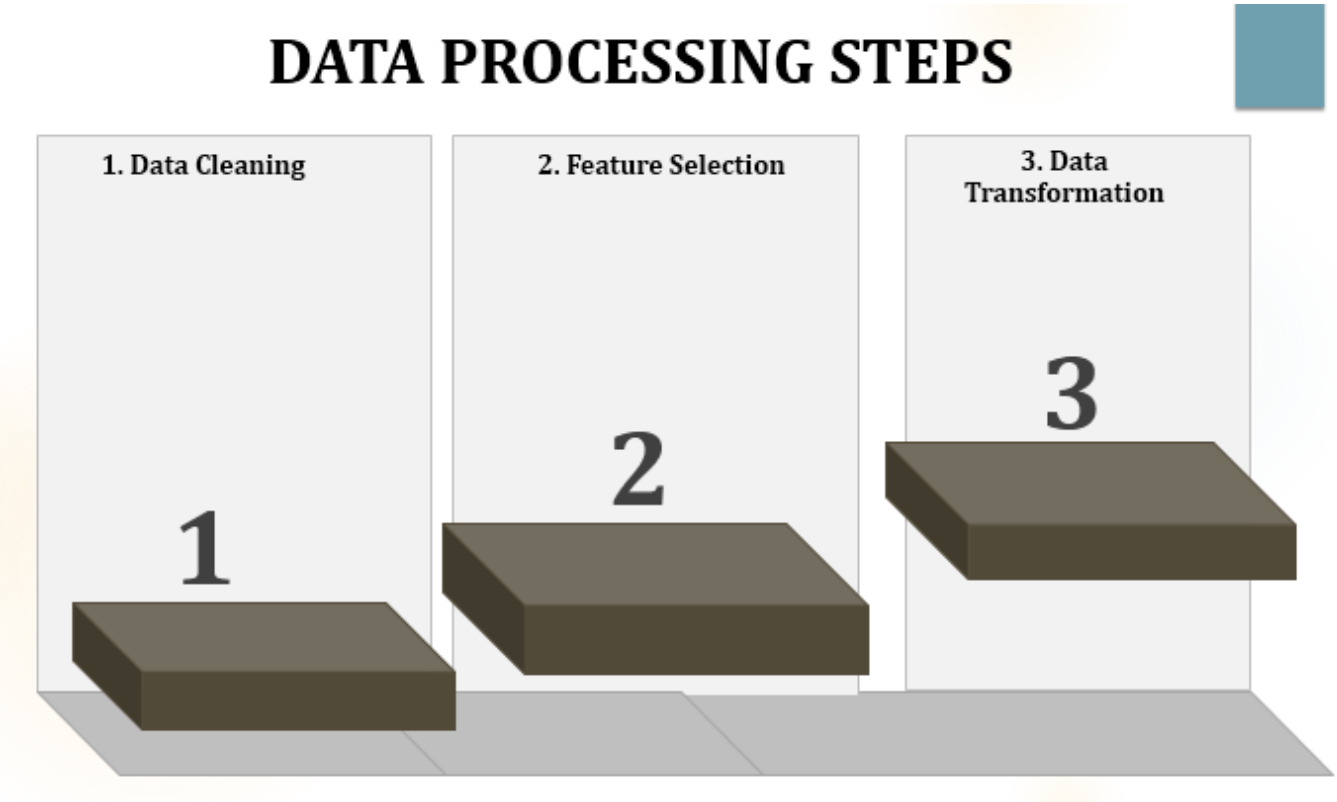

Fig 3.2 Data Processing Steps

3.2.1 Data Cleaning

This task involves filling of missing values, smoothing or removing noisy data and outliers along with resolving inconsistencies. As our data was collected from clinical record files of patients manually so there was very much chances inconsistencies and missing values. Following tasks are performed for data cleaning purpose.

- **Organizing Raw data**

  The data recorded from files is in the form of unorganized rows of patients data. So it needed to be transformed in some organized form. For this purpose the data is split in a number of columns specifying name, age, gender, marital status and others etc.

