MENTAL DEPRESSION DISORDER PREDICTION USING MACHINE LEARNING APPROACH

Abstract

Depression is a medical disorder that disturbs the positive thinking of a person, how they feel and act. It disturbs the peace of an individual and leads the individual to physical and emotional problems which in turn affect their health. A depressive person develops aversion to living and is prone to suicide in most cases. It is observed that depression is most likely to appear during the late teens to mid - 20's and women are more prone to depression than men. The reasons for depression may be many but one of the most significant one is the usage of social media. According to a study given by WHO it is estimated that depression affects 3.8 % of total population which includes 5.0 % among adults and 5.7 % among people above 60 years [1]. There has been tremendous increase in using Machine Learning (ML) for medical diagnosis. Most of the prediction models are based on imaging. This study aims to deploy a Machine Learning (ML) technique to measure DTI MRI scans of mood affected and non – affected individuals.[11] The Process is done using supervised methods of machine learning and support vector machine algorithm is used. This helps us in choosing the algorithm that provides the best accuracy. A large dataset is trained and classifiers are tested to distinguish affected and non – affected individuals based on the DTI MRI scans produced.

Keywords: Machine Learning (ML), Support Vector Machine (SVM), DTI MRI scans, Supervised Learning (SL), Dataset.

Authors

B Kamala
Department of Information Technology
Sri Sairam Engineering College
West Tambaram, Chennai, TamilNadu
India
kamala.it@sairam.edu.in

J Ranjani
Department of Information Technology
Sri Sairam Engineering College
West Tambaram, Chennai, TamilNadu,
India
ranjani.it@sairam.edu.in

K. R Saradha
Department of Information Technology
Sri Sairam Engineering College
West Tambaram, Chennai, TamilNadu,
India
saradha.it@sairam.edu.in

K. S Mahanaga Pooja
Department of Information Technology
Sri Sairam Engineering College
West Tambaram, Chennai, TamilNadu,
India
sec19it142.it@sairamtap.edu.in

N Priyadharshini
Department of Information Technology
Sri Sairam Engineering College
West Tambaram, Chennai, TamilNadu,
India
sec19it147.it@sairamtap.edu.in

I. INTRODUCTION

One of the most common but serious mood disorder is Depression. The known types of depression are Psychotic depression, depression caused by Postpartum, Bipolar disorder, Persistent depressive disorder, Seasonal affective disorder. Psychotherapies, Brain Stimulation therapies are most commonly used effective therapies that help the affected person to heal faster. According to the current research, the cause for depression could be a combination of bio-mimic, genetic, psychological and environmental factors. As per the reports generated by Substance Abuse and mental Health Service Association in 2018, Adolescents suffered the highest rate of depressive episodes 14.4% and people aged 18 to 25 years suffered 13.8%. The lowest rate was recorded by people aged 50 years and above. In most cases Depression can lead to Suicides. Unlike certain disorders, depression can be treated. The clinical procedures are aided by the adoption of Machine Learning.

Machine Learning is a subsystem of Artificial Intelligence where outcomes are predicted more accurately by the software applications without being explicitly programmed. The data is fed into the machine as input and using the provided data, the machine gets trained and output is predicted. The very aim of machine Learning is to predict accurate results without human intervention. The Machine Learning algorithms are categorized as Un-Supervised Machine Learning, Supervised Machine Learning, Semi-Supervised Machine Learning and Reinforcement Learning. The aim of our work is to detect depression from the MRI-Scans using supervised machine learning algorithm.

II. LITERATURE SURVEY

Sudha K, Sreemathi S, Nathiya B, RahiniPriya D gathered a dataset from the person who were diagnosed clinically depressed. The classification model was constructed with ML algorithms includes Naïve Bayes, KNN, Random Forest and Decision Tree algorithms. They concluded that the Decision Tree algorithm predicted the outcome with more accuracy. [1]

Social media channels such as Twitter, Facebook and Instagram profiles were used by Hatoon AlSagri, Mourad Ykhlef which in turn used to collect the tweets and chats to classify whether the user is in depressed state based on text processing. Algorithms like SVM, Naïve Bayes and Decision Tree were employed. Results showed that SVM classifier outperformed than other models. [2]

Prajwal Kharel, Kalpana Sharma, Sunil Dhimal and Sital Sharma reviewed various ML algorithms to diagnose MDD by using neuroimaging dataset. [3]

