

Bachelor of Science in Computer Science & Engineering



**Predicting Depression Level Based on Human  
Activities and Feelings**

by

Urmi Saha

ID: 1504046

Department of Computer Science & Engineering  
Chittagong University of Engineering & Technology (CUET)  
Chattogram-4349, Bangladesh.

August, 2021

# Predicting Depression Level Based on Human Activities and Feelings



Submitted in partial fulfilment of the requirements for  
Degree of Bachelor of Science  
in Computer Science & Engineering

by

Urmi Saha

ID: 1504046

Supervised by

Dr. Md. Iqbal H.Sarker

Associate Professor

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

The thesis titled ‘**Predicting Depression Level Based on Human Activities and Feelings**’ submitted by ID: 1504046, Session 2018-2019 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

## **Board of Examiners**

---

Chairman

Dr. Md. Iqbal H.Sarker  
Associate Professor  
Department of Computer Science & Engineering  
Chittagong University of Engineering & Technology (CUET)

---

Member (Ex-Officio)

Dr. Md. Mokammel Hoque  
Professor Head & Head  
Department of Computer Science & Engineering  
Chittagong University of Engineering & Technology (CUET)

---

Member (External)

Muhammad Kamal Hossen  
Associate Professor  
Department of Computer Science & Engineering  
Chittagong University of Engineering & Technology (CUET)

# Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

---

**Signature of the candidate**

**Date:**

# Acknowledgements

All praises are due to Almighty Allah, Most Gracious, Most Merciful who blessed us to be here at Chittagong University of Engineering Technology for pursuing the bachelors degree.

Our journey towards the bachelors degree and in particular this thesis would not have been possible without the help of many people. It is our great pleasure to take this opportunity to thank them for the support and advice that we have received.

Our deepest gratitude goes to our Supervisor, Dr. Md. Iqbal Hasan Sarker, PhD for his significant help and support throughout the course of this work. Without his continuous encouragement and guidance this thesis would not have been completed. We are truly grateful to have worked with such inspiring and friendly supervisors. We would also like to thank him for his valuable comments and feedback while designing the survey form and throughout the entire data collection and manipulation process.

We would also like to thank each and every one who were involved in the data collection process from my friends, faculty members, fellow classmates, juniors and all the respondents who took time out of their busy schedule to complete our survey. This research would not have been possible without their help.

We would also like to thank our family members, our parents, brothers and sisters who are the very reason of our existence. Without their unconditional love and support this research too could not be completed.

This thesis is dedicated to all of them.

# Abstract

Depression is a major illness and a growing issue that affects a person's way of life, affecting normal functioning and impeding thought processes while they may be completely unaware of their condition. Depression is particularly common among people in underdeveloped and developing countries. People challenges with their studies, careers, relationships, drugs, and family issues, all of which are major or minor contributors to depression. Furthermore, people are uneasy speaking about this condition, and they often confuse it as insanity. This study aims to gain useful insights into why people suffer from depression, in addition to predicting depression level in people for the purpose of referral to a psychiatric facility. A survey was created with the help of psychologists and professors to gather the data for this study. The survey was conducted using a Google survey form. Using the Pearson Correlation method and R-Squared method, 15 features were selected among 30 features, then fuzzy rules were generated for 5 membership function (poor, mediocre, average, decent and good). The data was analyzed to identify specific features related to depression and then to predict depression level based on those features. The aim of this study is to identify the causes and level of depression. People prefer to neglect depression, which causes it to build up and worsen over time. This study aims to predict depression at its early stages and ensure that patients recover quickly and if it crosses early stage, then preventing tragic events such as suicide. In comparison to existing methods, our proposed methodology has achieved a high test accuracy of 95.71 percent.

**Keyword:** Depression Level Prediction, Machine Learning, Feature selection, Statistical method, Pearson Correlation method, R-Squared method, Fuzzy Rules.

# Table of Contents

Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.1.1 Depression Types and Symptoms . . . . .	2
1.1.1.1 Fuzzy Rules . . . . .	3
1.2 Depression Level Prediction Framework . . . . .	3
1.3 Difficulties . . . . .	4
1.4 Applications . . . . .	5
1.5 Motivation . . . . .	6
1.6 Contribution of the thesis . . . . .	7
1.7 Thesis Organization . . . . .	7
1.8 Conclusion . . . . .	8
<b>2 Literature Review</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 Related Literature Review on Predicting Depression Level Frame- work . . . . .	9
2.2.1 History of Depression . . . . .	9
2.2.2 Related Work Based on Social Site Data . . . . .	10
2.2.3 Related work as a recommendation system . . . . .	12
2.2.4 Related work based on generating rules . . . . .	13
2.3 Conclusion . . . . .	13
<b>3 Methodology</b>	<b>14</b>
3.1 Introduction . . . . .	14
3.2 Diagram/Overview of Framework . . . . .	14
3.3 Detailed Explanation . . . . .	15
3.3.1 Data Collection . . . . .	15

3.3.2	Preparing Questionnaire . . . . .	15
3.3.3	Collection of respondents' data . . . . .	16
3.4	Pre-processing of Dataset . . . . .	17
3.5	Features Selection . . . . .	18
3.5.1	Pearson Correlation method . . . . .	18
3.5.2	R-squared method . . . . .	19
3.5.3	Selected Features . . . . .	19
3.6	Fuzzy Logic System . . . . .	20
3.6.1	Algorithm . . . . .	22
3.7	Membership Function . . . . .	22
3.8	Fuzzy Rules . . . . .	23
3.8.1	Sample Input . . . . .	24
3.8.2	Sample Output . . . . .	25
3.9	Conclusion . . . . .	26
<b>4</b>	<b>Results and Discussions</b>	<b>27</b>
4.1	Introduction . . . . .	27
4.2	Dataset Description . . . . .	27
4.3	Impact Analysis . . . . .	29
4.3.1	Social Impact . . . . .	29
4.3.2	Ethical Impact . . . . .	29
4.4	Evaluation of Framework . . . . .	30
4.5	Evaluation of Performance . . . . .	30
4.6	Conclusion . . . . .	31
<b>5</b>	<b>Conclusion</b>	<b>32</b>
5.1	Conclusion . . . . .	32
5.2	Future Work . . . . .	33



# List of Figures

1.1	Block diagram of Depression Level Prediction Framework . . . . .	4
3.1	The proposed system of Depression Level Prediction . . . . .	14
3.2	Raw sample of Dataset . . . . .	17
3.3	Pre-processed sample of Dataset . . . . .	18
3.4	Overview of Fuzzy Logic System . . . . .	21
3.5	Overview of Membership Function . . . . .	23
3.6	Fuzzy Inference rule generation. . . . .	23
3.7	Fuzzy Inference rule generation. . . . .	24
3.8	Fuzzy Inference rule generation. . . . .	24
3.9	Sample Input . . . . .	25
3.10	Sample Output . . . . .	25

# List of Tables

4.1	Performance Evaluation . . . . .	30
-----	----------------------------------	----

# Chapter 1

## Introduction

### 1.1 Introduction

The world is progressing at a rapid rate with the help of technology and human skills. Everyone is becoming so busy and materialistic that sometimes we forget to give some time to think about our mental health. To keep up with the rapid pace in the world people are constantly pushing themselves taking a lot of pressure both physically and mentally which have a adverse effect on their health especially their mental health. Depression is a silent killer which can harm a human being in a great way if not treated at the right time. It is a common mental illness and everyone at some point of their life is depressed. However, due to the lack of self-unawareness, the society and people's judgment this illness is considered as a taboo in many places across the world and people tend to make fun of this illness and tease people who are diagnosed with this problem. For instance, in Bangladesh, if someone has mental illness or is diagnosed with similar illness people assume that this person is mad.

