MINOR PROJECT

(June - December 2023)

CogniMind :

MDD Diagnosis using EEG Analysis

Submitted by:

SID	NAME
21103104	Divyanshu
21103100	Kartik Mittal
21103092	Peyush Jindal
20103086	Aditya Goel

Under the guidance of Dr. Trilok Chand

Department of Computer Science and Engineering

Punjab Engineering College (Deemed to be University), Chandigarh

DECLARATION

I hereby declare that the project work entitled **"CogniMind"** is an authentic record of my own work carried out at Place of work as requirements of Minor Project (CS 1504) of 5th semester for the award of degree of B.Tech. in the Department of Computer Science and Engineering, Punjab Engineering College (Deemed to be University), Chandigarh, under the guidance of **Dr. Trilok Chand** with designation during June-December 2023.

Divyanshu	**Kartik Mittal**	**Peyush Jindal**	**Aditya Goel**
(21103104)	(21103100)	(21103092)	(20103086)

Date: ____________________

Certified that the above statement made by the student is correct to the best of our knowledge and belief.

Dr. Trilok Chand

(Faculty Coordinator)

ACKNOWLEDGMENT

We would like to thank Dr. Trilok Chand for giving us the opportunity to undertake this project "CogniMind". This opportunity has helped us learn more about various domains of computer engineering and machine learning and has helped us grow as capable engineers.

His guidance has allowed us to navigate the complexities inherent in the implementation of various features which helped us complete the project and we are deeply thankful for that. We are also positive that the experience gained by this project will help us in getting opportunities in future endeavors and for that we are deeply thankful.

Other than this, we would like to thank our friends and family for their constant support.

TABLE OF CONTENTS

List of Figures :

List Of Tables:

List Of Abbreviations:

- EEG : Electroencephalogram
- MDD : Major Depressive Disorder
- KNN : K Nearest Neighbours
- SVM : Support Vector Machine
- CRM : Customer relationship management
- RBF : Radial basis function
- Mne : Minimum Norm Estimation
- PDF : Probability density function
- WT : wavelet transform
- CNN : Convolutional Neural Networks
- LSTM: Long Short Term Memory

ABSTRACT

This report provides information about the development of AI models for the diagnosis of Major Depressive Disorder using EEG signals. The dataset used in the study consists of EEG data from 34 MDD patients. The document also discusses the pre-processing of the data, including band-pass and notch filtering, as well as feature selection and extraction methods such as multilevel decomposition and spectral density estimation. The project aims to expand its scope to diagnose other mental health disorders and analyze sleep patterns. It highlights the potential use of EEG signals in depression detection and suggests future improvements and applications.

INTRODUCTION

WHO considers depression as the primary contributor to global disability, and it poses dangerous threats to approximately all aspects of human life. Among more than 300 million people suffering from some depressive disorder, just twenty percent receive professional help. Prevailing clinical practices do not use any physiological or biochemical tests to confirm in their diagnostic process, for example, the presence of specific biomarkers in a person's body before prescription of the medication. The diagnosis is still relying on a conversation only. Using EEG signals is a well-established and non-invasive method and is one of the oldest neuroimaging techniques. Many researchers prefer this approach due to its accessibility to a large number of patients and cost-effectiveness.

The brain is constantly active, generating electrical activity that is very subtle (significantly less than a 9V battery) but detectable with a suitable device. Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity on the scalp. It represents the macroscopic activity of the surface layer of the brain underneath. Neuroscientific research has obtained consistent findings and established well-accepted theories on how the EEG signals related to cognitive, affective, or attentional processing. Thus, EEG Signals are to help identify MDD or Major Depressive Disorder. However, the complex, nonlinear and non-stationary electroencephalogram (EEG) signals are very tedious to interpret visually and complicated to extract the significant features from them. In this project, our main objective is diagnosing Major Depressive Disorder by analyzing the EEG signals.

BACKGROUND STUDY

The project involves studying mental health disorders, EEG signals, and AI model development. This section provides an overview of existing research and techniques in the field. Current clinical practices do not use physiological or biochemical tests in the diagnostic process. Diagnosis relies solely on conversation. The non-invasive EEG method is a well-established neuroimaging technique, preferred by many due to its accessibility and cost-effectiveness.

Mental Health Disorders: The project aims to diagnose MDD. The goal is to develop a system to detect and evaluate disorders like bipolar disorder, PTSD, OCD, social anxiety, panic disorders, phobias, schizophrenia, borderline personality disorder, narcissistic personality disorder, and antisocial personality disorder.

EEG Signals: Electroencephalogram (EEG) signals are the primary data source. An EEG cap with nineteen sensors captures neural signals, providing insights into brain activity and mental health.

AI & ML Models Development: Deep learning models like CNN and CNN LSTM, and ML models like KNN and SVM, classify EEG signals. Trained on MDD outpatient data, feature selection methods like multilevel decomposition, spectral density estimation, entropy calculation, and fractal dimension analysis extract relevant information.

System Design Tools: Activity diagrams, sequence diagrams, state diagrams, UML class diagrams, and use-case diagrams define the system architecture. These tools organize information and identify gaps and conflicts in the system design.

Development Tools: React.js, Flask, MongoDB, and Flask-MongoEngine are employed. React.js creates dynamic client-side applications, Flask is a web framework, and PyMongo provides Pythonic language for database access. Flask-MongoEngine simplifies MongoDB integration with Flask.

The project aims to develop an accurate system for diagnosing mental health disorders using EEG signals and AI models by understanding these background aspects.

