A MINOR PROJECT REPORT

ON

"DETECTION OF PSYCHIATRIC DISORDER USING MACHINE LEARNING"

A project report submitted in partial fulfillment for the award of the degree of

Bachelor of Technology

In

Computer Science & Engineering

Submitted by:

FAZLUL KARIM CHOUDHURY (19RISTCS003)

IMDADUL HUSSAIN (19RISTCS004)

MD. ANDALEEB LASKAR DURRANI (19RISTCS008)

MD. SHAKIR AHMED (19RISTCS009)

Under The Guidance Of:

Mr. Karjan Basumatary

Head of Department

Department of Computer Science & Engineering, RIST

REGIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Approved by AICTE, New Delhi, Ministry of HRD, Govt. of India

Affiliated to North Eastern Hills University (NEHU), Shillong,

Techno city, Killing Road, Baridua, 9th mile, Ri-Bhoi, Meghalaya-793101

Academic year 2022

REGIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Run By: ERD Foundation-Guwahati

Approved by: AICTE, New Delhi & Affiliated to: NEHU Shillong

CERTIFICATE

This is to certify that this project entitled **"DETECTION OF PSYCHIATRIC DISORDER USING MACHINE LEARNING"** Submitted by **Fazlul Karim Choudhury** (19RISTCS003) **Imdadul Hussain** (19RISTCS004) **Md. Andaleeb Laskar Durrani** (19RISTCS008) **Md. Shakir Ahmed** (19RISTCS009), students of 7th Semester of Computer Science and Engineering Department, Regional Institute of Science and Technology,9th Mile, Baridua, for partial fulfillment the requirement for the award of the Bachelor of Technology (Department of Computer Science& Engineering)Degree of NEHU, Shillong, is a record of student's own study carried out under my supervision and guidance.

Project guide:	Head of Dept.	Project Co-ordinator:
Mr. Karjan Basumatary	Mr. K. Basumatary	Dr. Rashel Sarkar
Head of Department	Dept. of CSE	Asst. Professor
Dept. of CSE		Dept. of CSE

Examination on:

Internal Examiner: External Examiner:

Techno City, Killing Road, Baridua, GS Road, 9th Mile, Ri-Bhoi, Meghalaya-793101

PH: 0361-289507(fax)/097060-91836, E-mail: rist09@redifmail.com, website: www.rist.ac.in

DECLARATION

We hereby declared that the work entitled **"DETECTION OF PSYCHIATRIC DISORDER USING MACHINE LEARNING"** is submitted in partial fulfillment of the requirement for the award of the degree of B. TECH in Computer Science & Engineering, North Eastern Hills University, Shillong, is a record of our own work carried out by us during the academic year 2022 under the supervision and guidance Of Mr. Karjan Basumatary, Head of Department, Department of Computer Science and Engineering, Regional Institute of Science & Technology.

(Signature of Candidate)
FAZLUL KARIM CHOUDHURY
(19RISTCS003)

(Signature of Candidate)
IMDADUL HUSSAIN
(19RISTCS004)

(Signature of Candidate)
MD. ANDALEEB LASKAR DURRANI
(19RISTCS008)

(Signature of Candidate)
MD. SHAKIR AHMED
(19RISTCS009)

I certify that the declaration by the candidate is true

(Signature of Guide)
Mr. Karjan Basumatary
HOD, CSE, RIST

ACKNOWLEDGEMENT

A large encouragement and valuable guidance from many sides has been a constant source of inspiration for us to complete the project. By doing this project we learned many things to develop a project and it allowed and taught us to interact with real life situations. The theoretical knowledge is sometimes not sufficient, we need some practical knowledge, so this project allowed us to get lots of practical knowledge to design and develop a new project.

The Completion of this project is possible only because of the guidance and advice and considerate care of all our teachers. They let us understand the concept and the ways to interpret it in real life situation. A vast number of suggestions were given in order to let us complete this project successfully. It has been a pleasure working on the subject. We have benefitted ourselves a lot from the project in our future point of view.

Also, we also want to give special thanks to all our friends for their help and continuous unedited support in completing this project on time.

(Signature of Candidate)
FAZLUL KARIM CHOUDHURY
(19RISTCS003)

(Signature of Candidate)
IMDADUL HUSSAIN
(19RISTCS004)

(Signature of Candidate)
MD. ANDALEEB LASKAR DURRANI
(19RISTCS008)

(Signature of Candidate)
MD. SHAKIR AHMED
(19RISTCS009)

ABSTRACT

A psychiatric disorder is a mental illness that greatly disturbs our thinking, moods, and/or behavior and seriously increases our risk of disability, pain, death, or loss of freedom. A person diagnosed with a psychological disorder encounters many difficulties considering in a social aspect or individual perspective. Nowadays recent technologies made this possible to connect everything that exists around us on earth through the internet, this is what machine learning made possible which can find patterns automatically and reason about the data.

The early and accurate diagnosis of such disorders is crucial for recovery and improvement. Machine Learning can be implicated in these medical and clinical fields to bring efficient results and solutions.

We aim to use machine learning models into psychiatry to predict various psychiatric disorder so as to establish more accurate diagnosis and better mental health.

These technological solutions have great potential in improving the quality of life of an psychiatric patient as this helps to reduce pressure on healthcare and to maximize the operational cost.

Keywords : Psychiatric disorder, Mental disorder, Random Forest, ML, AI .

LIST OF CONTENTS

Content Name **Page No.**

LIST OF FIGURES

FIGURE NAME **PAGE NO.**

LIST OF TABLES

TABLE NAME PAGE NO:

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

A psychiatric disorder is a mental illness that greatly disturbs our thinking, moods, and/or behavior and seriously increases our risk of disability, pain, death, or loss of freedom.

The early and accurate diagnosis of such disorders is crucial for recovery and improvement. Machine Learning can be implicated in these medical and clinical fields to bring efficient results and solutions.

According to research around 450 million people have psychiatric disorders today, with 43.8 million cases every year. The number of people affected will get doubled every year to 43.1 million. The increasing rate of psychiatric disorder patients is not uniform for all over the world. Despite its increasing risk in the society, its detection and diagnosis are still not up to the mark that is needed or expected. People are still unable to get proper diagnosis that would have relieved them of major issues later on or required cure if detected early on.

We aim to use machine learning models into psychiatry to predict various psychiatric disorders so as to establish more accurate diagnosis and better mental health.

1.2 PROBLEM STATEMENT

The analysis of psychiatric disorder can be defined and its description presented in terms of two important characteristics i.e., Psychophysical characteristic and Signal characteristic

- The Psychophysical characteristic further divided into: Physiological factor, cognitive factor and psychological factor.
- The signal characteristic further divided into two factors i.e., Electroencephalogram (EEG) and Magnetic resonance image (MRI).
- Physiological factor contains problem in sleeping, weakness problem in muscle, problem in movement, problem in locomotion.

- Cognitive factor contains problem in learning, problem in reasoning, hyper activity problem.
- Psychological factor contains delusion problem, problem in frontal, problem in occipital.

1.3 MOTIVATION

One out of every four families has a family member with a serious psychiatric illness. Because such an illness is often disabling, it is no surprise that four out of the five major causes of disability world-wide for individuals between 15 and 45 years of age are neuropsychiatric disorders. The disability they create exceeds that caused by HIV, cancer, cardiovascular disease, and road accidents.

Motivated by such a strong impact, we decided to create efficient machine learning model to detect various psychiatric disorder based on personality, habit, character, behavior etc. of a subject. Our model can help people to understand and detect psychiatric disorder in early stages and maintain good mental health.

1.4 PROPOSED SYSTEM

In this project, we proposed to create a machine learning model that can detect Psychiatric Disorder based on character and behavior of a patient.

Our model will have dataset containing information like character, habit, personality etc. of patient along with the name of disorder.

Model will analyze the data using various algorithms and when a new case enters into the model, it matches with stored case and predicts the name of disorder depending on the symptoms.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

[1] George Karystianis, Aleo J.Nevado, Chi-Hun Kim, Azad Dehghan, Jhon A. Keane, Goran Nenadic. "Automatic mining of symptoms severity from psychiatric evaluation notes", Int J Methods Psychiatry Res. (2018)

- This paper gives an idea to find the best possible approaches to detect the severity of psychiatric disorder present in the patient by processing the free text.
- A free text is a text which is used to record the information in mental healthcare which also include few laboratories test(e.g., Measuring of blood pressure).
- Three automatic approaches are discussed in this paper. Rule based approach, Machine learning approach, Hybrid approach.
- Advantage:
    - Helps the patient to identify the severity of symptoms in the early stage.
- Disadvantages:
    - Here the training set used is very less.
    - Accuracy in rule-based approach 80.1%,in machine learning approach 73.3%, in hybrid approach 72.0%.