According to most recent statistics, more than 350 million people of all ages suffer from depression globally. Most of the people is being affected by depression in their lives either directly or indirectly. As an individual person's depression not only affect them but also affect their loved one and can make them depressed too. Women are affected by depression more than men. For instance, women are 70

In the era of Survival with Depression, the 'Psychiatric Disorder' actually is dominating a large number of people and that doesn't mean - depression is their weakness but a serious mental illness. The thing can be overcome. As depression detection is mostly based on emotion health and measuring depression level on

basis of so much intimate and personal questionnaire where giving hundred percent privacy without breaking any ethical law is the most challenging sector of this thesis.

### **1.1.1 Depression Types and Symptoms**

There are common types of depression disorder as example of seven types - Major Depressive Disorder, Persistent Depressive Disorder, Bipolar Disorder, Postpartum Depression, Premenstrual Dysphoria Disorder, Seasonal Affective Disorder, Atypical Depression. [4]

On the basis of some common symptoms which occurs mainly in every type of depression, they can be grouped into together to identify the main symptoms with its sub symptoms. Here ,the common types symptoms can be joined together like -

1. Sleep disturbances, including insomnia, sleeping over time and cry over night.
2. Frequent tiredness, fatigue or passing most of the time without moving or laying, feeling back pain or migraine.
3. Difficulty in remembering problem, concentration losing and also losing interest in hobbies or sex.
4. Change in appetite, losing weight or weight gain.
5. Optimistic about individual dream , unsatisfied result in working sector, negative thinking about future.
6. Feeling broken, unworthy or burden.
7. Relationship status, bonding with family, social relationship and feeling loneliness.
8. Past traumas, feeling guilty, losing or death of loving one.
9. Copying others or feeling low watching other people, losing self confidence

10. Losing of self-control, drug addiction, anger and anxiety disorder , dominated by sub-conscious mind .
  11. Thinking about death of yourself always or suicide, attempt to kill yourself.
- [1,3]

#### **1.1.1.1 Fuzzy Rules**

Our study relies on Fuzzy rules. In medicine [1], control [2, 3, 4], classification [5], prediction [6], inference [7], decision making [8][9], image segmentation [10], and medical diagnostics [11, 12, 13, 14, 15], fuzzy systems are increasingly employed in a variety of practical applications. Because they deal with if-then rules whose antecedents and effects are constructed of fuzzy logic statements rather than classical logic statements, fuzzy systems are an extension of traditional rule-based systems. A fuzzy system is a rule-based system in which fuzzy logic is utilized to express various types of knowledge about a topic as well as to simulate the interactions and relationships that exist between its variables. Fuzzy logic principles have been effectively used to a wide range of problems in various disciplines where uncertainty and vagueness appear in various ways due to this trait.

## **1.2 Depression Level Prediction Framework**

This research has been conducted in time span of 1 year and 6 months. However we are continuing with this research as we are planning to extend our research to discover more findings. We have started our work from 20th of September in 2019 which last till 20th March 2021. In first quarter of September we spent time on brainstorming about our ideas and we took around 1 month to come up with our initial idea. After that we started to read similar articles that is related to our idea. At that phase of time our idea was still in development process. At the same time we met with many counsellor to know more about depression and its related work. They were so helpful to us and helped us throughout whole time till the last stage of our thesis. Taking guidance from our supervisor we were able to finalize our idea and we started our studies on depression, to find out level

of depression and did research to find our relevant reasons and features which are related to depression in people based on their activities and feelings. We were really careful when we were preparing our survey form.

The survey was conducted using a Google survey form. Using the Pearson Correlation method and R-Squared method, 15 features were selected among 30 features, then fuzzy rules were generated for 5 membership function (poor, mediocre, average, decent and good). The data was analyzed to identify specific features related to depression and then to predict depression level based on those features. After training and testing the data set, Fuzzy Rules was found to have the accuracy of 95.71 percent for 140 participants. The block diagram of our proposed framework is represented below:-

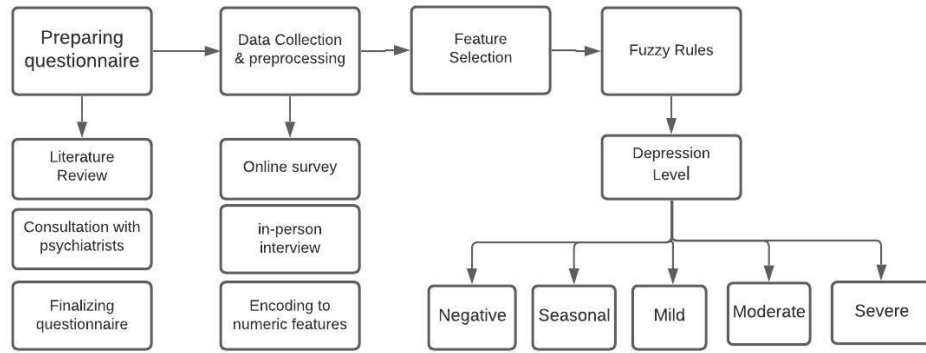


Figure 1.1: Block diagram of Depression Level Prediction Framework

## 1.3 Difficulties

In our work, we worked with more than 550 data collected from various people from different age and profession. The major challenges are enlisted below:-

1. **Hesitate to open up about their problem:** Despite the fact that the world is progressing and everyone is open to new experiences, mental illness is not well received in many cultures. It appears to be taboo in many Asian areas, as a result of which people who are experiencing it are less outspoken about their problem.
2. **Assume the problem as ignorant:** The majority of people are unaware

of their depression problem and do not treat it as seriously as other illnesses, assuming that it will go away on its own. So, they don't take care of their mental health. As a consequence, their condition can worsen, resulting in deterioration of their mental and physical health, as well as extreme depression, which may lead to suicidal behaviour and self-harm.

3. **Mixing up depression with common grief** In the other hand, some people become depressed because they believe they are depressed, although this is not the case. While conducting this study, we discovered that the majority of people behave in this manner and are not as articulate about their problems; some of them believe they are depressed, which is affecting their performance, but they are not, and some are unaware that they have moderate to extreme depression symptoms.
4. **Unaware of depression symptoms:** We've discovered that there are some who are unsure and want to know whether they're depressed or not so that they can get help. However, the nuanced questionnaire's inconsistencies and the embarrassment of disclosing all personal details to clinicians demotivated them from taking action to address their mental disorder.
5. **Identifying root cause of depression level:** We discovered through our study is that simply determining whether the person's depression would not help them recover and heal from this mental disorder. One of the top priorities should be identifying the root problems or level of depression in people. As a result, there is no effective forum for people to decide depression level by simply filling out some simple questionnaires

## 1.4 Applications

Depression level prediction can be used where it is necessary to predict someone's mental health condition. Some examples of how this system can be used today is given below:

1. **Individual use:** A person who feels hesitate to go to doctors or psychiatrists and open up about his personal information in front of them can easily rely on this system and can get to know about their situation. Besides because of Covid, people can't go outside, so this system can perform an important role.
2. **Mental Health Care service:** Mental health doctors and psychiatrists can use this model to be sure about their patients' situation as sometimes their experience can be wrong or they might be unsure about the root cause of their depression.
3. **To reduce anxiety:** People from the general population who suffer from mood and anxiety problems may benefit from improved nutritional status achieved with nutritional supplements. Our data set contains a questionnaire 'Can you eat when you are sad?' which can be used to predict their nutritional status and then can be used to reduce their anxiety as this also depends on lack of vitamin-D.[10]
4. **To reduce suicidal cases:** Depression is the most common mental disorders that at its worst can lead to suicide. Diagnosing depression in the early curable stage is very important. Various system like sleep disorders and alcoholism is in our questionnaire. This database can be used in Electroencephalogram Gram (EEG) signals and can help in classifying depression level. For this aim, the features are extracted from database.