PROPOSED WORK

Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity on the scalp that has been shown to represent the macroscopic activity of the surface layer of the brain underneath. It is typically non-invasive, with the electrodes placed along the scalp. EEG Signals can be used to help identify MDD or Major Depressive Disorder. However, the complex, nonlinear and non-stationary electroencephalogram (EEG) signals are very tedious to interpret visually and highly difficult to extract the significant features from them. In this project, our main objective is the diagnosis of major depressive disorder by analyzing the EEG signals.

In this Project, Doctor can create an account and can add patients he is currently viewing. He can update EEG signal recordings of the patient and can use the software to get their analysis. This website will not only allow Medical professionals to analyze the signals but also allows them to store the information of their patients. Patients can also log in and view their Report with remarks of the Doctor available with it.

IMPLEMENTATION DETAILS

K-Nearest Neighbors

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

For classification problems, a class label is assigned on the basis of a majority vote—i.e. the label that is most frequently represented around a given data point is used. While this is technically considered "plurality voting", the term, "majority vote" is more commonly used in literature. The distinction between these terminologies is that "majority voting" technically requires a majority of greater than 50%, which primarily works when there are only two categories. When you have multiple classes—e.g. four categories, you don't necessarily need 50% of the vote to make a conclusion about a class; you could assign a class label with a vote of greater than 25%.

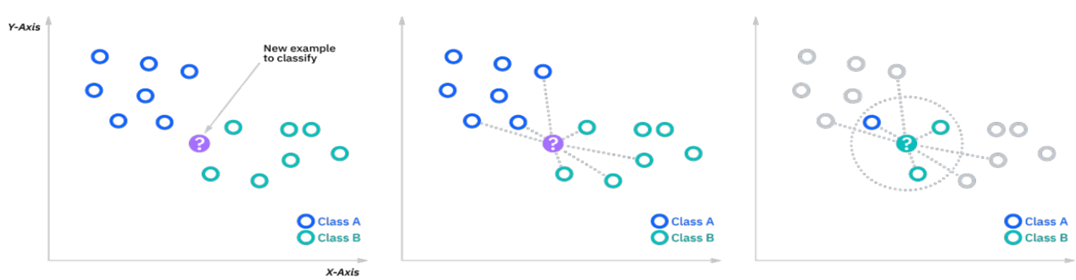


Fig-1: K Nearest Neighbours

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a robust classification and regression technique that maximizes the predictive accuracy of a model without overfitting the training data. SVM is particularly suited to analyzing data with very large numbers (for example, thousands) of predictor fields.

SVM has applications in many disciplines, including customer relationship management (CRM), facial and other image recognition, bioinformatics, text mining concept extraction, intrusion detection, protein structure prediction, and voice and speech recognition.

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane. Following this, characteristics of new data can be used to predict the group to which a new record should belong.

For example, consider the following figure, in which the data points fall into two different categories.

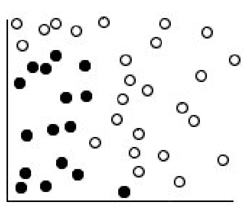


Fig-2: Original Dataset

The two categories can be separated with a curve, as shown in the following figure.

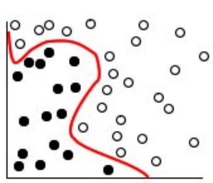


Fig-3: Data with separator added

After the transformation, the boundary between the two categories can be defined by a hyperplane, as shown in the following figure.

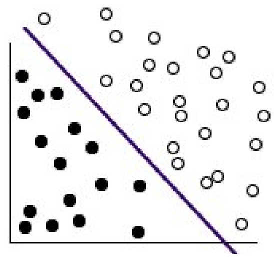


Fig-4: Transformed data

The mathematical function used for the transformation is known as the kernel function. SVM Model (in scikit-learn) supports the following kernel types:

- Linear
- Polynomial
- Radial basis function (RBF) (our model uses this)
- Sigmoid

Checking Similarity of two points X_1 and X_2

$$K(X_1, X_2) = exp(-\frac{||X_1 - X_2||^2}{2\sigma^2})$$

where,

1. ‘σ’ is the variance and our hyperparameter
2. $||X_1 - X_2||$ is the Euclidean (L_2-norm) Distance between two points X_1 and X_2

TECH STACK

Python: Utilized for face encoding and similarity comparison using the mne library. Employed the Pillow library for drawing boxes around recognized faces and displaying names. Integrated with cv2 for camera and face detection, capturing images sent to the face_recognition library for encoding and matching against known individuals.

Flask: Chosen as a web framework due to its simplicity and flexibility. Seamless integration with the face_recognition module, written in Python, eliminating the need for a JavaScript-based web app. Offers a small and easy-to-extend core.

Tailwind-CSS: Easy-to-use CSS framework. Enables writing CSS directly within HTML files, optimizing code by retaining only necessary CSS. Supports the use of traditional CSS, providing flexibility in choosing styles.

ReactJS: Declarative JavaScript framework for dynamic client-side applications in HTML. Facilitates building complex interfaces through simple components, connecting them to backend server data, and rendering them as HTML.

Flask-MongoEngine: Flask extension seamlessly integrating MongoDB with Flask applications. Simplifies MongoDB usage with Flask, offering convenient defaults and utilities for common tasks.

PyMongo: Python driver for MongoDB, a NoSQL database known for flexible, scalable, and high-performance document storage.

MongoDB: NoSQL database storing data in flexible, JSON-like documents.