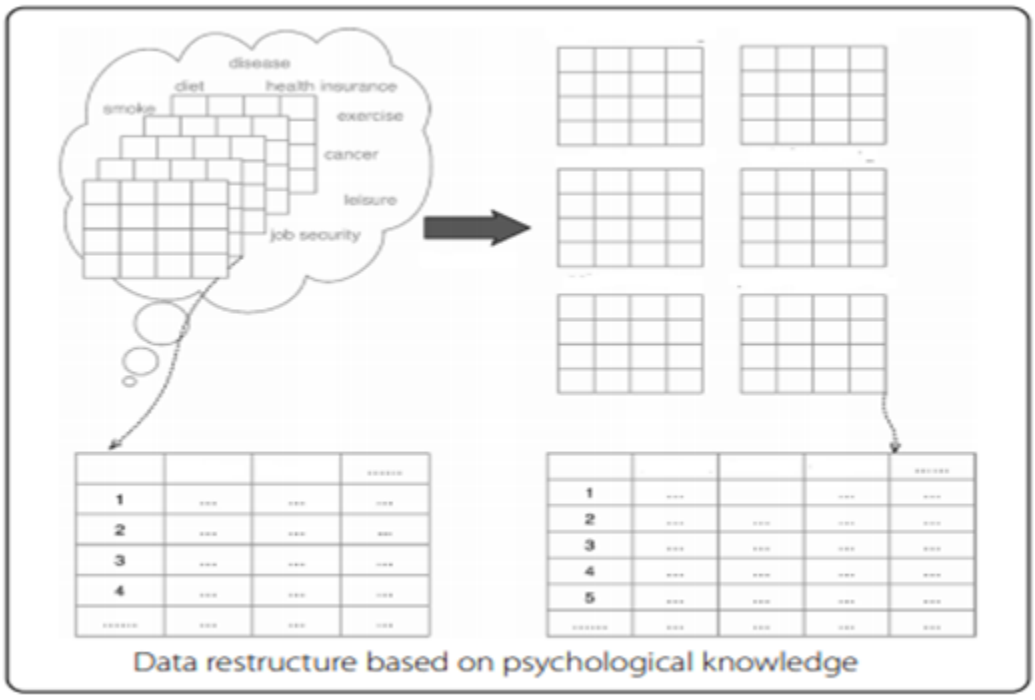

Fig 3.2.1: Data Restructure

- **Removal of duplicate data**

  The data may have possibility of presence of high record duplicates. This can lead to problems like: slow performance, degradation of data quality, waste of data storage and high operating cost.

  For the removal of duplicate entries the data is thoroughly explored and the duplicate entries are identified. The record is than Cross checked and verified. Finally the duplicate entries are replaced by correct values of data.

- **Removal of data inconsistency**

  Data Inconsistency can exist because our data is gathered from different files and formats. So there is quiet much chances of completely different and conflicting versions of identical information.

  For the removal of inconsistent data, the data is explored and cross checked with the original records of patients. The doubtful entries are also verified by the psychiatrist in supervision. After verification the correct and error free data is updated.

3.2.2 Feature Selection

Feature selection is considered an important and critical step in the development of machine learning programs. Feature selection is about selecting a set of features below the actual features. In order to:

- Reduce model complexity,
- Enhance the computational efficiency
- Reduce error due to noise by irrelevant features.