## 1.5 Motivation

The main motivation of this study is to determine whether or not a person is depressed using a machine learning and data analysis approach by simply filling out some simple depression-related questionnaires to predict depression level among people. Along with this, we want to make it easier for mental health counselors to identify the root causes of depression in human, as well as the causes that are unrelated to depression, so that they can better understand human psychology and have the best guidance and treatment for their issues. We used two algorithms



to run our model to see whether it could predict the level of depression. Genetic algorithm and random forest algorithm were used in this system. In chapter 3 and chapter 4 of our paper, we showed to pick optimal features or reasons that are related to depression and exclude irrelevant reasons that are not related to depression in general people.

## 1.6 Contribution of the thesis

The primary focus of our project is to develop a system which can predict depression level using machine learning. The contributions are as follows,

1. Sample data collection from various sectors of people of all age.
2. Collecting analyzing intimate facts or personal data ensuring ethical law.
3. Surveying on self-reporting depression questionnaire among people.
4. Evaluate the level of depressed people within short time to recover an individual.
5. To separate depression from natural grief or feeling sad.
6. To make grouping among the depressed and non-depressed people based on their activities and feelings.
7. To evaluate the performance of prediction model in real environment.

## 1.7 Thesis Organization

This report is organized into six chapters. Chapter one contains some introductory readings on depression, some challenges of implementation of our work, motivation of our work and the contributions we made. Chapter two contains brief discussion on previous works that is already implemented, their limitations and their role on using machine learning. Chapter three describes proposed system with necessary diagrams. An overall system architecture is given on this chapter. In chapter four our implementation of the project in details have been

illustrated. Chapter five focuses on the experimental results of the system. Evaluation measures and results of our system are described in this chapter. Chapter six consists of conclusion with the summary of our system and the future plan of our system.

## 1.8 Conclusion

In this chapter, a brief overview of our research work has been discussed. A summary of the proposed framework, as well as the motivation behind this thesis and its applications, have been represented. Finally, our research contributions along with the challenges faced are also stated. In the next chapter, background, literature reviews and the current state of the problem will be provided.

# Chapter 2

## Literature Review

### 2.1 Introduction

In this chapter we will briefly describe history of predicting depression and learn about different machine learning methods which are really useful for classifying depression. This chapter also contains brief discussion on related previous works.

### 2.2 Related Literature Review on Predicting Depression Level Framework

#### 2.2.1 History of Depression

Major depression, or simply depression, commonly referred to as major depressive disorder by many psychologist and Health care professionals, which has a long history and previously known as Melancholia and is now known as clinical depression. The word depression was derived from the Latin verb *deprimere*, which means "to press down". From the 14th century, "to depress" was used to express the idea of bringing down in spirits. The term also came into use in physiology and economics.

A half century ago, diagnosed depression was considered as endogenous (melancholic) or a biological condition, or reactive (neurotic), a reaction to stressful events. During the 1960s and 70s, manic-depression, one type of mood disorder (now as bipolar disorder) was distinguished from uni-polar depression. German psychiatrist Karl Kleist coined the terms of uni-polar and bipolar. A group of US clinicians introduced with Major depressive disorder in the mid-1970s for diagnostic criteria based on patterns of symptoms.

In the mid-20th century, researchers stated that a chemical imbalance in neurotransmitters in the brain was caused by the effects of reserpine and isoniazid in altering monoamine neurotransmitter levels and affecting depressive symptoms, which was rising day by day a rapid number in this era.

### **2.2.2 Related Work Based on Social Site Data**

Sharath Chandra Guntuk, David B Yaden, Margaret L Kern, Lyle H Ungar and Johannes C Eichstaedt, published on July 2017 ) worked on “Detecting Depression and mental illness on social media : an integrative review”, participants were recruited to take a depression survey and share their Facebook or Twitter data, self-declared mental health status. Based on that the paper declared the prediction of mental illness using Linear Regression and Support Vector system based on survey responses. The investigations inspected here recommend that downturn and other psychological instabilities are distinguishable on a few on the web conditions, however the generalizability of these investigations to more extensive examples and highest quality level clinical models has not been set up.[16]

Jude Mikal, Samantha Hurst, Mike Conway , published on August 2017 ) worked on “Investigating Patient Attitudes Towards the use of Social Media Data to Augment Depression Diagnosis and Treatment: a Qualitative Study”, based on the secondary analysis of a series of five focus groups with Twitter users, conducted by author JM. Two groups were made up of participants without a diagnosed history of depression and three of the groups were formed with participants with a diagnosed history of depression. This paper used NLP for mood tracking under the supervision of a mental health partitioner. Note that while there were a few questions communicated concerning the capacity of NLP calculations to effectively distinguish mental status from online media information (for example a technological limit) the majority of the conversation around accuracy fixated on inquiries of self-show in online media. Generally, members were enthusiastic about select in usage of social media with regards to clinician-drove emotional well-being care, yet in any event for certain members, there was some distrust identified with how well online media addresses the psychological wellness of clients.[17]

Maryam Mohammed Aldarwish et. al worked together also on “Predicting Depression Levels Using Social Media posts” using SVM and Naïve Bayes classifier. The first dataset is the training dataset which contains the manually trained 2073 depressed posts and 2073 not-depressed posts. It consists of three columns, the first one is binominal sentiment ( identified Depressed, Not-Depressed), the second column contains the depression category, and the third column contains the trained post. The second dataset consists of the patient SNS posts and it is changed for every individual to test the prediction of the model. The critical thought of their examination is to discover the affiliation among SNS clients’ exercises and psychological wellness disease. They proposed a web application that can arrange SNS client into one out of four sadness level, which could be utilized by therapists, family, and companions of the discouraged patient. The web application gathers the UGC from the patient’s Twitter as well as Facebook. It characterizes the patient into one out of four levels (Minimal, Mild, Moderate, or Extreme melancholy. They have made an originating before misery model utilizing RapidMiner to test two classifiers (SVM, and Naïve Bayes Classifier).[18]

Bridianne O’Dea,, Helen Christensen worked together on "Detecting Suicidality on Twitter", This examination planned to inspect whether the degree of worry for a self destruction related post on Twitter could be resolved in view of on the substance of the post, as decided by human coders and afterward duplicated by AI. The current examination showed that it is feasible to recognize the level of depression among self destruction related tweets, utilizing both human coders and a programmed machine classifier. Significantly, the machine classifier duplicated the exactness of the human coders. The discoveries affirmed that Twitter is utilized by people to communicate suicidality and that such posts evoked a degree of depression that justified further examination. Nonetheless, the prescient force for real self-destructive conduct isn’t yet known and the discoveries don’t straightforwardly distinguish focuses for mediation.[19]

They portray their prediction framework dependent on a gathering characterization approach, which joins supervise learning, information retrieval and features selection methods. During their test stage, they have performed starter tests to assess the use of three different techniques: simple rule-based classification using

a sentiment analysis library, deep learning-based classification using a Recurrent Neural Network (RNN), and topic extraction using Latent Dirichlet Allocation. Their lacking was enhancement of the IR-based resources of the system.[20]