About Our Dataset

The Dataset used in this study is provided by Mumtaz et al , with open access to everyone. A sample of 34 MDD outpatients (17 males and 17 females, mean age = 40.3 ± 12.9) was recruited. The EEG data were acquired according to the experiment design approved by the Hospital University Sains Malaysia (HUSM), Kelantan, Malaysia.

EEG cap with nineteen (19) electro-gel sensors was used to acquire EEG data. Electrodes Sensors from which Data acquired:- Fp1, F3, F7, Fz, Fp2, F4, F8, C3, C4, Cz, P3, Pz, P4, O1, O2, T3, T4, T5, T6.

Model Development

1. **Pre-Processing Data**
2. **Feature Selection/Extraction & Model Creation**
3. **Training on Dataset & Prediction**

- **Pre-Processing Data:**
    - EEG Signals are highly susceptible to noises. Signals that are picked up from the scalp are not necessarily an accurate representation of the signals originating from the brain, as the spatial information gets lost. Artifacts such as blinking or muscle movement can contaminate the data and distort the picture. Finally, we want to separate the relevant neural signals from spontaneous neural activity during EEG recordings.
- **Cropping :**

  Cropping the EEG signal involves selecting a specific portion of the signal for analysis or further processing.Cropping is performed to focus on specific time intervals within the EEG signal, eliminating irrelevant data.

- **Band-Pass Filter :**
    - We First use Band-Pass Filter (0.1 to 40Hz) to remove external interference and collect helpful information.
- **Notch Filter :**
    - We then used a 50Hz Notch Filter to remove electrical interference caused by equipment.
- **ICA Decomposition (ruinica) :**
    - Interference caused by muscles and eyes is removed so that only brain signals are used for the training model-threshold set at 80%.

Feature Selection and Extraction Methods

- **Multilevel Decomposition (DWT method): *Role:*** Integral in recognition and diagnostics, compressing time-varying biomedical signals into a concise parameter set. EEG signals being non-stationary, wavelet transform (WT) is employed for its time-frequency domain capabilities. WT expresses general functions as an infinite series of wavelets.
- **Spectral Density Estimation: *Purpose:*** Estimate power spectral density of a random signal in statistical signal processing. Characterizes frequency content, identifying periodicities through peaks at corresponding frequencies.
- **Power Spectral Density: *Measure:*** Quantifies signal power content versus frequency. Characterization: Used for broadband random signals, with amplitude normalized by spectral resolution.
- **Entropy Calculation: *Significance:*** Quantifies uncertainty or randomness in EEG time series patterns. Information Content: Roughly equivalent to the amount of information held within the signal.
- **Fractal Dimension: *Definition:*** Statistical index of complexity measuring changes in pattern detail with varying scales. Application: Provides insight into the complexity of EEG time series.
- **Kurtosis: *Purpose:*** Measures how distribution tails deviate from a normal distribution. Identification: Flags distributions with extreme values in their tails.
- **Skew: *Symmetry Indicator:*** Reveals the symmetry of amplitude probability

density function (PDF) in a time series. Interpretation: Zero skewness for equal large and small values; positive skew for more small values, negative for more large values. These methods collectively form a comprehensive approach to extracting relevant features from EEG signals, enabling effective analysis and diagnosis in the field of mental health disorders.

Algorithms Used

CNN (Convolutional Neural Network) Model: ***Operation:*** Applies convolution to signals, transforming and extracting relevant features. Filter Usage: 128, 64, or 32 filters in various layers. Convolutional Layers: First three layers use a filter size of 5, with sizes of 3 and 2 in subsequent layers.

CNN LSTM Model: ***Hybrid Approach:*** Combines CNN and LSTM for effective sequential data learning. LSTM Role: Learns from training, remembers learned information, and processes features for long-term dependencies.

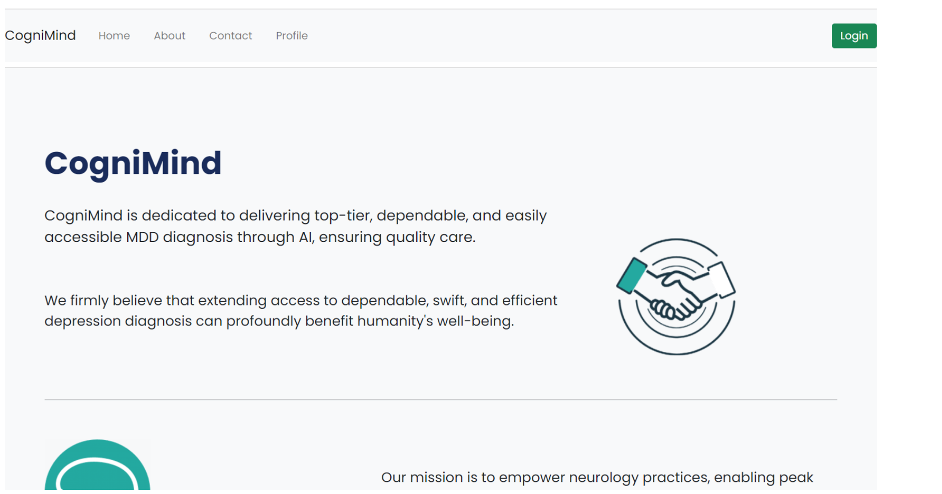
SVM (Support Vector Machine): ***Pattern Recognition:*** Analyzes input data, recognizing patterns in multidimensional space (hyper-plane). Data Representation: Samples represented as points, mapped into space for clear separability. Prediction: Based on established hyper-planes.