[2] Fani Deligianni,Yao Guo, Guang-Zhong Yang. "From Emotion to Mood Disorder: A Survey on Gait Analysis Methodology", Imperial College London (2019)

- This paper gives a brief study on the gait analysis caused by mood disorder. Here they review evidence that demonstrates the relationship between gait, emotion and mood disorder.
- They also observed how gait and body movement are affected due to mood disorder.

- They focus on clinical gaps and unmet needs along with technologies that can be used to address these problems .
- In aged people the disorder may cause loss of life or disability. In young age if this disorder is untreated then it can cause decrease in the quality of life.

[3] Yousra Bouaiachi , Mohamed Khaldi , Abdellah Azmani "Neural Network-based Decision Support System for Pre-diagnosis of Psychiatric Disorders", Third IEEE International CIST (2014)

- This paper shows a pre-diagnosis approach to simplify the modelling of a decision support system using neural networks.
- With artificial neural network as a decision tool case-based reasoning technique is applied.
- Back-propagation algorithm is applied.
- The proposed decision support system consists of 3 parts-knowledge base, user interface and a decision model (multi-layer perceptron).
- The efficiency of the pre-diagnosis neural network reaches the accuracy of 90%.
- ADVANTAGES:
    - Provide accurate results with minimal data outputs.
    - Training cycles are better optimized.

[4] Luciano Comin Nunes, Placido Rogerio Pinheiro, Mirian Caliope Dantas Pinheiro, Marum Simao Filho , Rafael Espindola Comin Nunes,and Pedro Gabriel Caliope Dantas Pinheiro "Automatic Detection and Diagnosis of Neurologic Diseases" IEEE (2019)

- This paper presents a model that facilitates the process of early diagnosis of various psychological disorders from the qualitative and comparative analysis of events and criteria , using Multicriteria Methodology.
- Hybrid approach is done combining multicriteria decision support methodology and an expert system generated by ExpertSINTA is presented.

- MACBETH method (measuring attractiveness by a categorized based evaluation technique) is used.
- DSM-5 book is used as a reference for classifying and identifying the disorders in the model.
- ADVANTAGES:
  - Facilitates the process of early diagnosis of various psychological disorders.
  - Much advancement needs to be done for the automation of decision-making processes involving multicriteria analysis.

[5] Vanishri Arun, Prajwal V, Murali Krishna, Arunkumar B. V, Padma S. K, Shyam V "A Boosted Machine Learning Approach for Detection of Depression" Symposium Series on Computational Intelligence SSCI (2018)

- A novel approach to detect depression using extreme Gradient Boosting (XGBoost) is presented in this paper.
- XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework.
- The training and learning of the dataset are faster in XGBoost model as compared to other predictive models.

COMPARISON OF RESULTS USING MYNAH COHORT

	ACCURACY (%)	TRAINING TIME IN SECONDS
XGBoost	97.80	0.059
AdaBoost	91.62	0.133
Random Forest	92.61	0.152
Decision Tree	89.54	0.166

[6] I. Kinder, K. Friganovic, J. Vukojevic, D. Mule, T. Slukan, D. Vidovic, P. Brecic, M. Cifrek "Comparison of Machine Learning Methods in Classification of Affective Disorders" MIPRO (2020)

- The goal of this paper is to process and analyze resting state EEGs of individuals with depressive disorders as well as healthy control subjects.
- 475 different features are extracted from each EEG and used in the evaluation of different binary classification methods.
- The support vector machine model shows the highest classification accuracy, followed by the K-nearest neighbor algorithm.

Method	Acc.	F1 score
SVM	**0.7667**	**0.7586**
KNN	**0.7667**	**0.7586**
Naïve Bayes	0.7000	0.7097
Linear Discriminant	0.7000	0.6400
Tree	0.5333	0.5625
Logistic Regression	0.5333	0.5333

[7] Shafeeq Ahmed, Preeti Singh, Prakash Singh. "Case Based Reasoning Model in the Diagnosis of Psychiatric Disorder" International conference on Communication and Electronics System (ICCES) (2016)

- In this research work, case-based reasoning technique applied in the diagnosis of psychiatric disorder.
- It consists of retrieval of cases that are already stored in the case base and nearest neighbor method applied for the matching of stored cases with the new case.
- The basic idea behind the case matching is that, similar types of cases produce similar type of results.

- Test Case Result :

Test cases	250
Correctly Predicted Cases	240
Wrongly Predicted	10
Accuracy	96.0%

[8] Nivedhitha Mahendran, P. M. Durai Raj Vincent, Kathiravan Shrinivasan, Vishal Sharma, Dushantha N. K. Jayakody. "Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults" IEEE (2020)

- This paper gives an idea of a Stacking Generalization Model for improving the accuracy in predicting Major depressive disorder (MDD).
- The stacking generalization is made by combining three low learners' Multilayer perceptron(MLP), support vector machine (SVM), and random forest(RF) and then averaging them into a Meta level learner.
- The classifiers are also implemented individually to compare the results. The accuracy of individual classifiers MLP, SVM, RF is 96.38%, 95.06%, and 96.90%, respectively.
- The accuracy of the stacking generalization model is 98.16% which show that the stacking generalization model has improved the prediction accuracy when compared with individual base learners.

2.2 SURVEY OUTCOME:

Sl.No.	Name of The Paper	Author and Year	Algorithm	Findings
1	Automatic mining of symptoms severity from psychiatric evaluation notes	G.Karystianis, A.J.Nevado, Chi-Hun, A.Dehghan, J.A. Keane, G.Nenadic (2018)	Rule based, ML Hybrid approach	Some approaches are mention which helps to find the severity of psychiatric disorder.
2	From Emotion to Mood Disorder: A Survey on Gait Analysis Methodology	Fani Deligianni,Yao Guo, Guang-Zhong Yang(2019)	Gait Analysis	Different mood disorder is addressed here which affect gait and body movement.
3	Neural Network-based Decision System for Pre-diagnosis of Psychiatric Disorders	Yousra Bouaiachi , Mohamed Khaldi , Abdellah Azmani (2014)	Back-propagation algorithm	The efficiency of the pre-diagnosis Neural N. reaches the accuracy of 90%.
4	Automatic Detection and Diagnosis of Neurologic Diseases	L.C.Nunes, P.R.Pinheiro, M.C.D.Pinheiro, M.S.Filho, R.Espindola C.Nunes, P.G.C.D.Pinheiro (2019)	MACBETH	Helps classify various psychological disorders through analysis of events and criteria
5	A Boosted Machine Learning Approach for Detection of Depression	V.Arun, Prajwal V, Murali Krishna, Arunkumar B. V, Padma S. K, Shyam V (2018)	XGBoost	Training and learning of the dataset are faster in XGBoost model
6	Comparison of ML Methods in Classification of Affective Disorders	I. Kinder, K. Friganovic, J. Vukojevic, D. Mule, T. Slukan, D. Vidovic, P. Brecic, M. Cifrek (2020)	KNN, SVM, Tree, Naïve Bayes, Regression	SVM model shows the highest classification accuracy, followed by the KNN algorithm
7	Case Based Reasoning Model in the Diagnosis of Psychiatric Disorder	Shafeeq Ahmed, Preeti Singh, Prakash Singh. (2016)	Case Based Model, KNN	Case based model using KNN algorithm has accuracy of 96%
8	Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults	Nivedhitha Mahendran, P.M.D.R Vincent, K.Shrinivasan, V.Sharma, Dushantha N. K. Jayakody (2020)	Stacking Gen. Model, MLP, SVM, RF	Stacking Gen. Model has improved accuracy than individual low-level learners like SVM, RF, MLP

Table no. 6.1: Survey outcome

CHAPTER 3

METHODOLOGY

3.1 FEASIBILITY STUDY

This paper emphasizes on the fact that the technique which is incorporated here to predict psychiatric disorder of the patient as well as enables caretakers and doctors to monitor mental condition of the patient. This paper is designed to develop an application which can predict psychiatric disorder of the patient by analyzing the character and behavior of the patients.