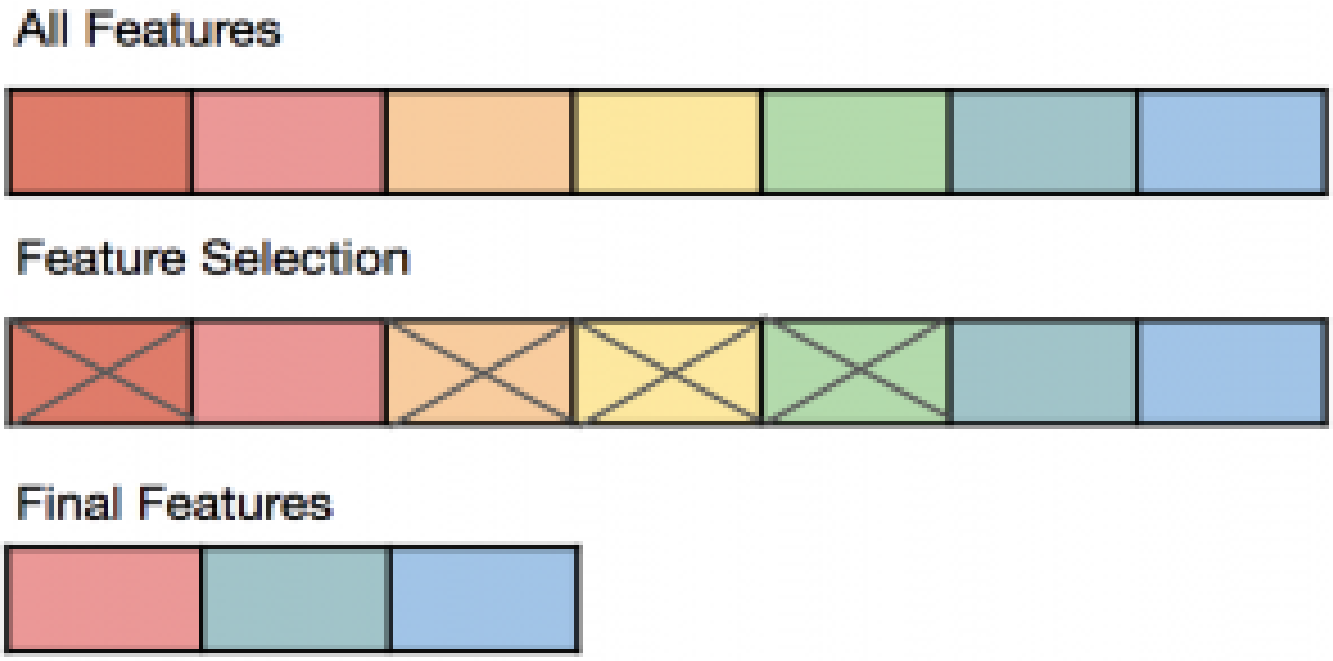

Fig: 3.2.2 : Feature Selection

The relevant features are selected and all the irrelevant and descriptive attributes are removed. As the data was majorly comprised of descriptive details of patients condition so it needed to be precise and selected on the basis of the relative importance of the feature selected in the diagnosis of the psychological disorder taken under consideration.

The feature selection is made by using deep knowledge of psychological domain and experience of the psychiatrist in supervision. Features targeting right factors and symptoms of the targeted patients are taken under consideration. Finally 24 Out of 32 features are

selected.

Features	Type	Attributes
Name	String	Full Name of Patient
Age	Integer	Age in years
Sex	String	Male=M Female=F
Mood	String	Low Normal
Appetite	String	Reduced Over eating Normal
Hallucinations	String	Yes No
Speech	String	Relevant Irrelevant
Fatigue	String	Yes No
Delusions	String	Yes No
Feeling	String	Anxious Irritability Normal
Past Failure	String	Yes No
Genetic Psy History	String	Yes No

Table 3.2.2-1: Features selected

Social Status	String	Friendly Not Friendly
Financial status	String	Normal Poor Rich
Suspiciousness	String	Yes No
Sleep	String	Disturbed Normal In Excess
Criminal Activity	String	Yes No
Temper	String	Aggressive Normal
Drugs	String	Yes No
Stress Coping	String	Less Normal High
Forgetfulness	String	Yes No
Insomnia	String	Yes No
Suicidal Tendency	String	Yes No
Diagnosis	String	MDD SCHZ BP-1

Fig 3.2.2-2: Selected Features

3.2.3 Data Transformation

Data is modified to make it more organized. Converted data can make it easier for both people and computers to use. Well-formatted and validated data improves data quality and protects applications from potential land mines such as useless values, unexpected duplicates, incorrect identification, and inconsistent formats.

For the clinical data we have taken under consideration transformation is important because the data taken from multiple record files needs to be in one standard form.

For all the 24 feature selected a few attributes are defined along with the format in which the attributes has to be specified.