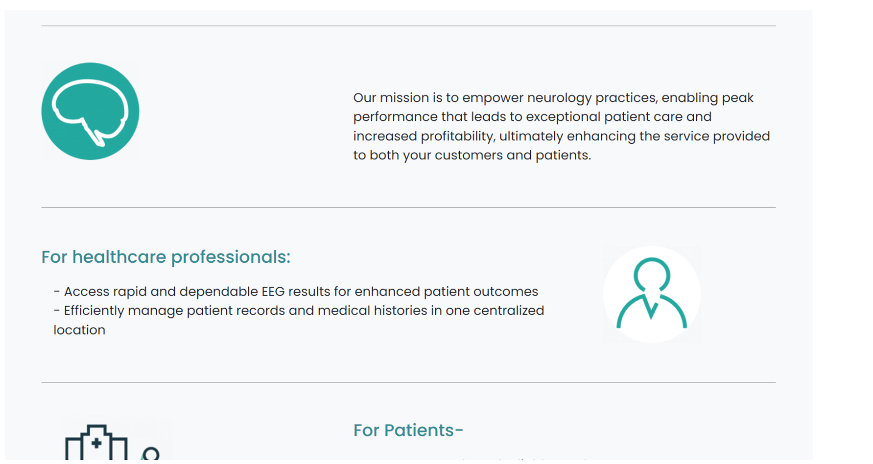
Weighted KNN (K-Nearest Neighbors): ***Prediction Principle:*** Predicts values for new samples based on feature similarity. Similarity Measure: Typically uses Euclidean distance.

Decision Tree: ***Tree Structure:*** Classifier structured like a tree, classifying instances based on feature values. Classification Process: Tests attributes from the root node to leaf nodes for instance classification. Random Forest: Ensemble Learning: Constructs multiple decision trees during training. Classification: Output determined by majority class selected by the ensemble.

ANN (Artificial Neural Network) Model: Neural Network Architecture: Based on connected units or neurons, modeling biological neurons. Signal Transmission: Each connection transmits a signal to other neurons. Learning Process: Weights of neurons and edges adjust during learning.

This diverse set of algorithms, ranging from CNNs and LSTMs for sequential data to SVMs and decision trees for pattern recognition, contributes to a comprehensive system for EEG signal classification and analysis in the context of mental health disorders.





Depression, identified by the WHO as a leading cause of global disability, affects over 300 million individuals worldwide, with only 20% receiving professional assistance.

Currently, diagnosis heavily relies on verbal assessments. Electroencephalography (EEG), a method monitoring scalp electrical activity, shows promise in detecting Major Depressive Disorder (MDD) by capturing brain surface activity. However, interpreting the intricate EEG signals, which are complex, nonlinear and non-stationary, poses a significant challenge.

Our initiative, CogniMind, focuses on diagnosing MDD through Machine Learning analysis of EEG signals. Our system offers a reliable and swift diagnosis, simplifying the process for healthcare professionals. By providing an intuitive interface for uploading patient EEG data and accessing ML-based analyses, we centralize patient records for ease of reference.

CogniMind aims to enhance neurological practices, ensuring superior patient care and improved operational efficiency. Our goal is to expand access to dependable and rapid depression diagnoses, ultimately benefiting humanity.