- **Requirement Gathering and Analysis-**All possible requirement of the system to be developed are captured in this phase and documented in a requirement specification document.
- **System Design-**The requirement specification is studied in this phase and the system designed is prepared. This system design helps in specifying in hardware and system requirement and helps in defending the overall system architecture.
- **Implementation-**With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to unit testing.
- **Integration and Testing-**All the unit developed in the implementation phase are integrated into a system after testing of each unit. Post integration are the entire system is tested for any fault or failure.
- **Deployment system-** Once the functional and non- functional testing is done the product is deployed in the customer environment or released into market.
- **Maintenance-**There are some issues which come up in the client environment. To fix those issues, patches are released. Also, to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

3.1.1 Minimum System Requirement

3.1.1.1 Minimum Software Requirement

- Operating system: windows 10(64 bit)
- Google colab

- Jupyter notebook
- Python 3
- Visual studio code

3.1.1.2 Minimum Hardware Requirement

- CPU (i5 intel)
- 8gb RAM
- 128gb SSD

3.2 METHODOLOGY

- **Data Collection:** In this step, data of different psychiatric disorder patients are collected for training and testing of model.
- **Pre-Processing:** Raw data collected is further processed to enhance performance of the model. Here the data is cleaned and transformed before feeding it to the algorithm.
- **Training Data:** The processed data is divided into training data which is used to train the model.
- **Testing Data:** The testing data is used to test the accuracy of the model.
- **Feature Extraction:** The data is transformed into numerical features that can be processed while preserving the information in the original data set.
- **Train using RF Model:** In this model data is trained using Random Forest (RF) Model.
- **Trained Model:** After training the model testing data is passed to check the accuracy of the model.
- **Deploy Model:** After checking the accuracy of the model, the model is then deployed and result is shown.

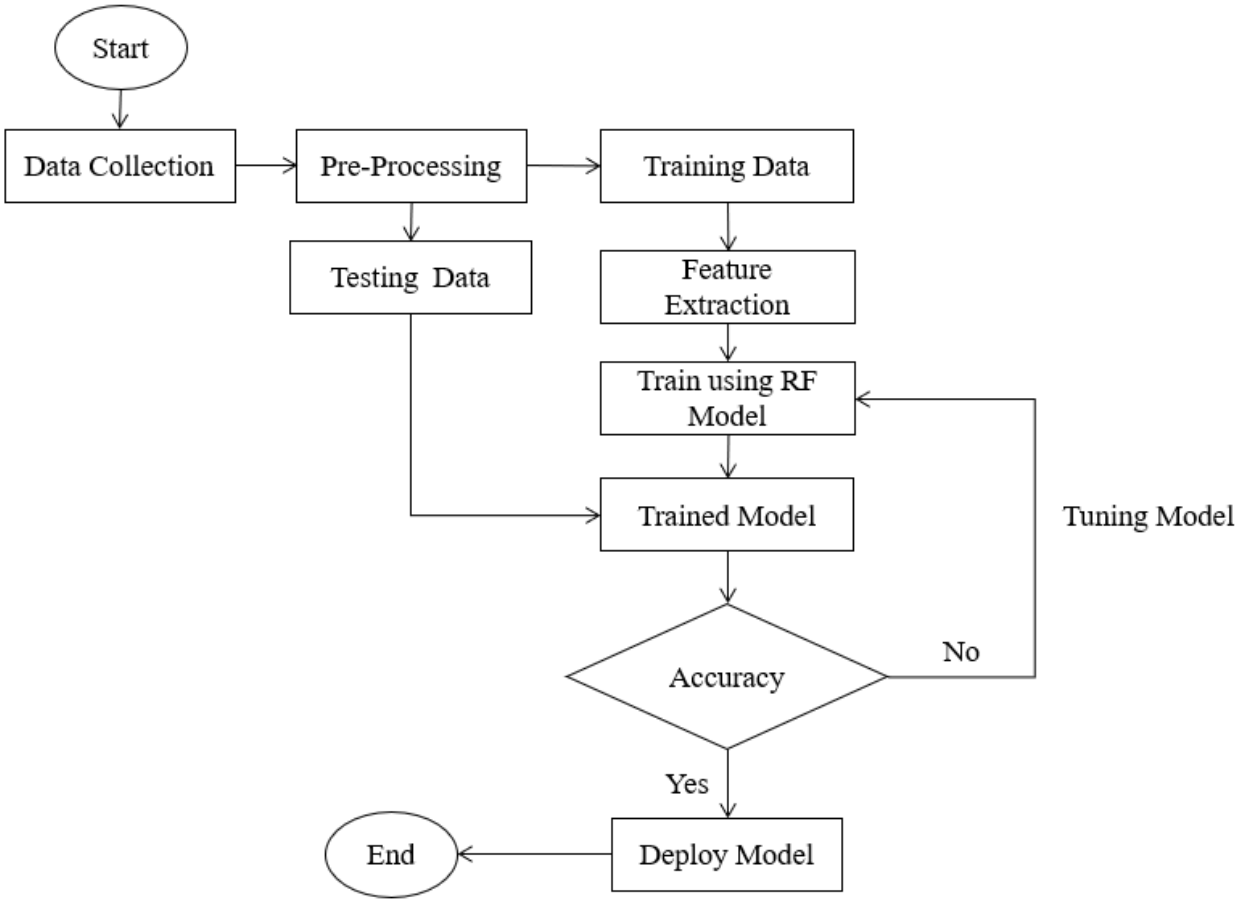


Fig 3.1: Flow Chart of PDD Model

CHAPTER 4

DESIGN

4.1 LOW LEVEL DESIGN

Low level- design fills in a portion of the gaps to give additional detail that is essential before designer can begin composing codes. It gives an increasingly explicit direction for how the pieces of the framework will work and how they will cooperate. It refines the meaning of the databases; the significant classes, and outer interfaces.

4.2 SDLC MODEL

Iterative Waterfall Model is used in our proposed system.

Iterative Waterfall Model

The Software Development Process model that we plan on using for the development of this project is the Iterative Waterfall model. This SDLC model is used as it provides the mechanism of error correction because there is a feedback path from one phase to its preceding phase. This model is simple to understand and use and also provides parallel development. Changing the plan or requirements in the model is very cost-effective.

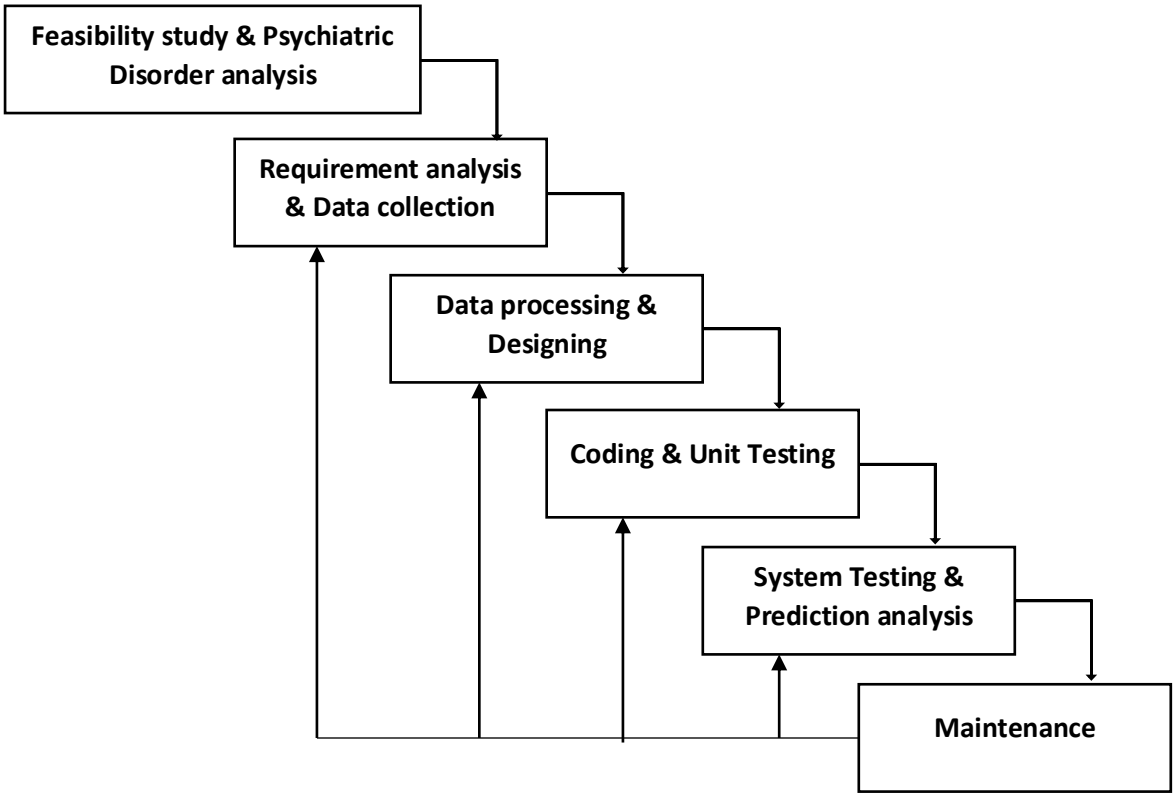


Fig 4.1: Iterative Waterfall Model

Stages of life cycle:

A typical Software Development Life Cycle consists of the following stages-

Requirement gathering & analysis: The business requirements are gathered during this phase of the iterative model. Then, an analyst determines whether they can be met within the financial constraints. This phase details the business needs, and system information (hardware or software) is acquired and assessed for viability. The kaggle dataset on Depression Anxiety Stress Scale Response Dataset was discovered and found to be exactly what was required. The dataset was collected and prepared accordingly. Exploratory analysis was conducted. Local machines were setup for training the ML-Ensemble and the Google Collab account was set and configured with appropriate libraries for the purpose of developing a straight forward interface, streamlit was choosen platform.

Design: During this phase of the iterative model, the project team receives the complete list of criteria for starting work in a specific direction. Then, they use various diagrams, like a data flow diagram, class diagram, activity diagram, state transition diagram, and so on, to gain explicit knowledge of the program design and to help them progress with development. The aforementioned information was then utilized as reference for the project to be developed. Upon exploratory analysis of the data, various lightweight neural network models were trained on the peak amplitude envelopes of the available dataset while a set of popular machine learning models were trained on the features extracted by random forest classifier.

Implementation: At this point in the project, according to the iterative model, the actual coding of the system begins. This stage will be influenced by the Design Stage's analysis and design. All needs, planning, and design plans have been carried out. The chosen design will be implemented by the developer using predefined coding and metrics standards.

Testing: After completing the coding phase, software testing starts using different test methods. There are many test methods, but the most common are white box, black box, and grey box test methods. This sort of testing includes performance testing, stress testing, security testing, requirements testing, usability testing, multi-site testing, disaster recovery testing, and so on. The final build was tested on the local machine.

Deployment: After completing all the phases, software is deployed to its work environment. Here the model is deployed in the form of web application using streamlit.

Review: In this phase, after the product deployment, review phase is performed to check the behavior and validity of the developed product. And if there are any errors found then the process starts again from the requirement gathering.

Maintenance: In the maintenance phase, after deployment of the software in the working environment there may be some bugs, some errors or new updates are required. Maintenance involves debugging and new addition options. The software will require continuous patching with suggested enhancements. As the project is for educational purpose, there is no further maintenance planned.

4.3 HIGH LEVEL DESIGN

High- level- design defines an application structure in general terms. It recognizes the framework's general i.e., equipment working framework's etc. engineering, for example, client/server, service- oriented. It recognizes the framework's significant parts, for example, detailing modules, and top- level classes. It ought to likewise outline on how the bits of the framework will interact. The high- level- design centers on what.

4.4 DATA FLOW DIAGRAM

DFD Level 0:

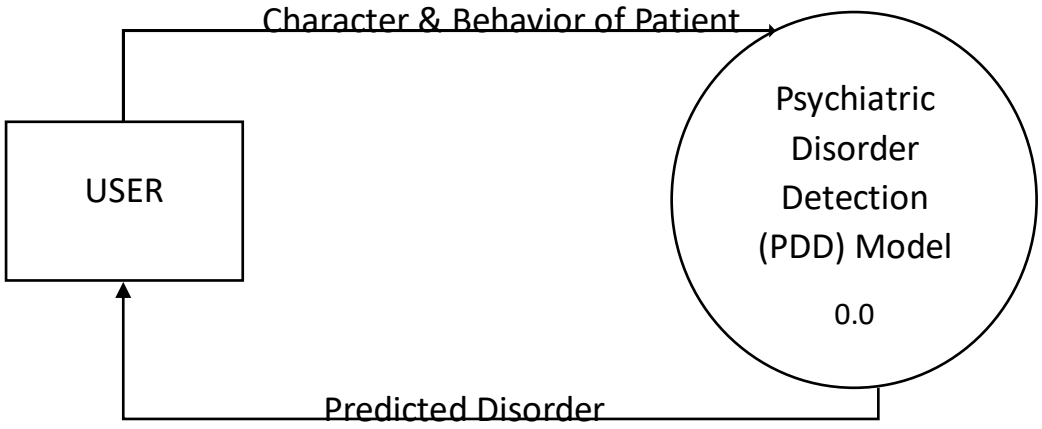


Fig 4.2: Level 0 DFD of PDD Model

EXPLANATION OF DFD Level 0:

In the above zero level (context diagram) DFD there is one entity “user” and whole system, "Psychiatric Disorder Detection (PDD) Model" is presented with a single bubble. A bubble in DFD Represents a Process.

DFD Level 1:

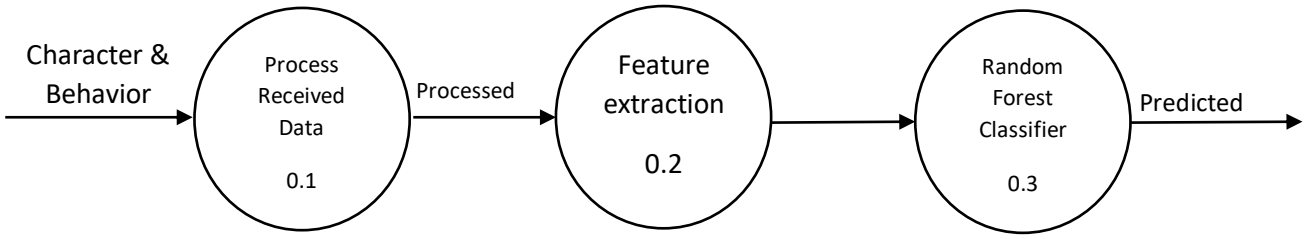


Fig 4.3: Level 1 DFD of PDD Model

EXPLANATION OF DFD Level 1:

In the above Level I DFD, there are three processes, the data received from the users or patients are processed through the model and features are extracted and classified into groups. These are then converted into numeric data and transferred to the next process i.e., random forest classifier, hence predicting the disorder.

DFD Level 2:

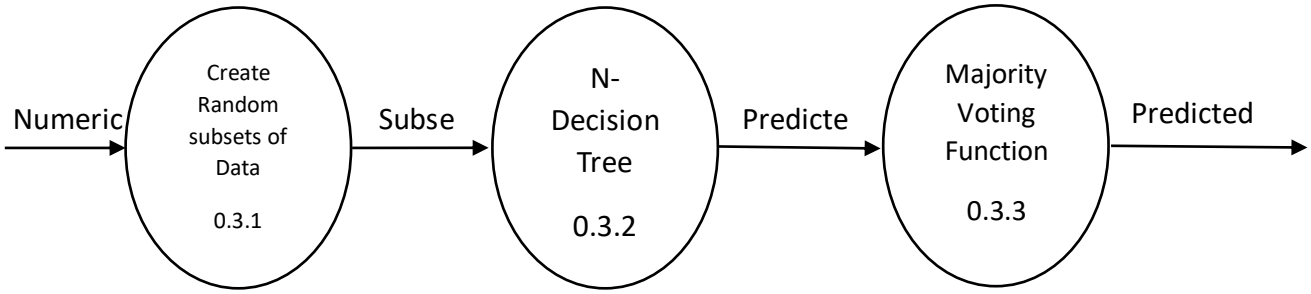


Fig 4.4: Level 2 DFD of PDD Model

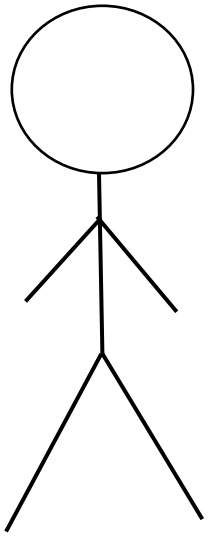
EXPLANATION OF DFD Level 2:

After the Numeric data is transferred to the random forest classifier, from these data, decision trees are built associated with the selected data points or subsets. After which n-decision trees are selected and for new subsets, predictions of each decision trees are found out and these new subsets are assigned to the category that wins the majority votes.