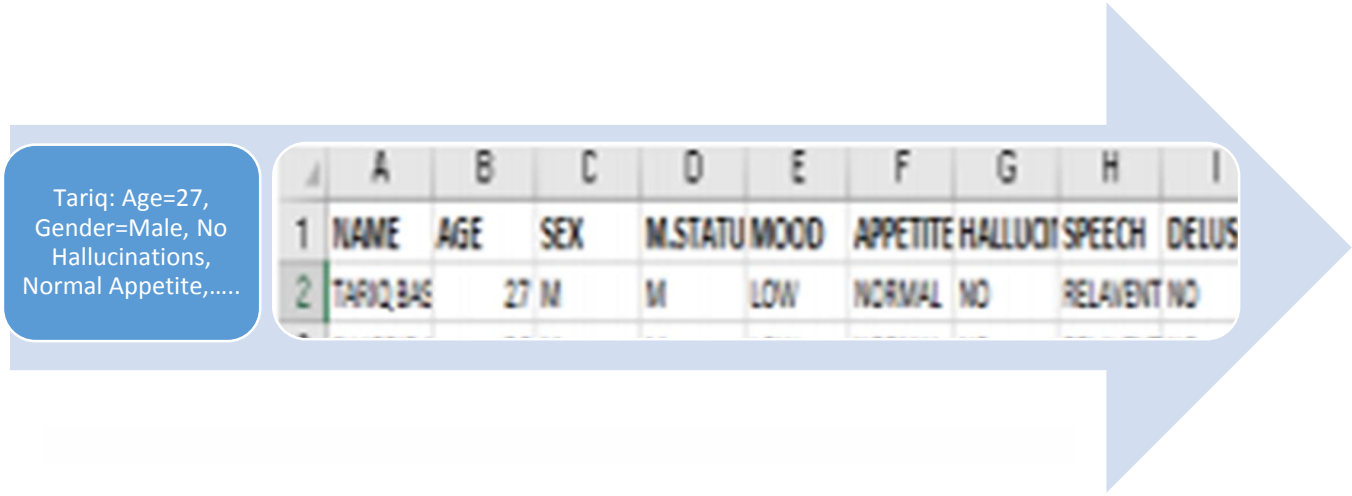

Fig 3.2.2-1: Data Transformation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	NAME	AGE	SEX	M.STATU	MOOD	APPETITE	HALLUCI	SPEECH	DELUSIO	FEELING	FATIGUE	SUICIDAL	GENETIC	PAST FA
2	TARIQ BAS	27	M	M	LOW	NORMAL	NO	RELAVENT	NO	ANXIOUS	YES	NO	NO	NO
3	SHABBIR A	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	NO	YES	NO
4	NAGAR FA	33	M	S	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO	NO
5	SHAFAQAT	40	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	YES	NO	NO
6	M. JAMAL	20	M	S	LOW	REDUCED	NO	IRRELAVEI	NO	IRRITATED	NO	YES	NO	NO
7	GHULAM I	29	M	M	LOW	NORMAL	NO	RELAVENT	NO	HELPLESS	YES	YES	NO	NO
8	UMAR ZAI	40	M	D	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	NO	NO	NO
9	ABDUL KA	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO	NO
10	ABDUL RA	42	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO	NO
11	SULTAN	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO	NO
12	M. EJAZ	23	M	S	LOW	REDUCED	NO	RELAVENT	NO	IRRITATED	YES	NO	NO	NO
13	TUFAIL AH	34	M	M	LOW	NORMAL	YES	IRRELAVEI	NO	IRRITATED	NO	YES	NO	NO

Fig 3.2.2-2: Transformed Data

3.3 Experimental Dataset

The preprocessed data is ready for the experimentation of our desired technique. For further processing of the psychological dataset, The XLXS datasheet is loaded in the ROSETTA system. Which is a software package for inducing rough-set based rule models. The system includes a large number of algorithms for discretization, reduct computation, rule pruning and classifier evaluation [21]. The Psychological dataset is split into two sets with a split factor of 0.90 i.e. The Training Set of the data comprises of 90% of the total data while test set comprises the rest of 10% of data.

	NAME	AGE	SEX	MARITAL STATUS	MOOD	APPETITE	HALLUCINATIONS	SPEECH	OCCUPATION	FEELING	FATIGUE	SUICIDAL TENDENCY	GENETIC PSYCHO
1	TAHIR BASHIR	27	M	M	LOW	NORMAL	NO	RELAVENT	NO	ANXIOUS	YES	NO	NO
2	SHABBIR ALI	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	NO	YES
3	NASAR FAR	38	M	S	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO
4	SHAFAQAT	40	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	YES	NO
5	M. JAMAL	29	M	S	LOW	REDUCED	NO	IRRELAVENT	NO	IRRITATED	NO	YES	NO
6	GHULAM MO	29	M	M	LOW	NORMAL	NO	RELAVENT	NO	HELPLESS	YES	YES	NO
7	UMAR ZAMA	40	M	D	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	NO	NO	NO
8	ABDUL KAR	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO
9	ABDUL RAU	42	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO
10	SULTAN	36	M	M	LOW	NORMAL	NO	RELAVENT	NO	IRRITATED	YES	NO	NO
11	M. EZAZ	29	M	S	LOW	REDUCED	NO	RELAVENT	NO	IRRITATED	YES	NO	NO
12	TUFAIL AHM	34	M	M	LOW	NORMAL	YES	IRRELAVENT	NO	IRRITATED	NO	YES	NO
13	M. AKRAM	38	M	M	LOW	NORMAL	NO	IRRELAVENT	YES	IRRITATED	YES	YES	NO
14	MOHSIN ALI	30	M	S	LOW	REDUCED	YES	RELAVENT	NO	IRRITATED	NO	YES	YES