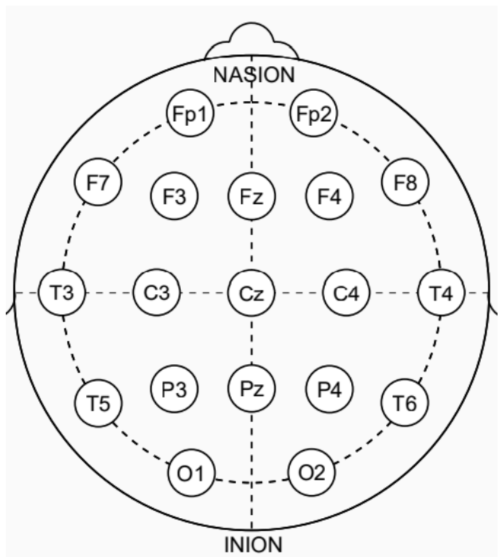
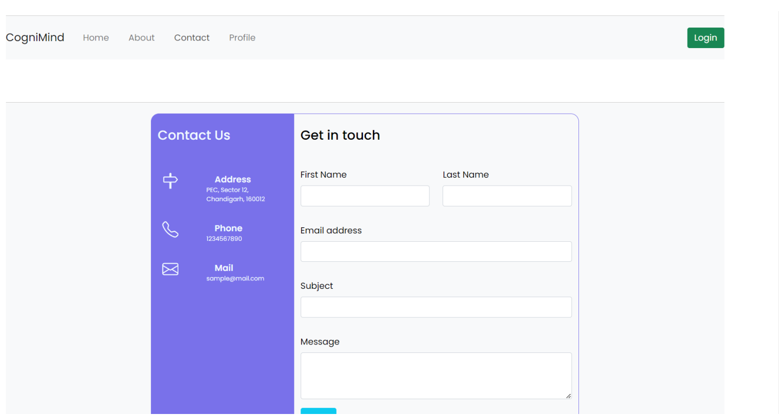
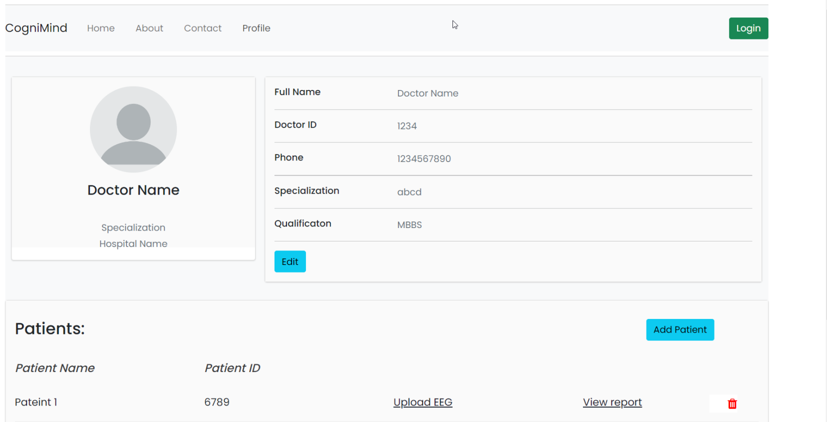
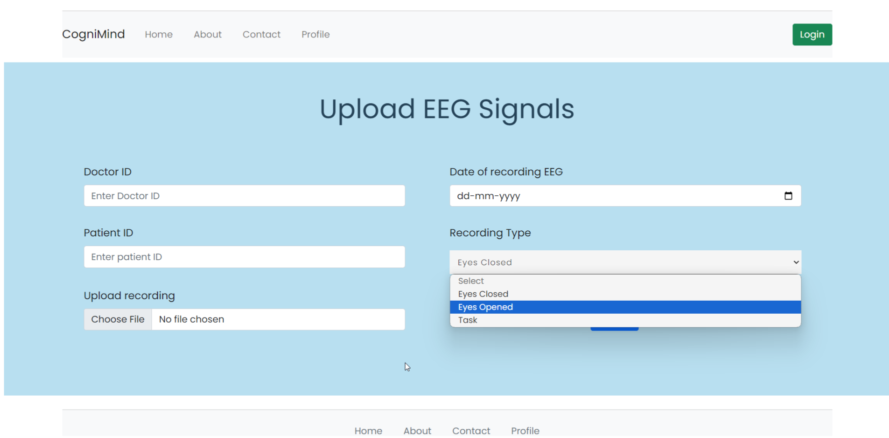
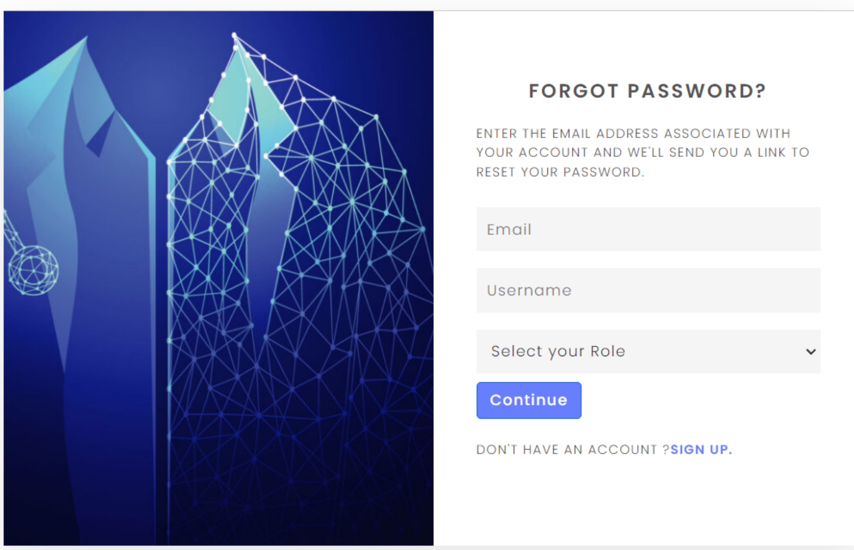


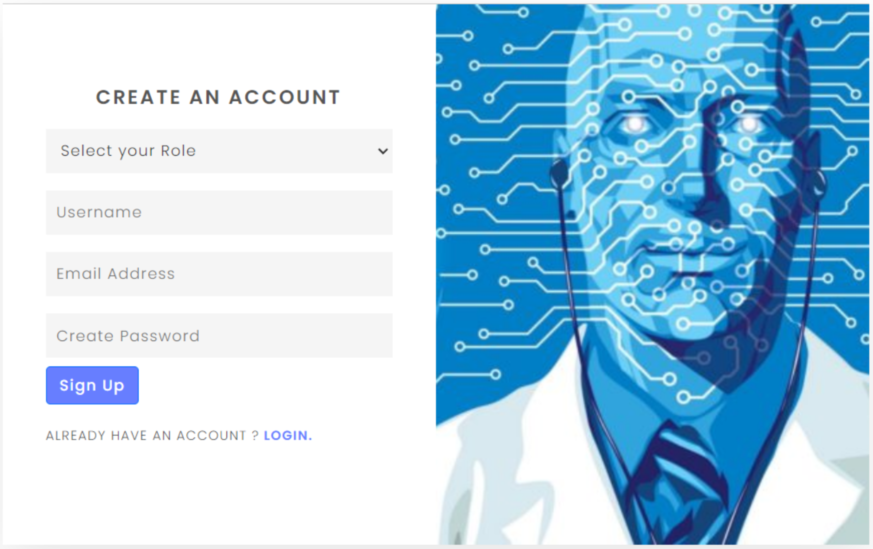
Fig-5: Brain Channel Mapping




CogniMind
Home
About
Contact
Profile
Login
Doctor Name
Specialization
Hospital Name
Full Name
Doctor Name
Doctor ID
1234
Phone
1234567890
Specialization
abcd
Qualificaton
MBBS
Edit
Patients:
Add Patient
Patient Name
Patient ID
Pateint 1
6789
Upload EEG
View report