Depression is hard to diagnose, bringing about high under-conclusion. Diagnosing depression is regularly founded on self-announced encounters, practices announced by family members, and a psychological status assessment. At present, specialists use studies and polls to recognize people who might be at hazard of sorrow. This cycle is tedious and expensive. They propose a mechanized framework that can recognize in danger clients from their public web-based media action. All the more explicitly, we recognize in danger clients from Twitter. To accomplish this objective we prepared a client level classifier utilizing Support Vector Machine (SVM) that can recognize in danger clients with a review of 0.8750 also, an accuracy of 0.7778. They additionally prepared a tweet-level classifier that predicts if a tweet shows trouble. These papers were based on depression detection and prediction level based on Social Media data where the feasibility of social media based assessment of mental illness creates numerous ethical questions. And all participants don't share what is happening through post. [21]

### **2.2.3 Related work as a recommendation system**

Shiqi Yang, Ping Zhou, Kui Duan. September worked on "emhealth: Towards Emotion Health Through Depression Prediction and Intelligent Health Recommender System". This project paper analyses the feelings of a person by a mobile app survey and based on that recommend health care or other medications to improve the person's condition. But there was not enough recommend contents and any level prediction, so most of the time it was difficult to give proper therapy.[22]

The fundamental objective of this framework is to get the suitable determination of infection and the right treatment and give the fitting strategy for treatment through a few hints that worry the sickness and how to treat it. In this paper the plan of the proposed Expert System which was created to help Psychologist in diagnosing depression disease. The proposed master framework is represented

considerable authority in the analysis just wretchedness with the accompanying side effects: a loss of energy, an adjustment of hunger, sleeping more or less, anxiety, reduced focus, uncertainty, sensations of uselessness, guilt or hopelessness and thoughts of self-harm or suicide.[23]

#### **2.2.4 Related work based on generating rules**

This paper builds up a standard based technique to recognize a bunch of hazard prescient examples from individual level longitudinal sickness estimations by coordinating the information change, rule revelation and rule assessment. They further broaden the identified rules to make rule-based observing procedures to adaptively screen people with different illness severities. They applied the standard put together technique with respect to an electronic wellbeing record (EHR) dataset of sorrow treatment populace containing individual level longitudinal Patient Health Questionnaire (PHQ)- 9 scores for evaluating sorrow seriousness. 12 danger prescient guidelines are identified, and the standard based prognostic model dependent on identified rules empowers more exact expectation of illness seriousness than other prognostic models including RuleFit, calculated relapse and Support Vector Machine. Two rule based observing methodologies beat the most recent PHQ-9 based checking procedure by giving higher affectability and specificity. [24]

### **2.3 Conclusion**

In this chapter we have discussed previously developed method of depression prediction system. We have come to know about various symptoms of depression, approaches to predict depression and generating rules based on features to predict depression. We have also found out some method's advantage and flaws.

# Chapter 3

## Methodology

### 3.1 Introduction

This chapter presents the proposed methodology for predicting depression level and explains in details with its constraints. The data collection and development procedure also presented in this chapter. We also go through the data pre-processing, features selection and depression prediction using fuzzy rules. The detailed description will be included here.

### 3.2 Diagram/Overview of Framework

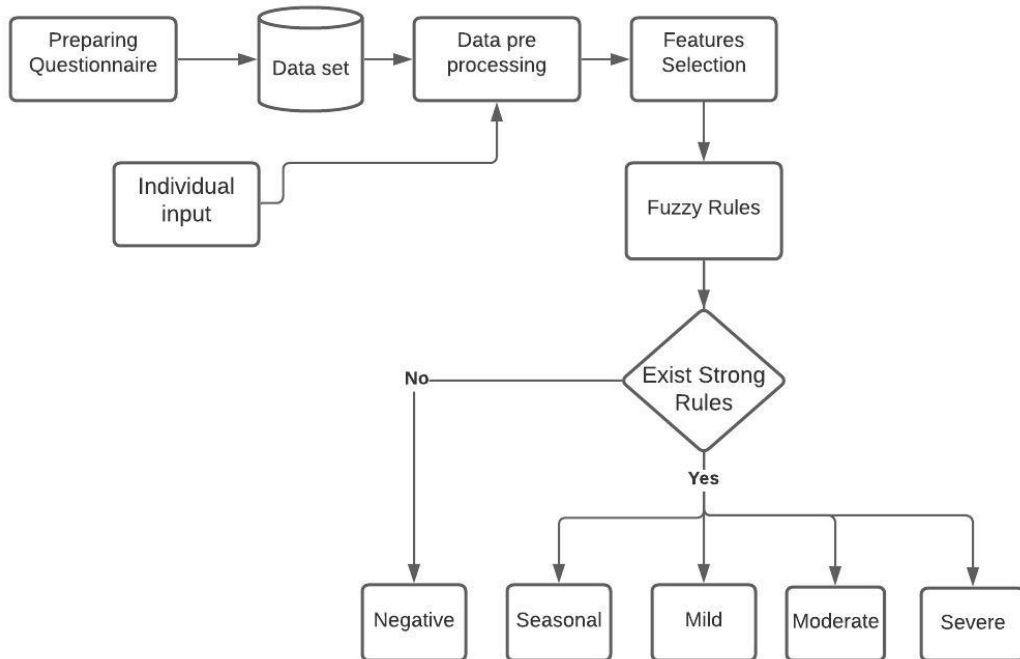


Figure 3.1: The proposed system of Depression Level Prediction



Figure 3.1 provides a complete walk-through of the entire study. First data collection was necessary to obtain a usable data-set. Data collection was completed after into two major steps, 1. Preparing questionnaire and 2. Data collection. Questionnaire was prepared by reviewing literature and consultation with psychiatrists as shown in section 3.3.1.1. With this questionnaire data was collected using two different approaches as described in section 3.3.1.2. Collected data was pre-processed into usable dataset mentioned in section 3.3.2. Once the final dataset was available, redundant features were eliminated using the algorithm in table 3.1 as described in section 3.3.3. Finally, fuzzy rules were applied to predict depression level.

### **3.3 Detailed Explanation**

Different block mentioned in figure 3.1 are described in the subsections of this section. Data collection starts off the process, followed by feature selection and concludes with fuzzy rules.

#### **3.3.1 Data Collection**

Data was collected using two separate steps by preparing questionnaire and collection of data from respondents. This two steps are described below:

#### **3.3.2 Preparing Questionnaire**

Before building this model, relevant literature was reviewed and there were discussions with local psychiatrists working on mental health. The primary objective is to figure out the commonly involved causes responsible for the problem. Once the necessary literature was reviewed, some common causes are identified: Sleep disturbances, including insomnia, sleeping over time and cry over night, Frequent tiredness, fatigue or passing most of the time without moving or laying, feeling back pain or migraine, Difficulty in remembering problem, concentration losing and also losing interest in hobbies or sex, Change in appetite, losing weight or weight gain, Optimistic about individual dream , unsatisfied result in working

sector, negative thinking about future, Feeling broken, unworthy or burden, Relationship status, bonding with family, social relationship and feeling loneliness, Past traumas, feeling guilty, losing or death of loving one, Copying others or feeling low watching other people, losing self confidence, Losing of self-control, drug addiction, anger and anxiety disorder , dominated by sub-conscious mind, Thinking about death of yourself always or suicide, attempt to kill yourself etc. These causes can vary geographically due to differences in social structure and verified with the help of local psychiatrists working on this issue. The questionnaire is prepared to find necessary information from an individual about the causes identified as key reasons behind depression among people of all ages. Thirty Multiple choice questions (MCQ) formulated the questionnaire assessing reasons, local situations, and other aspects. Questions focus on family relationships, career, financial situation, social condition and individual's mental health. All of these questions are designed in a sophisticated way to extract the information needed to predict a classification model. Once the draft questionnaire has been prepared, it is further examined by specialized psychiatrists. Corrections have been made by making few additions and exclusions to the draft. To make sure this questionnaire is actually capable of obtaining the appropriate information we need for our study, we have performed a pretesting among 20 participants. Once the pretesting has been completed, the questionnaire was ameliorated for the last trial to get the final questionnaire used in the study. Finally, the questionnaire was completed to be used for data collection.