Meenal J Patel, Alexander Howard, Aizenstein had a study on MRI modalities that are associated with depression and mental illness.[4]

A deep learning model was developed by Ezekiel Victoria, Zahra M. Aghajan, Amy Sewart and Ray Christian using video data, audio data and classified whether the participants were depressed or not. [5]

Marzieh Sadat Mousavian aimed for automated detection of discriminative features for MDD based on MRI data. The results concluded that Naïve Bayes classifier showed superior performance in comparison with other methods. [6]

Zhao Zhang, Guangfei Li, Yong Xu, Xiaoying Tang summarized the current advancement in research of Machine Learning and Deep Learning techniques for the classification human brain magnetic resonance images. Then those techniques were employed to some neurological and psychiatric diseases. [7]

Shuang Gao, Vince D. Calhoun, Jing Sui predicted the diagnostic biomarkers that diagnose MDD from Bipolar Disorder in the early stage. [8]

III. METHODOLOGY

The methodology employed in our work is as follows:

1. The data need to be preprocessed initially
2. The preprocessed data will be supplied to the ML model
3. The Flask web framework is then get the input from the user
4. The framework splits the file into Pickle and JSON file
5. The resultant model will be delivered to the user using the same Flask web framework

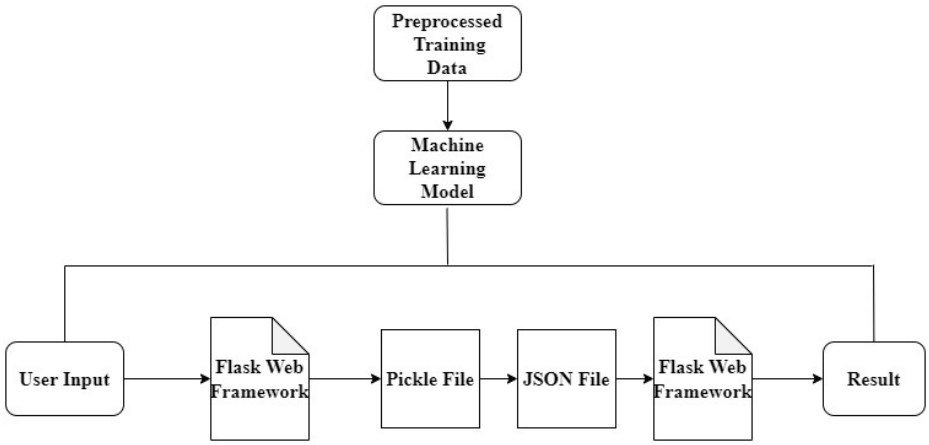

Figure 1: Working Model

IV. SYSTEM ARCHITECTURE

The following architecture explains how to detect depression using MRI Scans in a step by step process. First the MRI scan images of are given as input to the system. Then data preprocessing step is done with the MRI scanned images for removing noisy and unwanted image segments from the original image. Normalization and image resizing can be done in this data preprocessing phase. Then the image is splitted into various image segments.

Usually the images are stored in multidimensional format. Before the training process, the image needs to be flattened. The flattening process is used to convert the

multidimensional array into one dimensional array. Then the train-test split process is done. The commonly used 80:20 train-test split methodology is used to split the data into training and test set.

Model evaluation is done to evaluate the predicted model. Evaluating the model gives us some form of metrics to measure the model's performance. Based on the metrics and the accuracy level, mood affected and healthy and normal persons can be classified using the prediction results.