CogniMind
Home
About
Contact
Profile
Login
Upload EEG Signals
Doctor ID
Enter Doctor ID
Date of recording EEG
dd-mm-yyyy
Patient ID
Enter patient ID
Recording Type
Eyes Closed
Select
Eyes Closed
Eyes Opened
Task
Upload recording
Choose File
No file chosen
Home
About
Contact
Profile


FORGOT PASSWORD?
ENTER THE EMAIL ADDRESS ASSOCIATED WITH YOUR ACCOUNT AND WE'LL SEND YOU A LINK TO RESET YOUR PASSWORD.
Email
Username
Select your Role
Continue
DON'T HAVE AN ACCOUNT ?SIGN UP.



SYSTEM DESIGN

Systems design defines elements of a system like modules, architecture, components, and their interfaces and data for a system based on the specified requirements. It is the process of defining, developing, and designing systems that satisfy a business or organization's specific needs and requirements. These system design tools are needed to organize the information, determine gaps in understanding, and identify conflict areas.

Modeling techniques used in the system development are:

- **Activity Diagram**
- **Sequence Diagrams**
- **State Diagram**
- **UML Class Diagram**
- **Use-Case Diagram**

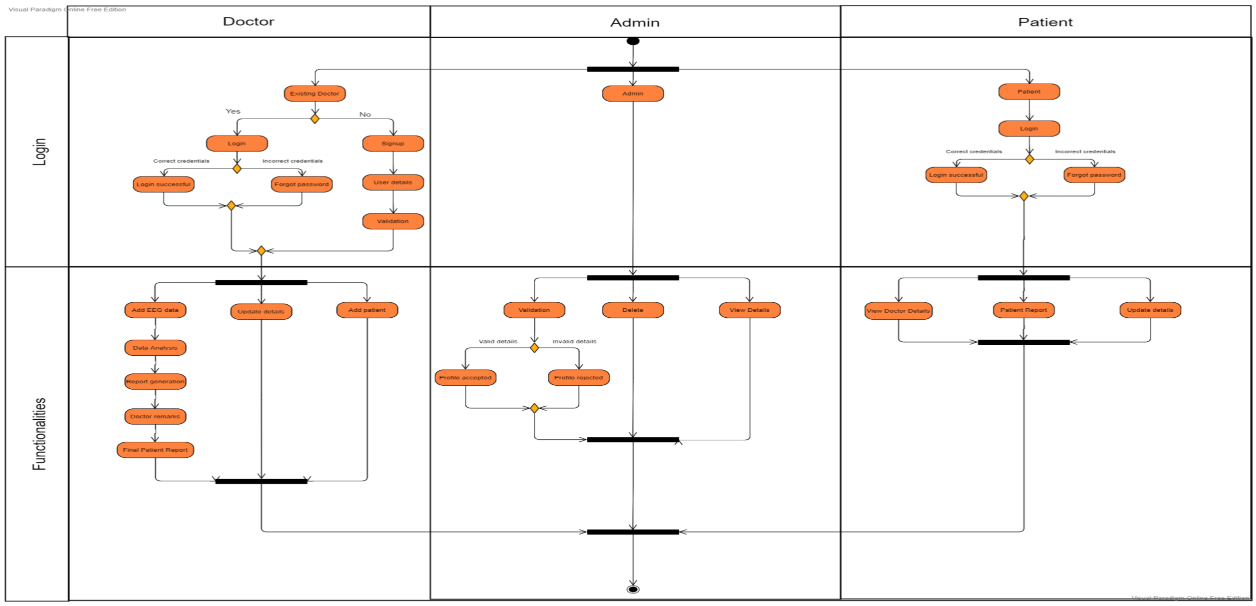


Fig-6: Activity Diagram

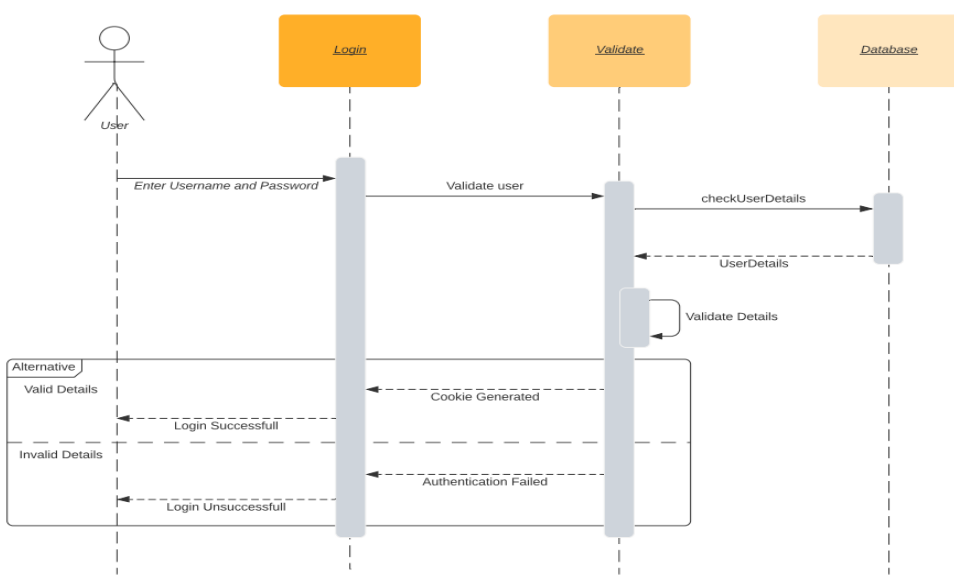


Fig-7: Sequence Diagram (Registration)