### **3.3.3 Collection of respondents' data**

The key objective of data collection is to find answers to identical questions from different sectors and different aged people , the questionnaire was prepared in that manner. Responses have been collected through an online survey and in-person interviews and patients respectively. Different approaches are adopted for obvious reasons which are explained below. As for the patients suffering from depression, data have been collected through in-person interviews because of the social stigma and the difficulty in verifying or identifying an individual as a patient. It is inevitably difficult to get a voluntary response from people suffering from this

problem because of the negative attitude of society towards an individual known as a patient. Consequently, recognizing and collecting enough individual data required for this study voluntarily by SAQ (self-assessment questionnaire) or other means e.g. online survey from patients is implausible. Each interview takes approximately 20minutes to be conducted. Data have been collected from people through a google form delivered personally to the recipients by email or via social media from voluntary participants. Link to the form have been attached. The form contains prefatory statements describing the purpose of the study and asking for permission to let their information be used for the research and ethical law has been ensured.

### 3.4 Pre-processing of Dataset

What is your gender?	How old are you?	What is your occupation?	Do you have sleeping disturbances at night?	Addicted to social media?	Addicted to drug?	What's your relationship status?
Female	23.0	Student	yes	Yes	no	in a relationship
Female	23.0	Student	yes	Yes	no	in a relationship
Male	22.0	Student	yes	Yes	no	in a relationship
Female	24.0	Student	yes	Maybe	no	Single
Female	22.0	Student	yes	Yes	no	Single

Figure 3.2: Raw sample of Dataset

The spreadsheet constitutes each column in the data-set and the answers of every single question are replaced with categorical encoding into values between '0','1', '2', '3' or '4' depending on the question being binary or a 5 point-Liker scale question. Finally, a target variable titled "goal" is added to the data-sets to predict depression level( Negative, Seasonal, Mild, Moderate and Severe). Values within the column "goal" 5 point-Liker scale: '0' represents Negative , '1' represents Seasonal, '2' represents Mild, '3' represents Moderate and '4' represents Severe

are assigned to the variable ‘goal’ in the data-set. Finally, we have 30 variables in our data-set with 29 feature variables and one target variable. then the missing data was handled with average values within that column. The features of the dataset are mentioned in detail in section 1.1.

What is your gender?	How old are you?	What is your occupation?	Do you have sleeping disturbances at night?	Addicted to social media?	Addicted to drug?	What's your relationship status?
0	23.0	1	2	2	1	2
0	23.0	1	2	2	1	2
1	22.0	1	2	2	1	2
0	24.0	1	2	0	1	0
0	22.0	1	2	2	1	0

Figure 3.3: Pre-processed sample of Dataset

## 3.5 Features Selection

Features selection was performed based on two statistical method ( Pearson correlation method and R-squared method).These will be discussed in subsection 3.5.1 and 3.5.2.

### 3.5.1 Pearson Correlation method

Pearson’s relationship, which creates a score that can change from  $-1$  to  $+1$ . Two objects with a high score (close  $+1$ ) are profoundly similar. Two uncorrelated items would have a Pearson score almost zero. Two objects that connected contrarily would have a Pearson score close  $-1$ .

The Pearson connection for two objects, with matched traits, entries the result of their disparities from their item means, and partitions the entirety by the result of the squared contrasts from the object means. Formula for Pearson’s correlation:-

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Pearson's correlation is parametric, as in it depends intensely on the "signify" boundary for the two objects. This implies that Pearson's correlation may have higher legitimacy for an ordinary dispersion, with a focused mean, than for a dissemination that isn't ordinarily appropriated, like a Zipf distribution.

Python's Scipy module offers a Pearson work. As well as registering Pearson's connection, the Scipy work creates a two-followed p-esteem, which gives some sign of the probability that two absolutely uncorrelated objects may deliver a Pearson's relationship esteem as outrageous as the determined worth.

### 3.5.2 R-squared method

R-squared is an integrity of-fit measure for around the fitted regression line. This measurement indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. R-squared measures the strength of the connection between your model and the reliant variable on a helpful 0 – 100

R-squared assesses the spread of the information focuses around the fitted regression line. It is additionally called the coefficient of determination, or the coefficient of multiple determination for multiple regression. For a similar informational collection, higher R-squared qualities address more modest contrasts between the noticed information and the fitted qualities. R-squared is the percentage of the dependent variable variation that a linear model explains-

$$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}} \quad (3.1)$$

### 3.5.3 Selected Features

The name of the features and their description along with the statistical methods are described in the following:-

Correlation Methods	Features Name	Total Features
Pearson correlation method	Do you have sleeping disturbances at night? Addicted to drug? What's your relationship status? Your Economical status? Do you have past traumas? Your Health? Can you eat when you are sad? Do you feel yourself a burden? Do you have a lot of pressure on your working sector? Do you have self-control? When do you often feel bad ? Is your subconscious mind dominating you? Do you ever attempt suicide?	15
R-squared method	Do you have sleeping disturbances at night? Addicted to social media? What's your relationship status? Pass most of the time laying on bed without any body movements? Your Family bonding? Your Economical status? Social relationship? Do you feel yourself funworthy? Do you have past traumas? Do you cry overnight? Do you feel lonely? Your Health? Can you eat when you are sad? Do you feel guilt in your life? Do you feel yourself a burden? Do you have a lot of pressure on your working sector? Do you have self-control? When do you often feel bad ? Is your subconscious mind dominating you? Satisfied Result in your working sector?	20

### 3.6 Fuzzy Logic System

Fuzzy Logic (FL) is a strategy for thinking that looks like human thinking. The methodology of FL mimics the method of dynamic in people that includes all middle of the road prospects between advanced qualities YES and NO.

The traditional rationale block that a PC can comprehend takes exact info and produces an unmistakable yield as TRUE or FALSE, which is comparable to human's YES or NO.

The creator of fuzzy rules, Lotfi Zadeh, saw that dissimilar to PCs, the human

dynamic incorporates a scope of potential outcomes among YES and NO such as - CERTAINLY YES POSSIBLY YES CANNOT SAY POSSIBLY NO CERTAINLY NO The fuzzy rules deals with the degrees of conceivable outcomes of contribution to accomplish the clear yield. It has four primary parts as displayed

**Fuzzification Module** It changes the framework inputs, which are fresh numbers, into fuzzy sets. It parts the information signal into five stages, for example,

- LP x is Large Positive
- MP x is Medium Positive
- S x is Small
- MN x is Medium Negative
- LN x is Large Negative

**Information Base** It stores IF-THEN standards given by specialists.

**Surmising Engine** It recreates the human thinking measure by making fluffy deduction on the data sources and IF-THEN guidelines.