4.5 USE CASE DIAGRAM

A use case diagram at its simplest is a representation of user interaction with the system that shows the relationship between the user and the different use case in which the user is involved.

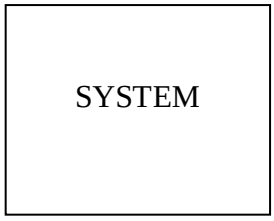
ACTOR: Actor in a case diagram is any entity that performs role in one given system. This could be a person, organization or an external system and usually drawn like skeleton shown below.



ACTOR

USE CASE: A use case represents a function or an action within the system. It's drawn as an oval and named with function.

SYSTEM: The system is used to define the scope of the use case and drawn as a rectangle. This is an optimal element but useful when you've visualizing large system.



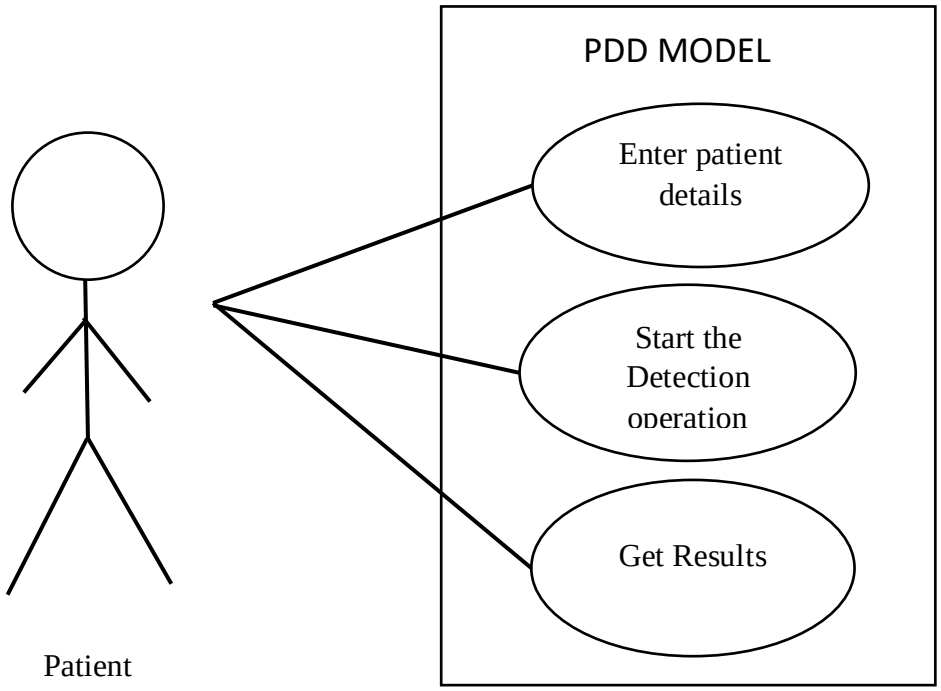


Fig 4.5: Use Case Diagram of the PDD Model

Description of use case:

In this use-case diagram, we have built a “psychiatric disorder detection system”. Here the user will be entering the required details through answering of the questionnaire given and these data will be feed into the model after which the detection operation will be started and final output will be presented predicting if the patient has any disorder or not. The actors are the Patients/users.

CHAPTER 5

IMPLEMENTATION

5.1 DESCRIPTION

The Psychiatric Disorder Detection system has been deployed using python module called PICKEL and a GUI is created using STREAMLIT library.

5.2 DATA PROCESSING

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff

from google.colab import drive
drive.mount('/content/drive')

# from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.ensemble import VotingClassifier
from sklearn.metrics import recall_score, confusion_matrix
from sklearn import preprocessing
```

```
df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/data1.csv',sep= ',')
# get the number of missing data points per column
missing_value_count = (df.isnull().sum())
print(missing_value_count[missing_value_count > 0])
# percent of data that is missing
total_cells = np.product(df.shape)
total_missing_value = missing_value_count.sum()
print('Percentage of missing value in Data Frame is: ',total_missing_value / total_cells *100 )
print('Total number of our cells is :',total_cells)
print('Total number of our missing value is :',total_missing_value)
```

```
country       2
major      11403
dtype: int64
Percentage of missing value in Data Frame is:  0.16479189694907417
Total number of our cells is : 6920850
Total number of our missing value is : 11405
```

```
# Delete 'major'&'country' and 'serial no.' column
df = df.drop('major', axis = 1)
df = df.drop('country', axis = 1)
df = df.drop('Unnamed: 0', axis = 1)
print("New shape of our data", df.shape)
# build list of not required features to be deleted from dataset

removedFeatures = [f'Q{i}E' for i in range(1, 43)] # add feature 'Q1E' to 'Q42E' to be removed
removedFeatures.extend([f'Q{i}I' for i in range(1, 43)]) # add feature 'Q1E' to 'Q42E' to be
removed
```

```
removedFeatures.extend([f'VCL{i}' for i in range(1, 17)]) # add feature 'VCL1' to 'VCL16' to be removed

removedFeatures.extend([ 'source', 'introelapse', 'testelapse', 'surveyelapse', 'engnat', 'hand', 'orientation',

    'voted', 'screensize', 'uniquenetworklocation'])


# remove features from the dataset

df = df.drop(removedFeatures, axis=1)


# Again removing some feature

rf = ['education','urban', 'gender', 'race', 'religion' , 'married','familysize','age']

df = df.drop(rf, axis=1)


# Removing TIPI features

rf2 = ['TIPI1','TIPI2','TIPI3','TIPI4','TIPI5','TIPI6','TIPI7','TIPI8','TIPI9','TIPI10']

df = df.drop(rf2, axis=1)


# Declare feature vector and target variable

x = df.drop(['target'], axis = 1)

y = df['target']

# Split data into separate training and test set

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 777)

# find best scored features

from sklearn.feature_selection import RFE

# Create the RFE object and rank each pixel

clf_rf = RandomForestClassifier()

rfe = RFE(estimator=clf_rf, n_features_to_select=30, step=1)

rfe = rfe.fit(x_train, y_train)


print('Chosen best 30 feature by rfe:',x_train.columns[rfe.support_])
```

```
Chosen best 30 feature by rfe: Index(['Q1A', 'Q3A', 'Q4A', 'Q5A', 'Q7A', 'Q8A', 'Q9A', 'Q10A',
'Q11A', 'Q12A',
       'Q13A', 'Q16A', 'Q17A', 'Q20A', 'Q21A', 'Q22A', 'Q24A', 'Q26A', 'Q27A',
       'Q28A', 'Q29A', 'Q30A', 'Q32A', 'Q33A', 'Q34A', 'Q36A', 'Q38A', 'Q39A',
       'Q40A', 'Q41A'],
      dtype='object')


X = df[x_train.columns[rfe.support_]]
y = df['target']
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size = 0.2, random_state = 777)
X_train.shape, X_test.shape, y_train.shape, y_test.shape


# create the classifier with n_estimators = 100


clf = RandomForestClassifier(n_estimators=100, random_state = 777)
# fit the model to the training set
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
print('Model Accuracy is: ',accuracy_score(y_pred,y_test))
cm = confusion_matrix(y_test,clf.predict(X_test))
# sns.heatmap(cm,annot=True,fmt="",cmap='Blues')


Model Accuracy is:  0.8981772470144563
```

5.3 MODEL DEPLOYMENT

```
import pickle
filename = 'model1.pkl'
pickle.dump(clf, open(filename,'wb'))
load_model = pickle.load(open('model1.pkl','rb'))
```

5.4 GUI using STREAMLIT

```
import numpy as np
import pickle
import streamlit as st

load_model = pickle.load(open('D:/Minor Project/model1.pkl','rb'))

def run_model(input):
  newinp = np.array([input])
  pred = load_model.predict(newinp)
  # print(pred)
  if(pred[0]==0):
    return "You have no psychiatric disorder"
  elif(pred[0]==1):
    return "You might have mild psychiatric disorder"
  elif(pred[0]==2):
    return "You might have moderate psychiatric disorder. We advise you to consult a psychiatrist"
  else:
    return "You might have severe psychiatric disorder. We advise you to consult a psychiatrist
immediately"

def main():
  st.title('Psychiatric Disorder Detection')
  result = ""
  n = st.text_input('Enter your name')
  st.write("")
  st.write("***All questions are necessary !!")
  st.write(" '1' >> (Didn't apply to me at all)")
  st.write(" '2' >> (Applied to me to some degree)")
```