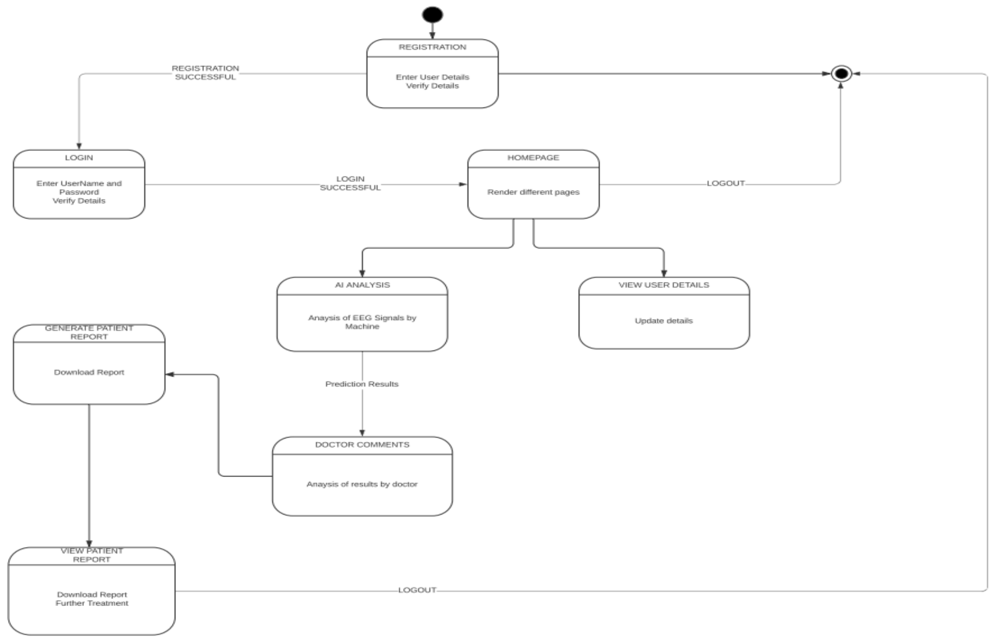


Fig-8: State Diagram

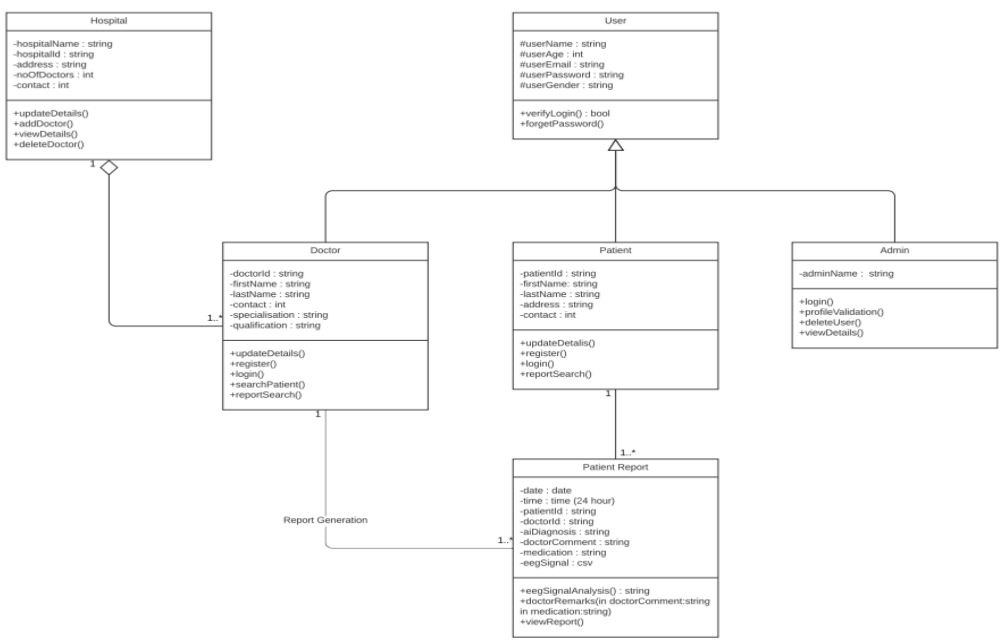


Fig 9: UML Class Diagram

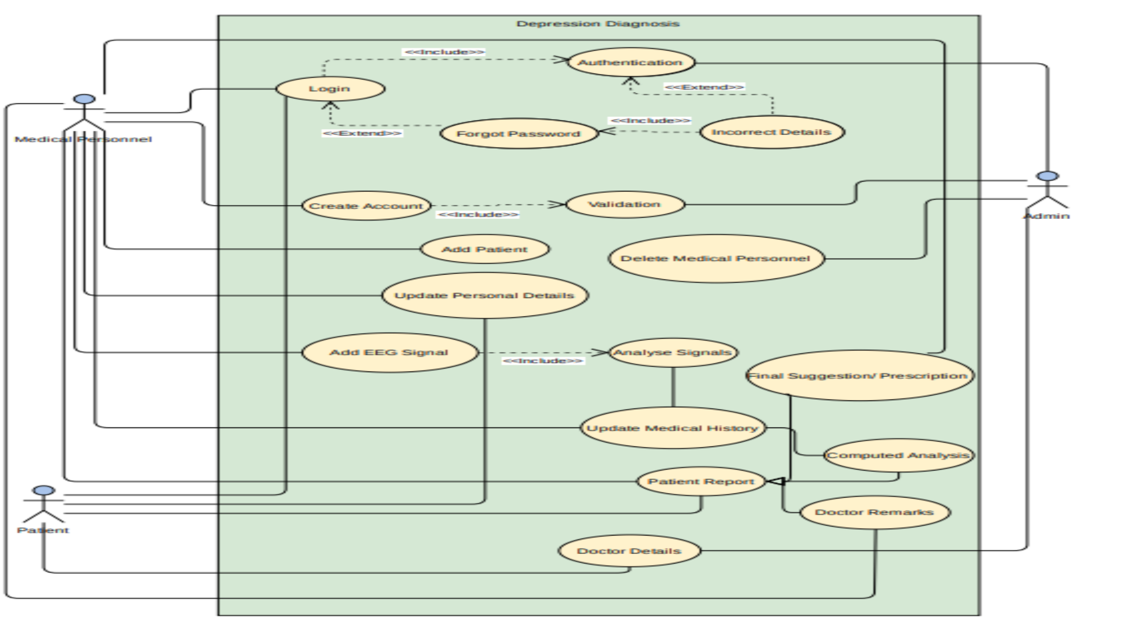


Fig-10: Use Case Diagram

RESULTS AND DISCUSSION

Data Model And Description	Accuracy	F1 Score
Plain CNN Model. Data directly passed to CNN model for Training and Prediction	84%	0.83
Logistic Regression	68%	0.76
Gaussian NB	50%	0.54
XGBoost Classifier	70%	0.75
Support Vector Machine using Entropy of Data	89%	0.88
Support Vector Machine using Fractal Dimension of Data	84%	0.84
Decision Tree using Entropy, Fractal Dimension, Kurtosis and Skew of Data	90%	0.91
Random Forest using Entropy, Fractal Dimension, Kurtosis and Skew of Data	95%	0.95
KNN using Entropy, Fractal Dimension, Kurtosis and Skew of Data	84%	0.83
ANN model using Entropy, Fractal Dimension, Kurtosis and Skew of Data	80%	0.80

CONCLUSIONS

Our Project indicates the potential use of EEG signals in depression detection. Many deep learning and machine learning models have been used to attain high performance. Our presented models have a minimum accuracy of 80% with combined accuracy (majority selection) results in an accuracy of >99% for the dataset. In the future, these models can be updated or improved to diagnose the severity of depression and its different stages and as a complementary tool or second opinion in the diagnosis of depression made by clinicians. However, a more accurate and robust model can also be developed using a larger dataset and to detect other neurological disorders at the early stages.

FUTURE SCOPE

This Project can be further expanded to diagnose a lot more Mental health disorders like:- Bipolar disorder, Post-traumatic stress disorder (PTSD), Obsessive-compulsive disorder, social anxiety disorders, panic disorders, phobias, Schizophrenia, borderline personality disorder, narcissistic personality disorder, antisocial personality disorder.

This Project can also help be further expanded to check a person's sleep pattern. It can also help in detecting whether a person is lying or not and can also be used to see what next word a person is going to say (mind reading).

This Project can also help medical professionals to evaluate medication given to a person with a certain mental disorder. It can be used to also evaluate in which part of brain changes occur after intaking the medication to understand what was the cause of the disorder.