**Defuzzification Module** It changes the fluffy set acquired by the surmising motor into a fresh worth.

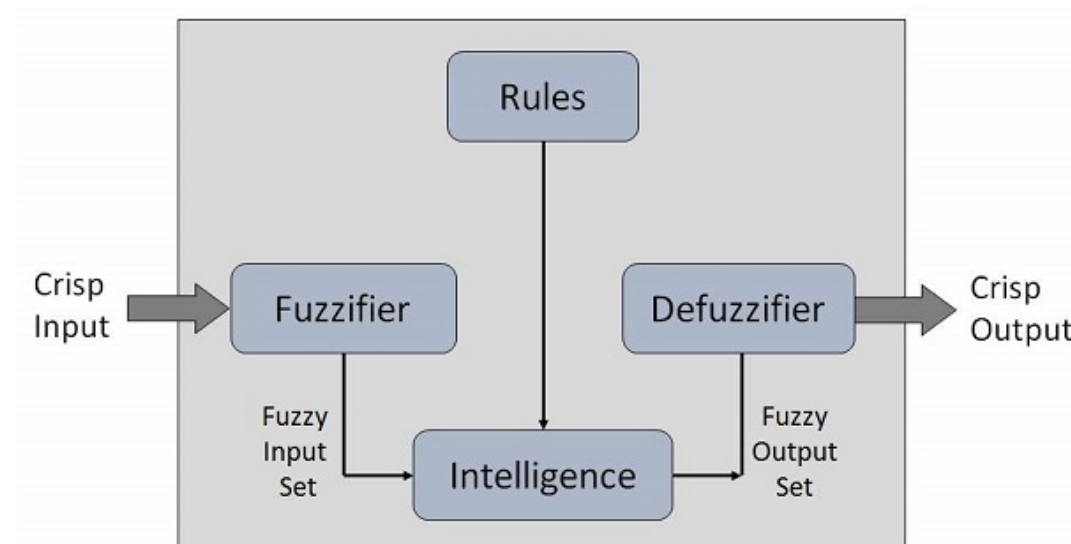


Figure 3.4: Overview of Fuzzy Logic System

### 3.6.1 Algorithm

**begin**

1. Define linguistic Variables and terms
2. Construct membership functions for them
3. Construct knowledge base of rules
4. Convert crisp data into fuzzy sets using membership functions (fuzzification)
5. Evaluate rules in the rule base
6. Combine results from each rule
7. Convert output data into non-fuzzy values (defuzzification)

**end**

The membership functions work on fuzzy sets of variables.

## 3.7 Membership Function

Membership Functions permit you to measure etymological term and address a fuzzy set graphically. An enrollment work for a fuzzy set An on the universe of talk X is characterized as  $A:X \rightarrow [0,1]$ .

Here, every component of X is planned to a worth somewhere in the range of 0 and 1. It is called participation worth or level of enrollment. It measures the level of enrollment of the component in X to the fuzzy set A.

Mathematically, triangular membership function is defined as below:

$$f(x,a,b,c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (3.2)$$



It can also be defined in the other way as:

$$f(x,a,b,c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (3.3)$$

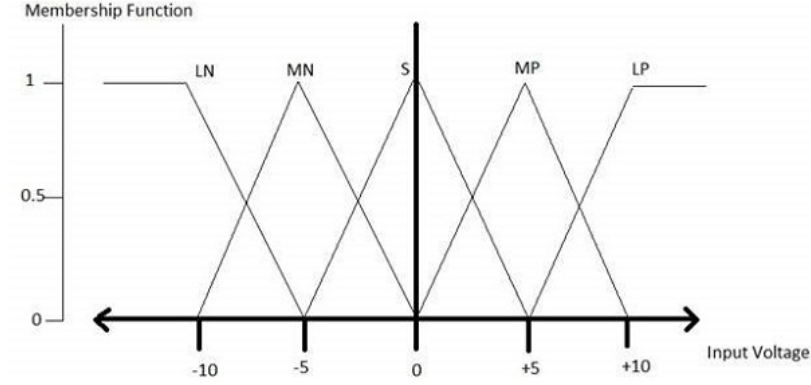


Figure 3.5: Overview of Membership Function

### 3.8 Fuzzy Rules

After defining membership functions we have generated fuzzy rules for predicting depression level. Some of them are showing below-

- If ['Do you have sleeping disturbances at night?'] is ['poor'] then ["Depression type"] is ['poor'] shown in figure 3.6

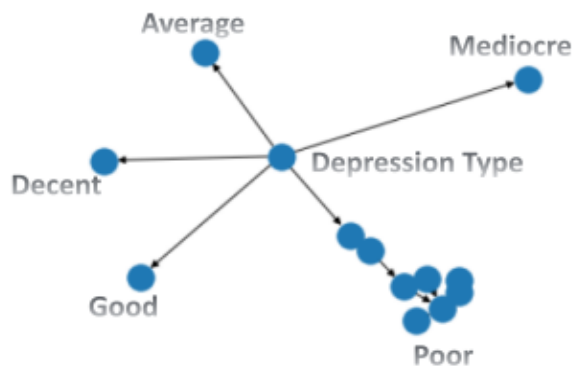


Figure 3.6: Fuzzy Inference rule generation.

- If ['Do you have sleeping disturbances at night?'] is ['good'] then ["Depression type"] is ['good']) shown in figure 3.7

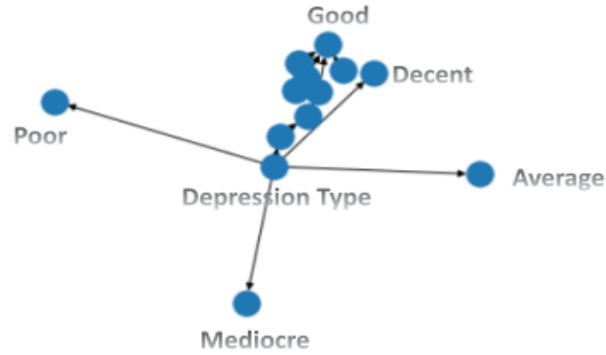


Figure 3.7: Fuzzy Inference rule generation.

- If ['Your Family bonding?'] is ['poor'] then ["Depression type"] is ['good']) shown in figure 3.7

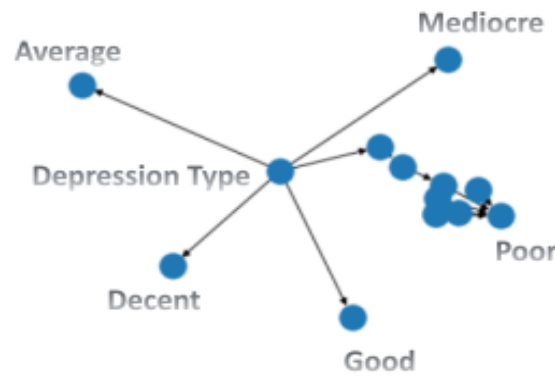


Figure 3.8: Fuzzy Inference rule generation.

### 3.8.1 Sample Input

Fuzzy inference mechanism is the fuzzy logic thinking measure that decides the yields relating to fuzzified inputs. The fuzzy rule-base is formed by IF-THEN rules like

- IF Economic Status is poor
- AND Past Traumas is good
- AND Satisfied Result at working sector is poor
- AND Subconscious mind dominating is average

- THEN Depression Type is Severe

Each fuzzy rule characterizes a fuzzy ramifications among condition and end rule parts. Utilizing fuzzy sets, the conduct about the article can be addressed as the type of fuzzy relations. These relations are made out of fuzzy articulations that are associated by fuzzy consistent administrators. Three significant consistent administrators are usually applied in a fuzzy connection: Intersection (AND), Union (OR) and Complement (NOT).

```

classify.input["How old are you?"] = 0.44444
classify.input["Do you have sleeping disturbances at night?"] = 0
classify.input["Addicted to drug?"] = 0
classify.input["What's your relationship status?"] = .6
classify.input["Your Economical status?"] = 0.5
classify.input["Do you have past traumas?"] = 0.3
classify.input["Your Health?"] = 1
classify.input["Can you eat when you are sad?"] = 0
classify.input["Do you feel yourself a burden?"] = 0
classify.input["Do you have a lot of pressure on your working sector?"] = 0.5
classify.input["Do you have self-control?"] = 0.5
classify.input["When do you often feel bad ? "] = 0.3333
classify.input["Is your subconscious mind dominating you?"] = 0.666667
classify.input["Do you ever attempt suicide? "] = 0.5
classify.input["Satisfied Result in your working sector?"] = 0

classify.compute()