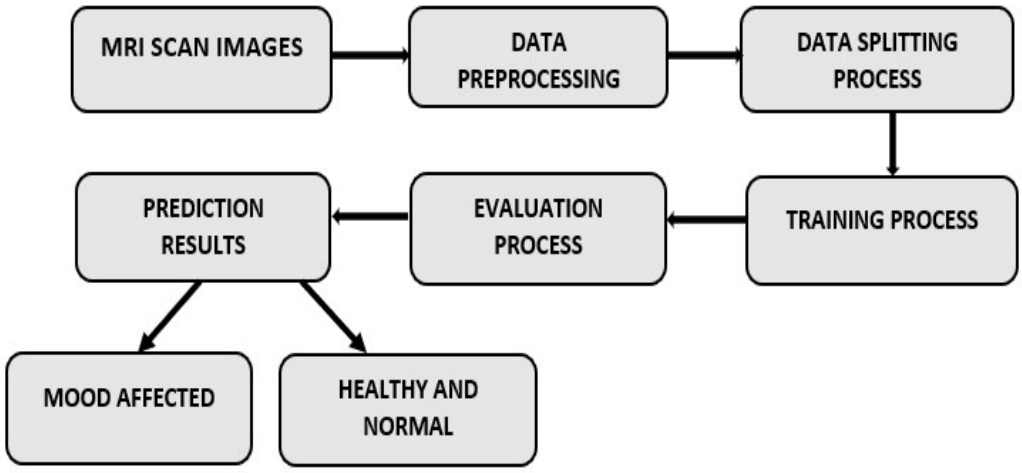

Figure 2: Depression Detection Methodology

1. **Dataset Collection:** The initial step in the machine learning pipeline is the collection of raw data for training the ML model. The precision of the results produced by the system depends on the data we feed to train them. The more the raw data, more precised are the results. Data collection allows us to keep track of past events so that we can utilize data analysis to uncover repeating patterns. Different machine learning algorithms are deployed to create predictive models that might help in forecasting the future changes and helps in finding trends based on these patterns. The predictive strength of the model lies on the data on which they are built. The accuracy of the high performing models is because of the uniform data collecting procedures. The information must be error-free (garbage in, garbage out) and pertinent to the task at hand. The dataset employed in this study contains MRI brain images of patients. The dataset is classified into two types: 'depressed' and 'normal'. In Python, imread() function imported from the mainstream library 'matplotlib', is used to read an image from a file into an array.

2. **Data Preprocessing:** To enable the ML model to easily parse the data, Data preprocessing is employed. It involves steps which we need to follow for data manipulation, encoding the raw data. It is not always the case where we get clean and formatted data. The data used in Real-world might contain unpleasant noise, void values, and is present in a format that cannot be used directly in machine learning models. Data preprocessing is a crucial step, which involves cleaning the data and making it fit for a

machine learning model to process it. Data preprocessing improves the model's accuracy and efficiency. The methods for preprocessing employed in this study are following.

3. **Image Resizing:** The images collected can be in different sizes which may cause difficulties in processing the data. So, the images have been resized to a particular size so that homogeneity is maintained across the data samples.

4. **Flattening:** Usually, real world images are in the form of 2-Dimensional arrays. They cannot be passed into a machine learning model. Here comes the role of flattening. Flattening is a technique used to convert multi-dimensional arrays to a one-dimensional array(1D array). This flattened data is then being fed into the model for classification.

5. **Train and Test Split:** Train and test split is used to measure the performance of any machine learning model. The train-test split will be useful for the regression and classification task in turn which will be more useful in supervised learning methods. The dataset is splitted into two sets called training and testing data set. The training set is the primary set which is used to fit the model. The secondary set is the test set which got input given to the model. After getting input to the test set, predictions are made and it compares them to the predicted values.

    - **Train Dataset:** Dataset for fitting the machine learning model.
    - **Test Dataset:** Dataset assess how well the machine learning model fits.

    We have used scikit-learn Machine Learning library (sklearn) to implement the test-train split procedure.