```
st.write(" '3' >> (Applied to me to a considerable degree)")
st.write(" '4' >> (Applied to me very much, or most of the time)")
st.write("")
st.write("Questions for detection of Psychological disorder (30):")
st.write("")

q1 = st.radio('Q1. I found myself getting upset by quite trivial things.', ['1', '2', '3', '4'])
q2 = st.radio("Q2. I couldn't seem to experience any positive feeling at all.", ['1', '2', '3', '4'])
q3 = st.radio('Q3. I experienced breathing difficulty in the absence of physical exertion.', ['1', '2', '3', '4'])
q4 = st.radio("Q4. I just couldn't seem to get going.", ['1', '2', '3', '4'])
q5 = st.radio('Q5. I had a feeling of shakiness', ['1', '2', '3', '4'])
q6 = st.radio('Q6. I found it difficult to relax.', ['1', '2', '3', '4'])
q7 = st.radio('Q7. I found myself in situations that made me so anxious I was most relieved when they
ended.', ['1', '2', '3', '4'])
q8 = st.radio('Q8. I felt that I had nothing to look forward to.', ['1', '2', '3', '4'])
q9 = st.radio('Q9. I found myself getting upset rather easily.', ['1', '2', '3', '4'])
q10 = st.radio('Q10. I felt that I was using a lot of nervous energy.', ['1', '2', '3', '4'])
q11 = st.radio('Q11. I felt sad and depressed.', ['1', '2', '3', '4'])
q12 = st.radio('Q12. I felt that I had lost interest in just about everything.', ['1', '2', '3', '4'])
q13 = st.radio("Q13. I felt I wasn't worth much as a person.", ['1', '2', '3', '4'])
q14 = st.radio('Q14. I felt scared without any good reason.', ['1', '2', '3', '4'])
q15 = st.radio("Q15. I felt that life wasn't worthwhile.", ['1', '2', '3', '4'])
q16 = st.radio('Q16. I found it hard to wind down.', ['1', '2', '3', '4'])
q17 = st.radio("Q17. I couldn't seem to get any enjoyment out of the things I did.", ['1', '2', '3', '4'])
q18 = st.radio('Q18. I felt down-hearted and blue.', ['1', '2', '3', '4'])
q19 = st.radio('Q19. I found that I was very irritable.', ['1', '2', '3', '4'])
q20 = st.radio('Q20. I felt I was close to panic.', ['1', '2', '3', '4'])
q21 = st.radio('Q21. I found it hard to calm down after something upset me.', ['1', '2', '3', '4'])
q22 = st.radio('Q22. I found it difficult to tolerate interruptions to what I was doing.', ['1', '2', '3', '4'])
q23 = st.radio('Q23. I was in a state of nervous tension.', ['1', '2', '3', '4'])
```

```
  q24 = st.radio('Q24. I felt I was pretty worthless.', ['1', '2', '3', '4'])

  q25 = st.radio('Q25. I was intolerant of anything that kept me from getting on with what I was doing.', ['1', '2', '3', '4'])

  q26 = st.radio('Q26. I felt terrified.', ['1', '2', '3', '4'])

  q27 = st.radio('Q27. I felt that life was meaningless.', ['1', '2', '3', '4'])

  q28 = st.radio('Q28. I found myself getting agitated.(troubled or nervous)', ['1', '2', '3', '4'])

  q29 = st.radio('Q29. I was worried about situations in which I might panic and make a fool of myself.', ['1', '2', '3', '4'])

  q30 = st.radio('Q30. I experienced trembling (eg, in the hands).', ['1', '2', '3', '4'])

  st.write("")

  if st.button("TEST"):

    val = [q1,q2,q3,q4,q5,q6,q7,q8,q9,q10,q11,q12,q13,q14,q15,q16,q17,q18,q19,q20,q21,q22,q23,q24,q25,q26,q27,q28,q29,q30]

    n= n+','

    result =n+' '+ run_model(val)

    st.success(result)


if _name_ == "_main_":