```

Figure 3.9: Sample Input

### 3.8.2 Sample Output

Now output for the following inputs have been shown below-

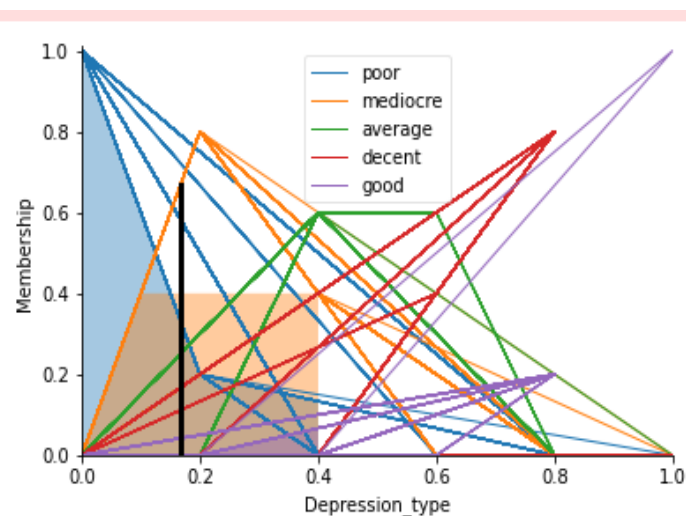


Figure 3.10: Sample Output

### 3.9 Conclusion

The sections above delineate the complete list of activities that was performed during this study. The list begin with questionnaire preparation and ends with prediction of depression level. The steps were planned in accordance with the introduction and literature review sections. The performance and efficacy of these methods adopted are analysed in detail in the following section. This process was adopted with a goal to creating a real-time data set, generating fuzzy rules and finally to identify depression leve of individual user.

# Chapter 4

## Results and Discussions

### 4.1 Introduction

Chapter 4 explains the details experimental analysis on developed data set for our approach. This Chapter also investigates the performance of implemented models. The details comparative analysis among various methods with existing techniques also illustrated in this Chapter.

### 4.2 Dataset Description

After reviewing literature and discussions with local psychiatrists to know about the local aspects of substance abuse, the questionnaire is formed based on the symptoms identified. Then we apply pretesting to finalize a total of 36 questions in the questionnaire. These 30 questions are MCQ (Multiple choice questions) with two or more answers are given below:

1. What is your gender?
2. How old are you?
3. What is your occupation?
4. How many hours do you manage to sleep?
5. Do you have sleeping disturbances at night?
6. Addicted to social media?
7. Addicted to drug?
8. What's your relationship status?

9. Pass most of the time laying on bed without any body movements?
10. Your Family bonding?
11. Your Economical status?
12. Nature of you?
13. Social relationship?
14. Do you feel yourself unworthy?
15. Do you have past traumas?
16. Do you cry overnight?
17. .Do you feel lonely?
18. Your Health?
19. Can you eat when you are sad?
20. Do you feel guilt in your life?
21. Do you feel concentration losing or remembering problem?
22. .Do you feel yourself a burden?
23. .Do you have a lot of pressure on your working sector?
24. Do you have self-control?
25. .When do you often feel bad ?
26. Is your subconscious mind dominating you?
27. Are you depressed?
28. Do you ever attempt suicide?
29. Satisfied Result in your working sector?
30. Depression type?

## **4.3 Impact Analysis**

As specified earlier this study was performed help prevention and to facilitate treatment of depressed people and also identifying depressed people. The study was designed to achieve a high degree of social and environmental impact. The impacts are described in the following sections.

### **4.3.1 Social Impact**

This research was performed to address a social issue. Depression and mental health problem has deep underlying social consequences. Left untreated this problem can cause severe trouble to loss a people of personal or family life and the members within a society. Social impacts of this study can be summarized as:

1. If more depressed people can be identified early by applying this research, it will be really helpful for the vulnerable individual as well as the society.
2. The depression level identified in this study can help to assess the condition of people within the society members and hence actions can be taken to get rid of this problem.

### **4.3.2 Ethical Impact**

Mental health problem is a problem that is interrelated to several ethical concepts. While in an imbalanced mental state as a consequence of depression, ethical qualities of a human can be compromised. The ethical impacts of this study can be summarized as:

1. This study itself is performed as part of an ethical responsibility: looking after the depressed people of the community. Thus it has greater ethical consequences.
2. While in an intoxicated state, studies have found people lose their ethical consciousness and valuable life by suicide like by hanging themselves or taking poisonous substance. These study aims at tackling the issue which

in turn can help solve this dangerous problem by acting upon the depressed individuals.

## 4.4 Evaluation of Framework

We have evaluated our proposed method under different environmental situations and illuminations. To evaluate the efficiency of our proposed system, we calculate the accuracy with the following equation.

$$Accuracy = \frac{No.ofcorrectedscenarios}{No.ofscenarios} * 100\% \quad (4.1)$$

For performance evaluation,from 50 participants data we got accurate results for 47 participants.So,

$$Accuracy = \frac{137}{140} * 100 = 95.71\% \quad (4.2)$$

## 4.5 Evaluation of Performance

We have compared the proposed mechanism to existing approaches that have been developed. In Table 4.1, we shows this comparison with the existing approaches. Our comparison is based on accuracy rate.

Table 4.1: Performance Evaluation

Research Name	Accuracy
SNS Based Predictive Model for Depression [25]	70
Predicting Depression via Social Media[26]	70
Predicting Depression Levels Using Social Media Posts[23]	63.3
Proposed System	95.71

The performance of the proposed method is to predict depression level by using fuzzy logic. The proposed method in [25] only depression pattern is identified. In [26]depression behaviour identification is done. In both proposed method there were no detection of depression type . From the comparison in conclusion we can say that, our proposed method is better than other existing frame work



## 4.6 Conclusion

This chapter shows the performance evaluation result of proposed predicting depression level based on human activities and feelings using fuzzy rules. Analyzing the performance we see that, our proposed methodology gives better accuracy of predicting depression level than other existing ones. In the next chapter, the conclusion is drawn on this thesis work.

# Chapter 5

## Conclusion

### 5.1 Conclusion

The basis of this study was predicting depression level based on human activities and feelings have been discussed in previous sections. Necessary information on this research culled from the related researches paved the way to figure out common symptoms of depression. Subsequently, to gain data from participants on these causes, the questionnaire was carefully designed safeguarding individual privacy and finalized after pre-testing. Responses from participants were converted into data-set after some pre-processing. Using the Pearson and R-squared statistics between feature variables and target variable, feature importance was calculated. Using only the 20 necessary features, out of the initial 30 helped acquire better accuracy. Participants were respondents aged between 18-60+ yrs. (largely youths) which provided a better opportunity to understand behavioral traits as similar age groups usually display similar patterns of behavior. Local impacts that occur uniquely to Bangladesh were taken into consideration to measure risk factors in a feasible way to improve the effectiveness of the model. Although the performance of the prediction model is slightly down in case of the class addicted, this shortcoming can be overcome by collecting more data from patients suffering from depression. As far as the overall performance is concerned, a good accuracy score was achieved because we prepared the questionnaire according to the information we needed and pretested it before finalizing leaving no scope for unnecessary questions to stay in the questionnaire. Data-set was created from the information gathered from the questionnaire resulting in an effective data-set. This study was performed among the people of Bangladesh which does not represent the whole scenario of depression in the country. Causes might vary for

changes in the target population in both home and abroad. An increase in the size of the data-set will reveal some more valuable insights which will help to predict the depression with a higher degree of accuracy. Due to the COVID-19 outbreak in midst of the study, collecting data from rehabilitation centers became increasingly difficult. A larger data-set can be collected in the future which will produce an even better accuracy in the addicted class to create a predictive model with even better performance and also incorporate an even larger population. It has already been addressed again and again, how serious and damaging this problem is to Bangladesh and all over the world. Sadly, the number of people suffering from Major Depressive disorder is increasing globally which is a portent for larger problems ahead. This study focuses on prevention by creating a model to assess depression level which can be a little step forward towards addressing this global problem. Finally, from this study, it can be said that data science has an important role to play in understanding these intricate behavioral traits of human beings to address important psychological issues. Health officials can use the huge potential of data to build a safer and better future.