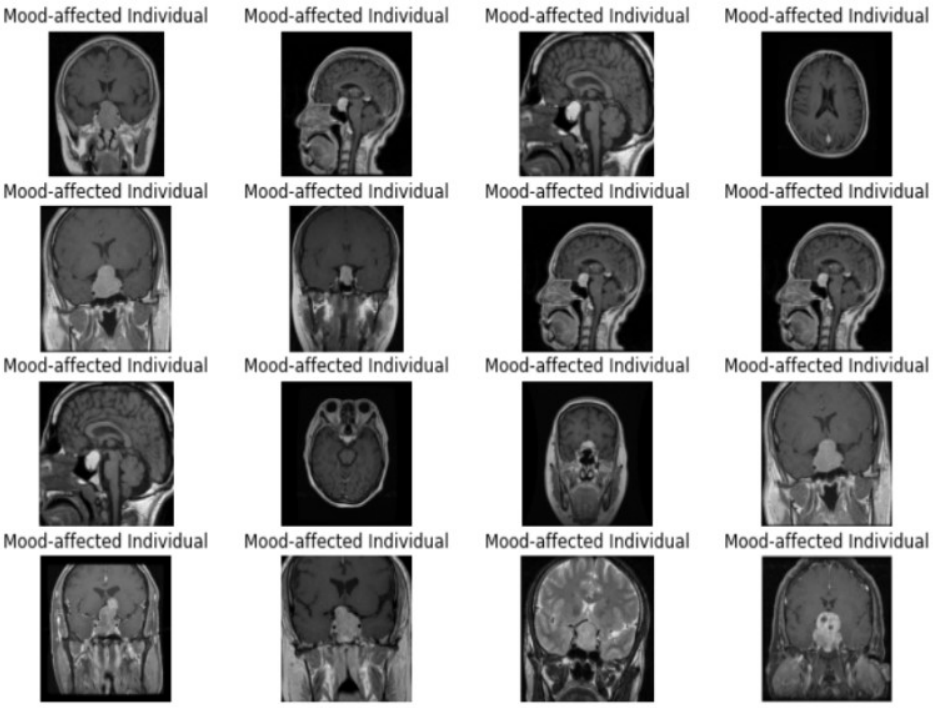

Figure 3: Mood-affected Individual

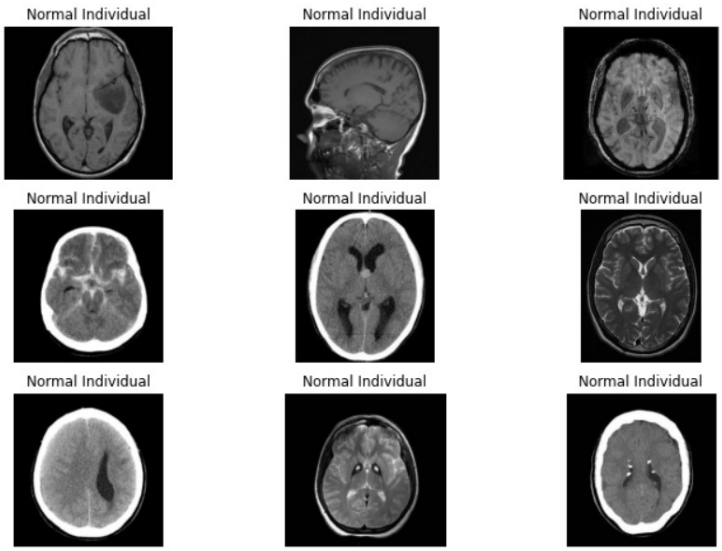

Figure 4: Normal Individual

V. CLASSIFICATION

Support Vector Machine (SVM) classifier and Logistic Regression supervised machine learning algorithms are used for Image classification. SVM is a Machine Learning Algorithm that is exclusively used for classification problems. It spots out a hyper-plane that separates two extreme classes in to N-dimensional space. This algorithm is applied by importing SVM from sklearn library. We have set three parameters namely C, kernel, and gamma values to perform the classification. These parameters are used to analyze the Image and classify the image into proper classification.

Logistic Regression is an algorithm that is mostly widely used for binary classification problems. C value for Logistic Regression is the inverse of regularization strength. It is set as a positive float value. The model is well trained with these algorithms using the dataset.

VI. PERFORMANCE EVALUATION

Performance evaluation is an integral part of model evaluation process. Model monitoring is used in performance evaluation to measure how well a model performs at the task for which it was intended. The accuracy scores are obtained for both the SVM and Logistic Regression algorithms and these scores are compared. A common way to assess model performance involves evaluating model accuracy. Accuracy refers to the percentage of correct model predictions, which is one approach to define machine learning performance.

The efficiency of SVM algorithm is found to be better comparatively to Logistic Regression. So we get higher and efficient results for depression detection using SVM algorithm with the accuracy of 97%. The performance of a predictive model is also calculated and evaluated by choosing the right metrics. Confusion matrix is one of the evaluation metrics which is a matrix or a table summarizes the performance of a model.