Fig 3.3-1: Experimental dataset

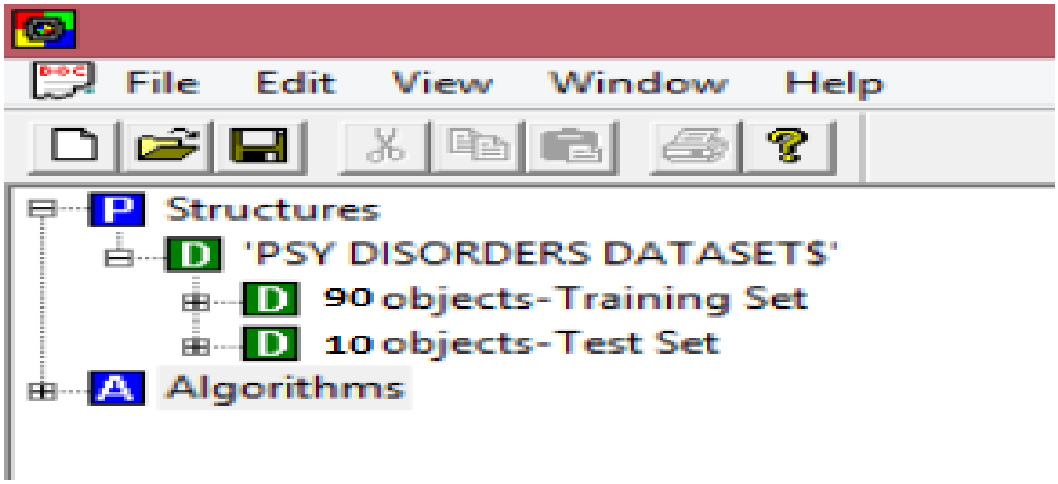

Fig: 3.3-2: Splitting dataset

3.4 Decision Rule Set Formation

The rule of thumb is a simple IF-THEN statement that contains status (also called antecedent) and prediction. The rules of the decision follow a general rule: IF the conditions are met THEN make some predictions. Decision rules are probably the most interpretive model. Their IF-THEN structure is similar to natural language and the way we think, as long as the structure is constructed from sound elements, the duration of the condition is short (minimum number of features = double number combined with AND) and there are not many rules. In the system, it is natural to write IF-THEN rules. What is new in machine learning is that the rules of decision are learned through an algorithm.

3.4.1 Generating Decision Rules

For generation of Decision rules and for further processing of the psychological dataset, The datasheet is loaded in the ROSETTA system. Which is a software package for inducing rough-set based rule models. The system includes a large number of algorithms for discretization, reduct computation, rule pruning and classifier evaluation [21]. The Psychological dataset is split into two sets with a split factor of 0.8 i.e. The Training Set of the data comprises of 80% of the total data while test set comprises the rest of 20% of data. To get decision rule sets, The Rosetta systems Johnson Algorithm is applied on the Training set of the data. After processing we will get the Reduct set and the Decision Rule Sets in the form as given.

	Reduct	Support
3	{ECONOMIC STATUS, TEMPER}	100
4	{FEELING}	100
5	{SPEECH, FORGETFULNESS}	100
6	{SPEECH, FATIGUE, DOMESTIC CONDITION}	100
7	{SPEECH, TEMPER}	100
8	{APPETITE, HALLUCINATIONS, DRUGS}	100
9	{HALLUCINATIONS, SOCIAL STATUS}	100
10	{DELUSIONS}	100
11	{APPETITE, HALLUCINATIONS}	100
12	{SOCIAL STATUS, INSOMNIA}	100
13	{APPETITE, FEELING}	100
14	{APPETITE, PAST FAILURE}	100
15	{HALLUCINATIONS, PAST FAILURE}	100
16	{APPETITE, HALLUCINATIONS, TEMPER}	100
17	{APPETITE, SPEECH, FATIGUE}	100
18	{SUSPECIOUSNESS, DRUGS}	100
19	{APPETITE, SUICIDAL TENDENCY, DRUGS}	100
20	{APPETITE, INSOMNIA}	100
21	{SPEECH, GENETIC PSY HIS, DOMESTIC CONDITION}	100
22	{APPETITE, SUICIDAL TENDENCY, SUSPECIOUSNESS}	100
23	{HALLUCINATIONS, FORGETFULNESS}	100
24	{APPETITE, FATIGUE, SUICIDAL TENDENCY}	100