REFERENCES

Analysis of Pisarenko Harmonic Decomposition-based sub Nyquist Rate Spectrum Sensing for Broadband Cognitive Radio (K. Chandrasekhar, Hamsapriye, and V.K. Lakshmeesha)

Deep learning for Electroencephalogram (EEG) classification tasks: A review

Deep Learning With Convolutional Neural Networks for EEG Decoding and Visualization

EEG-based workers' stress recognition at construction sites

An EEG Signal Processing Framework to Obtain High-Quality Brain Waves from an Off-the-Shelf Wearable EEG Device

Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals

EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks:A review

Ensemble approach for detection of depression using EEG features (Egils Avots, Kla̅vs Jermakovs, Maie Bachmann, Laura Paeske, Cagri Ozcinar, and Gholamreza Anbarjafari)

Automated major depressive disorder detection using melamine pattern with EEG signals (Emrah Aydemir & Turker Tuncer & Sengul Dogan & Raj Gururajan & U. Rajendra Acharya)

DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression (Geetanjali Sharma, Abhishek Parashar, Amit M. Joshia)

EEG Signal denoising using hybrid approach of Variational Mode Decomposition and wavelets for depression (Chamandeep Kaur , Amandeep Bisht , Preeti Singh , Garima Joshi)

DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG (Ayan Seal , Senior Member, IEEE, Rishabh Bajpai , Jagriti Agnihotri , Anis Yazidi , Senior Member, IEEE, Enrique Herrera-Viedma , Fellow, IEEE, and Ondrej Krejcar)

Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains (Amjed S. Al-Fahoum and Ausilah A. Al-Fraihat)

Depression Diagnosis by Deep Learning Using EEG Signals: A Systematic Review

[25] Yasin, S., Hussain, S. A., Aslan, S., Raza, I., Muzammel, M., & Othmani, A. (2021). EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks: A review. Computer Methods and Programs in Biomedicine, 106007.

Machine learning approaches in Detecting the Depression from Resting-state Electroencephalogram (EEG): A Review Study

A comparative analysis of signal processing and classification methods for different applications based on EEG signals (AshimaKhosla, PadmavatiKhandnor, TrilokChand)