SCREENING TEST RESULTS	DISORDER	NO DISORDER
POSITIVE	TRUE POSITIVE(TP)	FALSE POSITIVE(FP)
NEGATIVE	FALSE NEGATIVE(FN)	TRUE NEGATIVE(TN)

Screening Test Results	Disorder	No Disorder
Positive	88	12
Negative	78	9

Figure 5: Confusion matrix

Here from the above figure, it is clear that the screening test results includes positive and negative as predicted values and the mood affected value Disorder and No Disorder is classified as actual values. Disorder is coming under Positive test result and No Disorder is under Negative screen result. The following confusion matrix made for a classifier that classifies the Normal and abnormal people based on screening test results. The accuracy, precision and recall for the above model is,

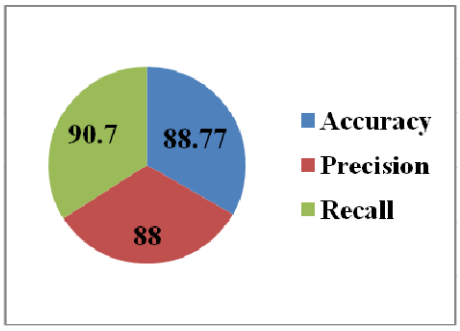

Figure 6: Accuracy, Precision and Recall for Depression Detection

VII. CONCLUSION

This application aims to detect the depression in an individual at an early stage with more accuracy. Early detection allows us to reduce risks and symptoms as they arise. The use of artificial intelligence with machine learning technology may help to automate the process and assist doctors for more prompt and better decision-making. The supervised models of machine learning methods like SVM and Logistic regression are used to predict the depression earlier. This application is trained with a large data set and the classifiers are improvised with the better accuracy. The suggested system's future advancements will concentrate on improving user accessibility for the application.

REFERENCES

[1] Sudha K, Sreemathi S, Nathiya B, RahiniPriya D (2020) "Depression Detection using MachineLearning", International Journal of Research and Advanced Development, Special Issue on AICTE Sponsored International Conference on Data Science & Big Data Analytics for Sustainability

[2] Hatoon AlSagri, Mourad Ykhlef (2020), "Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features", IEICE Transactions on Information and Systems, Volume E103.D (2020) Issue 8

[3] Prajwal Kharel, Kalpana Sharma, Sunil Dhimal and Sital Sharma (2019) "Early Detection of Depression and Treatment Response Prediction using Machine Learning: A Review", 2019 Second International Conference on Advanced Computational and Communication Paradigms

[4] Meenal J Patel, Alexander Howard, Aizenstein (2016) "Studying depression using imaging and machine learning methods", NeuroImage: Clinical, Volume 10

[5] Ezekiel Victor, Zahra M. Aghajan, Amy Sewart and Ray Christian (2019), "Detecting Depression Using a Framework Combining Deep Multi modal Neural Networks with a Purpose-Built Automated Evaluation", Psychol Assess, Volume 8

[6] Marzieh Sadat Mousavian (2021) "Machine Learning Methods for Depression Detection Using SMRI and RS-FMRI Images", LSU Doctoral Dissertations. 5546.

[7] Zhao Zhang, Guangfei Li, Yong Xu, Xiaoying Tang (2021) "Application of Artificial Intelligence in the MRI Classification Task of Human Brain Neurological and Psychiatric Diseases: A Scoping Review", Diagnostics (Basel), Volume 8

[8] Shuang Gao, Vince D. Calhoun, Jing Sui (2018) "Machine learning in major depression: From classification to treatment outcome prediction", CNS Neuroscience and Therapeutics, Volume 11

[9] SureshKumar M, Naik, D.A., Mohana, R.M., Ramu, G. (2022), "Analyzing histopathological images by using machine learning techniques", . Applied Nanoscience, Volume 13, Issue 3.

[10] T.Sheela and N.Sumathi (2017), "An Efficient Sentiment Analysis by Using Hybrid Naïve Bayes and SVM Approach in Banking Institution", International Journal of Civil Engineering and Technology, Volume 8, Issue 12

[11] Kalaiselvi, V.K.G., Shanmugasundaram, H., Aishwarya, E., Nandhini, C, Bhuvaneshwari, S.J.(2022), "Analysis of Pima Indian Diabetes Using KNN Classifier and Support Vector Machine Technique", Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems

[12] T.Sheela and M.S.Abirami (2017), " Improved Speedup Performance in Automated Segmentation of Kidneys on Abdominal CT Images and Extracting its Abnormalities", Volume 116 No. 21.