Fig 3.4.1: Reduct Set

	Rule
1	FEELING(ANXIOUS) AND FATIGUE(YES) => DIAGNOSIS(MDD)
2	FEELING(ANXIOUS) AND FATIGUE(NO) => DIAGNOSIS(BP1)
3	CRIMINAL ACTIVITY(YES) => DIAGNOSIS(MDD)
4	ECONOMIC STATUS(POOR) AND TEMPER(NORMAL) => DIAGNOSIS(MDD)
5	FEELING(HELPLESS) => DIAGNOSIS(MDD)
6	SPEECH(RELAVENT) AND FORGETFULNESS(YES) => DIAGNOSIS(MDD)
7	SPEECH(IRRELAVENT) AND FORGETFULNESS(YES) => DIAGNOSIS(SCH)
8	SPEECH(RELAVENT) AND FATIGUE(YES) AND DOMESTIC CONDITION(DISTURBED) => DIAGNOSIS(MDD)
9	SPEECH(RELAVENT) AND TEMPER(NORMAL) => DIAGNOSIS(MDD)
10	APPETITE(NORMAL) AND HALLUCINATIONS(NO) AND DRUGS(YES) => DIAGNOSIS(MDD)
11	HALLUCINATIONS(YES) AND SOCIAL STATUS(NOT FRIENDLY) => DIAGNOSIS(SCH)
12	DELUSIONS(YES) => DIAGNOSIS(SCH)
13	APPETITE(OVER EATING) AND HALLUCINATIONS(YES) => DIAGNOSIS(SCH)
14	APPETITE(REDUCED) AND HALLUCINATIONS(YES) => DIAGNOSIS(SCH)
15	APPETITE(OVER EATING) AND HALLUCINATIONS(NO) => DIAGNOSIS(BP1)
16	SOCIAL STATUS(FRIENDLY) AND INSOMNIA(YES) => DIAGNOSIS(BP1)
17	APPETITE(REDUCED) AND FEELING(ANXIOUS) => DIAGNOSIS(BP1)
18	APPETITE(NORMAL) AND PAST FAILURE(YES) => DIAGNOSIS(BP1)
19	HALLUCINATIONS(NO) AND PAST FAILURE(YES) => DIAGNOSIS(BP1)

Fig 3.4-2: Decision Rules Generated

CHAPTER 4: RESULTS AND PERFORMANCE EVALUATION

Now when we have already gone through data collection, pre-processing and finally generation of the decision rules, this chapter will focus on evaluating the effectiveness of our approach. We will apply the decision rules generated on the test set. We will also discuss the validation process through which the performance evaluation of the process is made.

4.1 Overview

Confusion matrix is used to represent the complete pictorial view of the attained accuracy. Actual and predicted results are compared in the form of a table. All the three disorders i.e. major Depressive disorder (MDD), Schizophrenia (SCH) and Bi-Polar1 (BP1) are listed and the figures of each is shown against the respective column of predicted and actual values.

4.2 Generated Results

The test set on which the decision rules are applied consists of 20 records of patients. Out of which 10 patients of MDD and 4 patients of BP1 were predicted true. Whereas 5 out of 6 patients of schizophrenia are predicted true. The pictorial view of the accuracy shown by the system is given in the figure below.

	Predicted			
Actual	MDD	SCH	BP1	
	10	0	0	1.0
	0	5	1	0.833333
	0	0	4	1.0
	1.0	1.0	0.8	0.95

Fig 4.2: Generated Results

The result shows 100% accuracy for both Major Depressive disorder (MDD) and Bi polar1 (BP1). Whereas 83.3% accuracy was shown for Schizophrenia (SCH). The overall performance of the applied methodology is 95%.

4.3 Performance Evaluation Model using Cross Validation

Cross validation is process which is used to estimate and evaluate the performance of a created and designed model. voting Operator is used to combine the predicting power of more than one classifiers to attain better results. The trained models are tested with the help of cross validation technique. The Value of K is selected in such a way that the about 80% of data has been used as training and remaining 20% data as testing. The experimental procedures are repeated K times.