## 5.2 Future Work

This study leaves a wide scope for future works to be done. Data in this research was maximum collected from urban youths and students of Bangladesh, with better resources and more time, the focused range of the population in this study can be increased which will enable this model to predict within a larger community. A similar process can be adopted in different countries where symptoms may vary a little. An automated counseling chat-bot online can be developed with the help of psychiatrists to suggest participants further actions according to the depression predicted from the prediction model. Different people can suffer from different reasons like family problems, economic crisis, etc. Once the specific root reasons can be identified it can be treated with better precision. In the future, we intend to collect more data in the future and construct a rule-based model using our rule discovery process. Furthermore, performing a user survey to assess the method's usability at the application level could be a future project.

# References

- [1] S. Alayón, R. Robertson, S. K. Warfield and J. Ruiz-Alzola, ‘A fuzzy system for helping medical diagnosis of malformations of cortical development,’ *Journal of biomedical informatics*, vol. 40, no. 3, pp. 221–235, 2007 (cit. on p. 3).
- [2] X. Chen, Y. Wang, S. Zhen, K. Huang, H. Zhao and Y.-H. Chen, ‘Robust control design of uncertain mechanical systems: A fuzzy approach,’ *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 24, no. 02, pp. 307–324, 2016 (cit. on p. 3).
- [3] S. Saeed and A. Niknafs, ‘Artificial bee colony-fuzzy q learning for reinforcement fuzzy control (truck backer-upper control problem),’ *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 24, no. 01, pp. 123–136, 2016 (cit. on p. 3).
- [4] C.-C. Lee, ‘Fuzzy logic in control systems: Fuzzy logic controller. i,’ *IEEE Transactions on systems, man, and cybernetics*, vol. 20, no. 2, pp. 404–418, 1990 (cit. on p. 3).
- [5] D. G. Stavrakoudis and J. B. Theocharis, ‘Handling highly-dimensional classification tasks with hierarchical genetic fuzzy rule-based classifiers,’ *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, no. supp02, pp. 73–104, 2012 (cit. on p. 3).
- [6] B. Bahrami and M. Shafiee, ‘Fuzzy descriptor models for earthquake time prediction using seismic time series,’ *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 23, no. 04, pp. 505–519, 2015 (cit. on p. 3).
- [7] K. M. Tay and C. P. Lim, ‘On monotonic sufficient conditions of fuzzy inference systems and their applications,’ *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 19, no. 05, pp. 731–757, 2011 (cit. on p. 3).
- [8] N. Kubota and A. Yaguchi, ‘Decision making of robot partners based on fuzzy control and boltzmann selection,’ *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 03, pp. 529–545, 2011 (cit. on p. 3).
- [9] J. Martinez-Moreno and L. Martinez, ‘A discrete time variable index for supporting dynamic multi-criteria decision making,’ *KNOWLEDGE-BASED SYSTEMS*, 2014 (cit. on p. 3).
- [10] H. Bustince, E. Barrenechea and M. Pagola, ‘Image thresholding using restricted equivalence functions and maximizing the measures of similarity,’ *Fuzzy Sets and Systems*, vol. 158, no. 5, pp. 496–516, 2007 (cit. on p. 3).

- [11] V. Khatibi and G. A. Montazer, ‘Intuitionistic fuzzy set vs. fuzzy set application in medical pattern recognition,’ *Artificial Intelligence in Medicine*, vol. 47, no. 1, pp. 43–52, 2009 (cit. on p. 3).
- [12] A. Adeli and M. Neshat, ‘A fuzzy expert system for heart disease diagnosis,’ in *Proceedings of international multi conference of engineers and computer scientists, Hong Kong*, Citeseer, vol. 1, 2010, pp. 28–30 (cit. on p. 3).
- [13] E. Ephzibah and V. Sundarapandian, ‘A neuro fuzzy expert system for heart disease diagnosis,’ *Computer Science & Engineering*, vol. 2, no. 1, p. 17, 2012 (cit. on p. 3).
- [14] R. Arunprakash, R. Bhavadharini, K. V. Kumar and T. Jayasankar, ‘Intelligent heart disease detection and classification method using optimal neuro-fuzzy with stochastic gradient descent model for smart societies,’ *Artificial Intelligence Applications for Smart Societies: Recent Advances*, p. 179, (cit. on p. 3).
- [15] K. Wisaeng, ‘Predict the diagnosis of heart disease using feature selection and k-nearest neighbor algorithm,’ *Applied Mathematical Sciences*, vol. 8, no. 83, pp. 4103–4113, 2014 (cit. on p. 3).
- [16] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar and J. C. Eichstaedt, ‘Detecting depression and mental illness on social media: An integrative review,’ *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017 (cit. on p. 10).
- [17] J. Mikal, S. Hurst and M. Conway, ‘Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: A qualitative study,’ in *Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality*, 2017, pp. 41–47 (cit. on p. 10).
- [18] M. M. Aldarwish and H. F. Ahmad, ‘Predicting depression levels using social media posts,’ in *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)*, IEEE, 2017, pp. 277–280 (cit. on p. 11).
- [19] B. O’dea, S. Wan, P. J. Batterham, A. L. Callear, C. Paris and H. Christensen, ‘Detecting suicidality on twitter,’ *Internet Interventions*, vol. 2, no. 2, pp. 183–188, 2015 (cit. on p. 11).
- [20] H. Almeida, A. Briand and M.-J. Meurs, ‘Detecting early risk of depression from social media user-generated content.,’ in *CLEF (Working Notes)*, 2017 (cit. on p. 12).
- [21] Z. Jamil, ‘Monitoring tweets for depression to detect at-risk users,’ Ph.D. dissertation, Université d’Ottawa/University of Ottawa, 2017 (cit. on p. 12).
- [22] S. Yang, P. Zhou, K. Duan, M. S. Hossain and M. F. Alhamid, ‘Em-health: Towards emotion health through depression prediction and intelligent health recommender system,’ *Mobile Networks and Applications*, vol. 23, no. 2, pp. 216–226, 2018 (cit. on p. 12).

- [23] I. A. Alshawwa, M. Elkahlout, H. Q. El-Mashharawi and S. S. Abu-Naser, ‘An expert system for depression diagnosis,’ 2019 (cit. on pp. 13, 30).
- [24] Y. Lin, S. Huang, G. E. Simon and S. Liu, ‘Data-based decision rules to personalize depression follow-up,’ *Scientific reports*, vol. 8, no. 1, pp. 1–8, 2018 (cit. on p. 13).
- [25] J. Hussain, M. Ali, H. S. M. Bilal, M. Afzal, H. F. Ahmad, O. Banos and S. Lee, ‘Sns based predictive model for depression,’ in *International Conference on Smart Homes and Health Telematics*, Springer, 2015, pp. 349–354 (cit. on p. 30).
- [26] M. De Choudhury, M. Gamon, S. Counts and E. Horvitz, ‘Predicting depression via social media,’ in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, 2013 (cit. on p. 30).