  main()
```

5.5 SCREENSHOTS

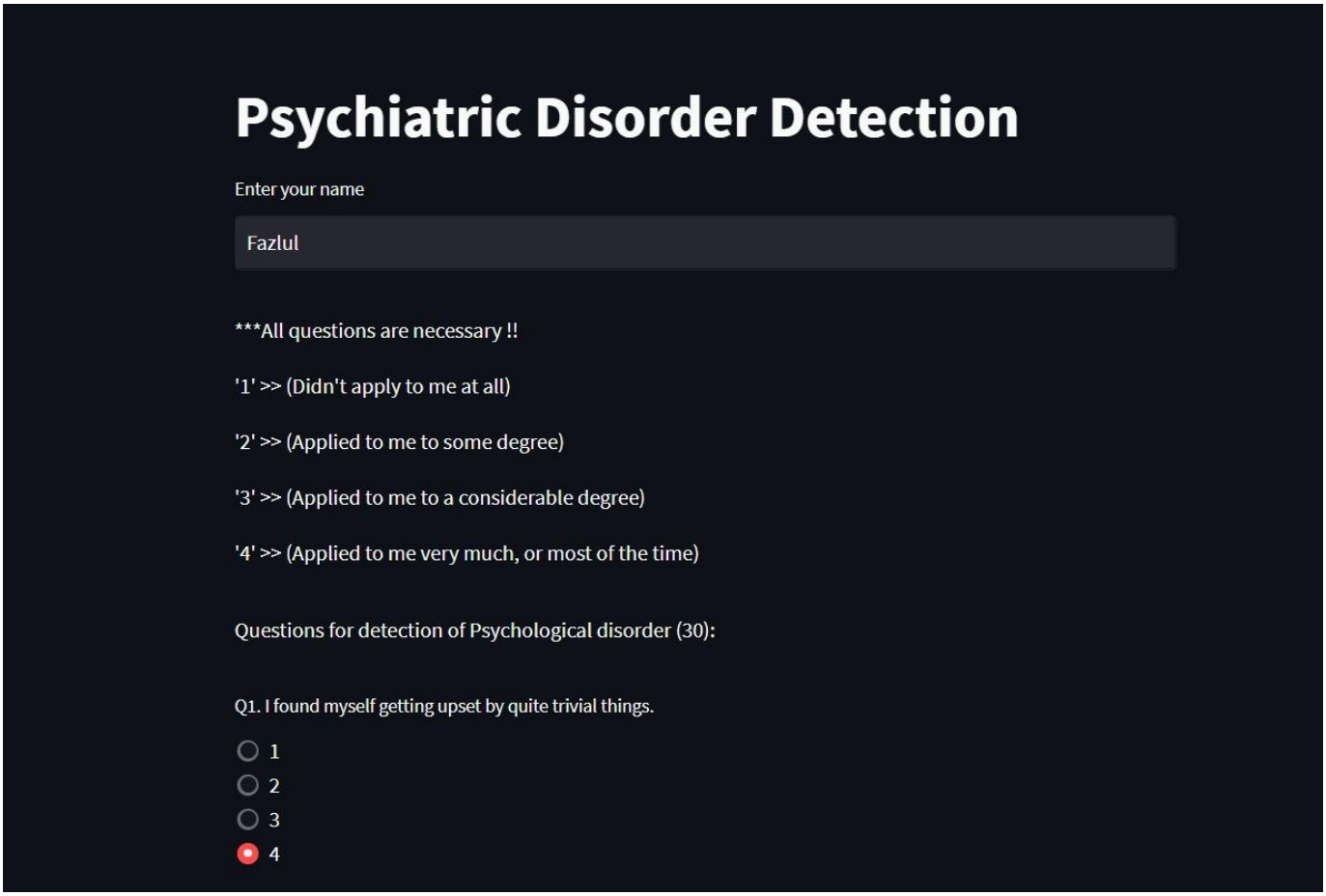


Fig. 5.5.1 GUI on Psychiatric Disorder Detection

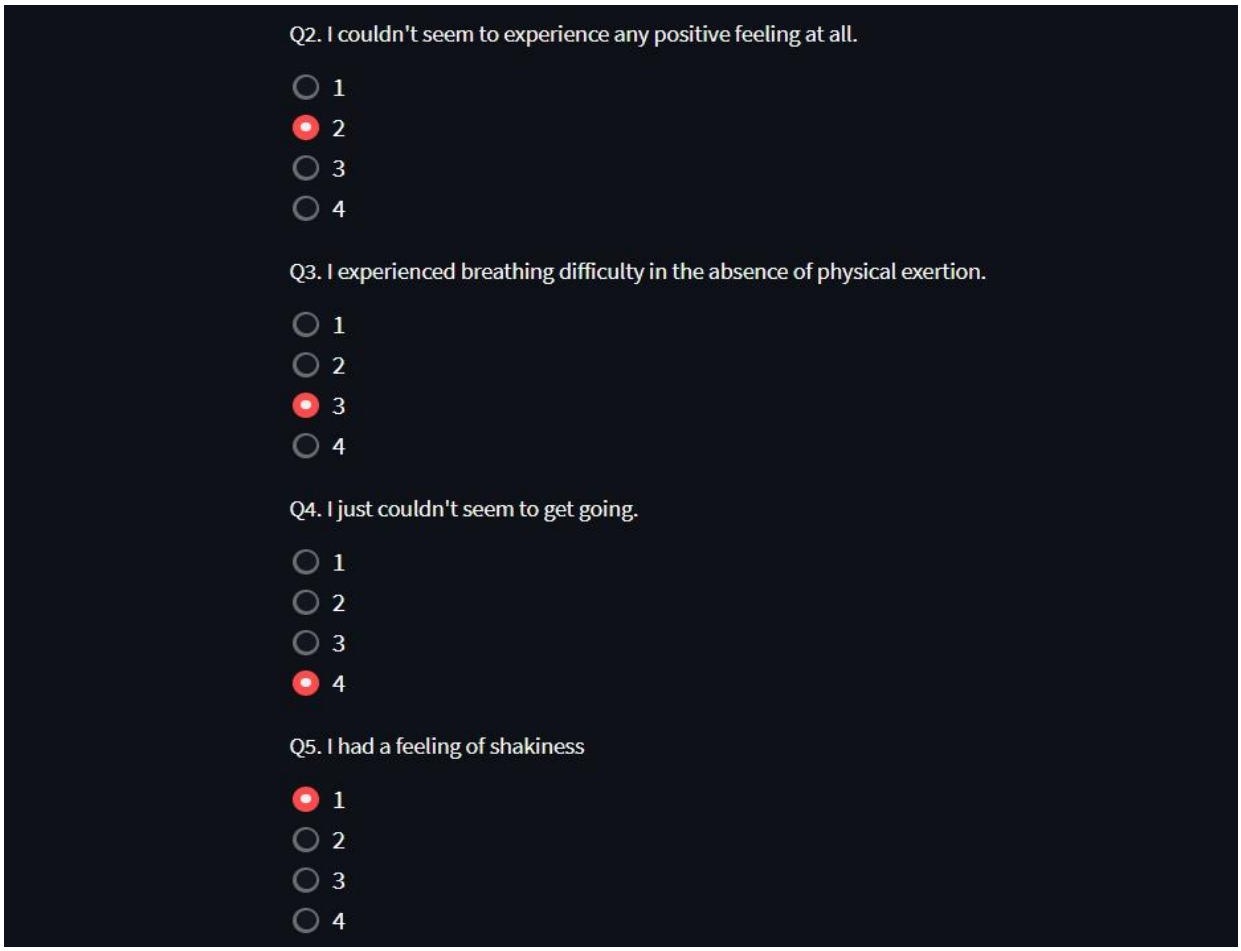


Fig. 5.5.2 GUI input Page

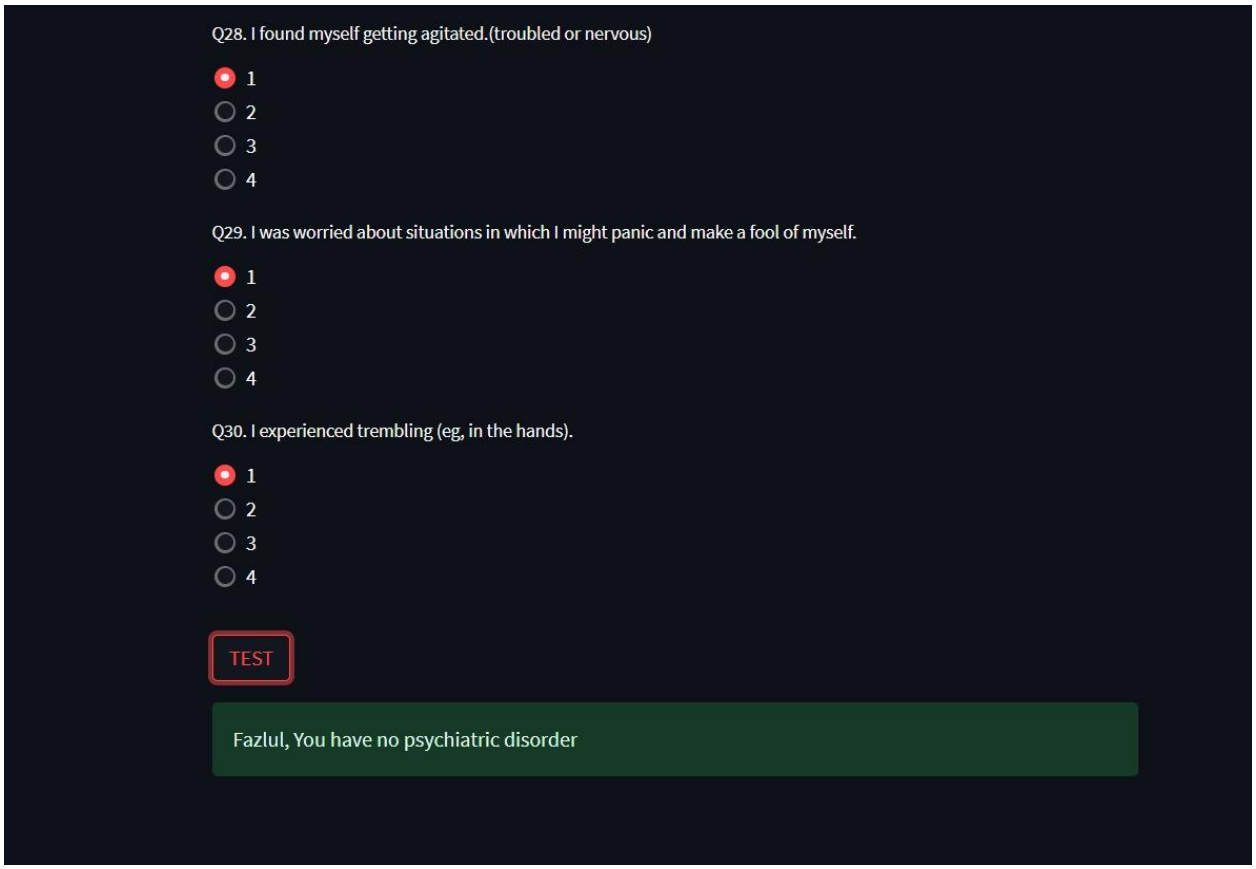


Fig. 5.5.3 GUI output Page

CHAPTER 6

TESTING

Testing is one of the most important phases in software development life cycle. The aim of the testing phases is to realize and identify all defects in a program. Although it is not possible to create software completely fault free, but through testing it can expose and rectify a large degree of errors and improve software quality.

6.1 TEST ENVIRONMENT

All modules of the "DETECTION OF PSYCHIATRIC DISORDER USING MACHINE LEARNING" were tested out under the following test environment:

Operating System: windows IO: 1909 built 64

Hardware environment:

- Ryzen 5 5600H series
- 8gbRAM
- NVIDIA graphics card

6.2 TESTING METHODOLOGY

Software testing methodology is defined as a strategies and testing types used to certify that the application under Test meets client expectations. Test methodology include functional and non-functional testing to validate the PDD. Examples of testing Methodologies are Unit Testing, Integration Testing, System Testing, Performance Testing etc.

6.2.1 UNIT TESTING

Unit Testing is a type of software testing where individual's units or components of software. The purpose is to validate that each unit of the software code performs as expected. Unit testing is done during the development (coding phases) of an application by the individual function, method, procedure, or object.

In SDLC, Iterative waterfall model, unit testing is first level of testing done before integration testing. Unit testing is a white box testing technique that is usually performed by the developer. The modules that constitute the "Detection of Psychiatric Disorder Using Machine Learning" system is: data collection, data pre-processor, model Deployment and GUI. The unit testing of each specific module has been discussed below.

Data collection

Serial no. of case	1
Name of the test	Data Collection
Feature being Tested	Feature Correlation
Description	Tested the data to increase accuracy
Input	Collected Data
Expected Output	Feature correlation matrix
Actual Output	Correlation matrix

Table no. 6.2: Unit-test case 1

Data Pre-Processor

Serial no. of test case	2
Name of the Test case	Data Pre-process
Feature being Tested	Data pre-processing capabilities
Description	The data collected from patients are feed into the model
Input	Raw data to be transformed into machine readable format
Expected output	The system shows data to main screen
Actual output	The system show data to main screen
Remarks	Test passed

Table no. 6.3: Unit-test case 2

6.2.2 INTEGRATION TESTING

The second phase of testing is integration testing. This phase is used to connect the individual's module and develop it into a complete system. The incremental manner of integration is given below.

Serial no. of case	1
Name of the test case	Data pre-processor module
Description	Integrate prediction module with the training module
Input	Collected data
Expected output	Prediction accuracy
Actual output	Prediction accuracy reached 90%
Remarks	Test passed

Table no. 6.4: Integration Test case 1

6.2.3 SYSTEM TESTING

System testing is carried out once the complete system has been integrated and the errors occurring during integration has been rectified.

Serial no. of case	1
Name of the test	System testing
Feature being Tested	Full functionality of the system
Description	Testing the data and based on the system prediction.
Input	Collected Data
Expected Output	Degree of predicted disorder
Actual Output	Degree of predicted disorder
Remarks	Test Passed

Table no. 6.5 : System-test case 1

CHAPTER 7

RESULT AND ANALYSIS

In our project, patient data are being used to predict the probability of having a psychiatric disorder. We use various behavioral questions to get information about patients which are processed into the model to predict the level of disorder (if any). The questions are being asked with the help of GUI which was made using Streamlit. These collected data are being processed by random forest algorithm for classification and finally getting the output in the patient screen. Due to this system, it will be easy for the health department and patient's guardian of any location to check the patient's mental health condition.

CHAPTER 8

CONCLUSION

8.1 CONCLUSION

Psychiatric disorder is one of the most common problem faced my people in today's era to be found among peoples of many age groups. In the last decade, for unknown reasons, the number of people with this disease has been rising worldwide. Numerous studies have been already conducted by researchers around the world, but the real reasons are still unknown. Despite the availability of many psychologist, the reason cause of having a psychiatric disorder is not accurately found out through the manual methods available. Also it takes a lot of time and effort to detect accurately the cause of it. On the other hand, machine learning has rapidly received popularity throughout the world in the last few decades. This technology can accurately detect automatically how much the particular person is suffering from the specific disorder. Our proposed system classifies the data provided by the patients through the questionnaire and accurately determines the degree to which the patient is exposed.

8.2 FUTURE SCOPE

In future this project can be modified by adding few more things like:

- Currently we are using local host, so its functions are limited. In future we can make globally usable with more advanced features.