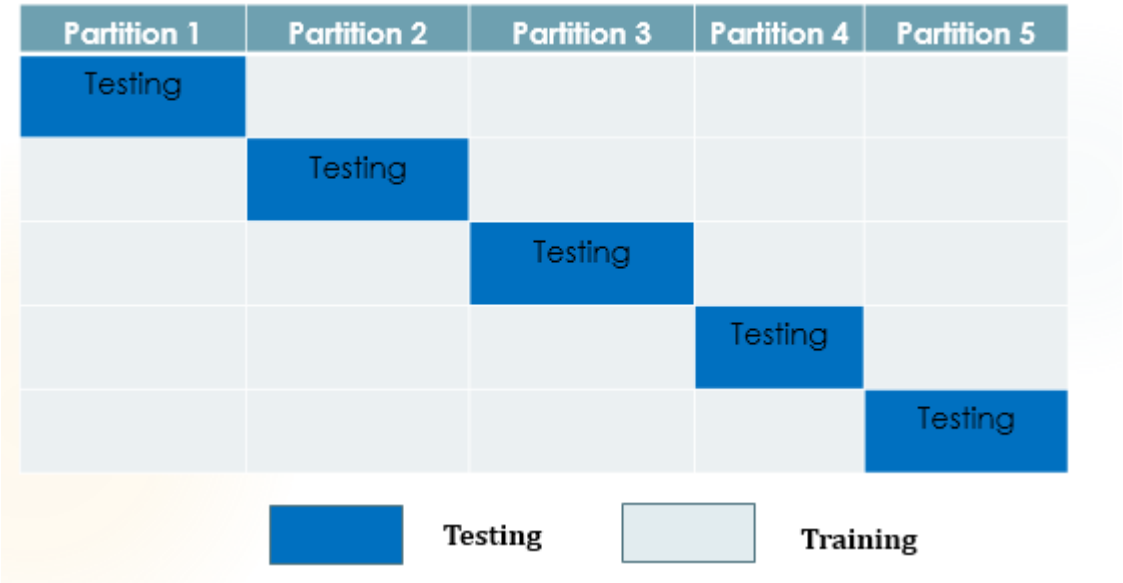

Fig 4.3: Partitions of Dataset

4.3.1 Confusion Matrix for Partition-1:

	PREDICTED			
	MDD	**SCH**	**BP1**	
ACTUAL	10	0	0	1.0
	0	5	1	0.833
	0	0	4	1.0
	1.0	1.0	0.80	0.95

Fig: 4.3.1 Confusion Matrix for partition-1

4.3.2 Confusion Matrix for Partition-2

	PREDICTED			
	MDD	**SCH**	**BP1**	
ACTUAL	8	0	0	1.0
	0	6	1	0.857
	0	0	5	1.0
	1.0	1.0	0.833	0.95

Fig 4.3.2: Confusion Matrix for Partition-2

4.3.3 Confusion Matrix for Partition-3

	PREDICTED			
	MDD	**SCH**	**BP1**	
ACTUAL	9	0	0	1.0
	0	4	0	1.0
	0	1	6	0.857
	1.0	0.8	1.0	0.95

Fig 4.3.3 Confusion Matrix for Partition-3

4.3.4 Confusion Matrix for Partition-4

	PREDICTED			
	MDD	**SCH**	**BP1**	
ACTUAL	7	0	0	1.0
	0	6	1	0.857
	0	0	6	1.0
	1.0	1.0	0.857	0.95

Fig 4.3.4: Confusion Matrix for Partition-4

4.3.5 Confusion Matrix for Partition-5:

	PREDICTED			
	MDD	**SCH**	**BP1**	
ACTUAL	9	0	0	1.0
	0	4	1	0.80
	0	0	6	1.0
	1.0	1.0	0.857	0.95

Fig 4.3.5: Confusion Matrix for Partition-5

CHAPTER 5: CONCLUSION AND FUTURE WORK

In this chapter, we will conclude our work and some future advise will also be given in this chapter. In Section 5.1, conclusion of our research work is made and in section 5.2 future work is explained.

5.1 Conclusion

Various data mining methods are designed to detect mental disorders. But it is designed for the effective treatment of mental illness so that the disease can be properly diagnosed. Although clinical diagnostic guidelines are well established, i.e., through the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (DSM), the Mental Disorders Diagnostic Manual provides a very good variation of various psychiatric disorders. [15] But still there is a point where certain mental disorders are passed due to similarities in the occurrence of their symptoms. Therefore, it is almost impossible for a psychiatrist to diagnose these problems without making a mistake. Even for some psychologists it is difficult to distinguish between certain mental states. In this study our main focus was on making such a database that could provide accurate indications for the diagnosis of Hyper Depression Disorder, Schizophrenia and Bi-Polar Affective under the guidance of trained psychologists at the Arms Forces Institute of Mental Health, Rawalpindi, Pakistan. We then adjusted the clinical database and installed the Johnsons Algorithm to obtain Decision Rule sets of Training data. Finally, decision rules are applied to the test set to determine the final result and accuracy of the method used. The result shows 100% accuracy of HDD patients with Schizophrenia, while 90% accuracy of BP Affective Disorder. Achieving a complete 95% accuracy which is a quiet percentage that keeps looking at previous strategies to date.

5.2 Future Work

Going forward, this study presents a method that can help in the initial diagnosis of depressive cases in a large number of potential cases before a formal clinical diagnosis. What is important is that we show that by decision-making rules a variety of mental disorders can be better identified. However, the reliability and sensitivity of this system requires testing on additional data sets. In particular, incorporating other aspects of the mental set will provide further evidence of the importance of mental functioning. Psychological factors appear to be the most important factors in predicting depression, however, further examination of these sub-sets in isolation will improve the functioning of isolation and understanding about the relationship between factors and depression. Future research indicators for using our method include using rich clinical and factual information from various other mental health care units operating in the country and those working abroad.

In addition the results of Bi-Polar Affective can also be improved if we apply the process to big data. This will not only lead to the development of a set of decision rules but will also improve the overall accuracy of the approach. The study also suggests that the development of decision-making rules is the most effective method used in the medical field and will provide assistance and assistance to physicians in diagnosing complex diseases.

REFERENCES

[1] "Psychiatry Organization" [Online].Available: https://www.psychiatry.org/patients-families/what-is-mental-illness

[2] "International Journal of Emergency Mental Health and Human Resilience" [Online].Available: https://www.omicsonline.org/open-access/mental-health-pakistan-optimizing-brains-1522-4821-17-160.php?aid=37919

[3] "Mental Illness", http://www.who.int/mediacentre/factsheets/fs300/en/, [last retrieved: 22nd Aug, 2021]

[4] Paetz, J, Brause, R. *A frequent patterns tree approach for rule generation with categorical septic shock patient data.* In: Proceedings of the international symposium on medical data analysis, Madrid, 8–9 October 2001, pp. 207–213

[5] Brossette, SE, Sprague, AP, Hardin, JM, et al. *Association rules and data mining in hospital infection control and public health surveillance.* J Am Med Inform Assoc 1998; 5(4): 373–381

[6] Ohsaki, M, Sato, Y, Yokoi, H, et al. *A rule discovery support system for sequential medical data, in the case study of a chronic hepatitis dataset.* In: Proceedings of the workshop notes of the international workshop on active mining, at IEEE international conference on data mining, Brussels, 10 December 2002, pp. 121. New York: IEEE

[7] Ordonez, C, Ezquerra, N, Santana, CA. *Constraining and summarizing association rules in medical data*. Knowl Inform Syst 2006; 9(3): 1–2.

[8] Ordonez, C, Omiecinski, E, De Braal, L, et al. *Mining constrained association rules to predict heart disease*. In: Proceedings of the IEEE international conference on data mining, San Jose, CA, 29 November–2 December 2001, pp. 433–440. New York: IEEE.

[9] Nahar, J, Imam, T, Tickle, KS, et al. *Association rule mining to detect factors which contribute to heart disease in males and females.* Expert Syst Appl 2013; 40(4): 1086–1093.

[10] Ordonez, C . *Comparing association rules and decision trees for disease prediction*. In: Proceedings of the international workshop on healthcare information and knowledge management, Arlington, VI, 11 November 2006, pp. 17–24.

[11] De Choudhury M, Counts S, Horvitz E (2013) *Social media as a measurement tool of depression in populations.* https://doi. org/10.1145/2464464.2464480

[12] Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H (2015) *Recognizing depression from twitter activity*. pp. 3187–3196. https:// doi.org/10.1145/2702123.2702280

[13] Peng Z, Hu Q, Dang J (2017) *Multi-kernel svm based depression recognition using social media data. Int J Mach Learn Cybern.* https://doi. org/10.1007/s13042-017-0697-1

[14] Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S (2017) *Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression.* IEEE, New York. https:// doi.org/10.1109/ICTC.2017.8190959

[15] Fatima I, Mukhtar H, Ahmad HF, Rajpoot K (2018) *Analysis of usergenerated content from online social communities to characterise and predict depression degree.* J Inform Sci 44(5):683–695. https://doi. org/10.1177/0165551517740835

[16] Reece AG, Danforth CM (2017) *Instagram photos reveal predictive markers of depression. EPJ Data Sci 6(1):15.* https://doi.org/10.1140/ epjds/s13688-017-0110-z

[17] Carlos Ordonez. (2006). "*Comparing Association Rules and Decision Trees for Disease Prediction"*. HIKM, Virginia, USA.

[18] Chang YS, H. W. (2014). *Depression diagnosis based on analogies and bayesian network.* IEEE.

[19] Daniel Jachyra, K. P. (2011). *"Classification of MMPI Profiles using Decision Trees*". Proceedings of the international workshop CS&P .

[20] Ghafoor, Y. H. (2015). *An intelligent approach to discovering common symptoms among depressed patients*. . Soft Computing.

[21] Hvidsten, T. (2010). *A Tutorial-Based Guide to the ROSETTA System: A Rough Set Toolkit for Analysis of Data*. Journal of Computer and Communications,.

[22] Jain, S. (2018). *A Data Mining Analysis Over Psychiatric Database for Mental health*. International Journal on Future Revolution in Computer Science & Communication Engineering

[23] Ron kohavi. (2011). *"Scaling up the Accuracy of Naive-Bayes Classifiers: a Decision Tree Hybrid*". Data Mining and Visualization. Silicon Graphics, Inc

[24] Shatte AB, Hutchinson DM, Teague SJ (2019) *Machine learning in mental health: a scoping review of methods and applications.* Psychol Med 49(9):1426–1448

[25] Hsieh W-H, Shih D-H, Shih P-Y, Lin S-B (2019) *An ensemble classifer with case-based reasoning system for identifying internet addiction.* Int J Environ Res Public Health 16(7):1233

[26] Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A (2018) *Depression detection from social network data using machine learning techniques.* Health Inform Sci Syst 6(1):8