- We can launch our model in real-time application to make it accessible to every person possible so that patients get quick and accurate results.
- Currently our project is limited only to showing how much the patient is exposed to the disorder, but in future we can improve the model by adding more features like showing accurately the disorder he is exposed to.

8.3 SOCIAL RELEVANCE

Medical strategies being currently deployed in-order to detect psychiatric disorders require manual intervention from a professional medical practitioner. Data from patients are being collected by psychiatrists manually and other various extensive medical tests are conducted. These tests may be quite expensive to a large part of the population. There is no automated system in usage that analyses and evaluates the symptoms collected from patients and detects the particular disorder in less time for diagnosis. This greatly limits the accessibility of the population to early psychiatric disorder detection as it is vital for treatment and avoiding later complications.

The project attempts to automate this process by applying machine learning strategies to Psychiatric disorder detection. The features collected from patients consist of many vital features and properties that are almost oblivious to a untrained or even trained psychiatrists/psychologists. These features can however be effectively utilized by ML modules in determining the disorder accurately.

This model has potential benefits for people suffering with serious psychiatric disorders providing it is early on in the care of an individual with the disease and is tailored to each individual's needs. It can enable people to diagnose their disorders as it provides a pre-diagnosis treatment to the patients for getting effective results after treatment. It also helps in enhancing the life for people with the disease.

This project aims to achieve this as it will serve effective and efficient alternative to the extensive and expensive medical processes. This project aims to nullify the requirement of these tests in diagnosis of psychiatric disorders. This in turn will enable the significant boost in the accessibility of the population in the early detection of psychiatric disorders. It also speeds up the process of detecting PD. Also due to nullification of the complicated medical examinations usually involved, this allows to conserve upon the vast wealth of resources consumed in detecting the disorders. This acts as a great boon upon conserving and optimizing the usage of medical resources.

REFERENCES

[1] George Karystianis, Aleo J.Nevado, Chi-Hun Kim, Azad Dehghan, Jhon A. Keane, Goran Nenadic. "Automatic mining of symptoms severity from psychiatric evaluation notes", Int J Methods Psychiatry Res. (2018)

[2] Fani Deligianni,Yao Guo, Guang-Zhong Yang. "From Emotion to Mood Disorder: A Survey on Gait Analysis Methodology", Imperial College London (2019)

[3] Yousra Bouaiachi , Mohamed Khaldi , Abdellah Azmani "Neural Network-based Decision Support System for Pre-diagnosis of Psychiatric Disorders", Third IEEE International CIST (2014)

[4] Luciano Comin Nunes, Placido Rogerio Pinheiro, Mirian Caliope Dantas Pinheiro , Marum Simao Filho , Rafael Espindola Comin Nunes,and Pedro Gabriel Caliope Dantas Pinheiro "Automatic Detection and Diagnosis of Neurologic Diseases", IEEE (2019)

[5] Vanishri Arun, Prajwal V, Murali Krishna, Arunkumar B. V, Padma S. K, Shyam V "A Boosted Machine Learning Approach for Detection of Depression" Symposium Series on Computational Intelligence SSCI (2018)

[6] I. Kinder, K. Friganovic, J. Vukojevic, D. Mule, T. Slukan, D. Vidovic, P. Brecic, M. Cifrek "Comparison of Machine Learning Methods in Classification of Affective Disorders" MIPRO (2020)

[7] Shafeeq Ahmed, Preeti Singh, Prakash Singh. "Case Based Reasoning Model in the Diagnosis of Psychiatric Disorder" International conference on Communication and Electronics System (ICCES) (2016)

[8] Nivedhitha Mahendran, P.M. Durai Raj Vincent, Kathiravan Shrinivasan, Vishal Sharma, Dushantha N. K. Jayakody. "Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults", IEEE (2020)

[9] B.S. Mahanand, S. Suresh, N. Sundararajan, M. Aswatha Kumar, “Identification of brain regions responsible for Alzheimer’s disease using a Self-adaptive Resource Allocation Network”, Neural Networks (2012)

[10] Leonore M. de Wit, Annemieke van Straten, Marieke van Herten, Brenda WJH Penninx, and Pim Cuijpers, “Depression and body mass index, a u-shaped association”, BMC Public Health, Longitudinal assessment of neuropsychological functionin major depression. Aust. N. Z. J. Psychiatry(2009).

[11] S. D. Puthankattil and P. K. Joseph, "Analysis of EEG Signals Using Wavelet Entropy and Approximate Entropy: A Case Study on Depression Patients," Int. J. Medical, Heal. Biomed Bioeng Pharm, Eng,(2014).

[12] G.Ulrich, E. Renfordt, G.Zeller, and K.Frick, "Interrelation between changes in the EEG and psychopathology under pharmacotherapy for endogenous depression . A contribution to the predictor question,” Pharmacopsychiatry, (1984) .

[13] A. Ahmed, R. Sultana, M. T. R. Ullas, M. Begom, M. M. I. Rahi, and M. A. Alam, "A machine learning approach to detect depression and anxiety using supervised learning," in Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast ,Australia ,(2019).

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

[15] D. Spronk, M. Arns, K. J. Barnett, N. J. Ccx»per, and E. Gordon, "An investigation of EEG, genetic and cognitive markers of treatment to antidepressant medication in patients with major depressive disorder: A pilot study," J. Affect. Disord. ,(2011).