e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

DEPRESSION DETECTION AND MENTAL HEALTH TRACKER USING MACHINE LEARNING

Sunil D. Kale*1, Gaurav Rasal*2, Pratiksha Ganjave*3, Rushikesh Markad*4, Yash Kalekar*5

*1Professor, Department Of Computer Engineering, Pimpri Chinchwad College Of Engineering, Pune, India.

*2,3,4,5Student, Department Of Computer Engineering, Pimpri Chinchwad College Of Engineering, Pune, India.

ABSTRACT

Depression is very commonly found and a quite serious type of mental disturbance. The thing about depression is it's one of the curable medical illnesses but the worst part is that the people who are suffering from depression won't get to know that they are suffering from some sort of mental illness. So, a system to predict depression in its early stage can be a savior for people and they can consult a psychologist and find a suitable cure for it.

Looking towards the significance of the problem statement, we have done an extensive survey of recently published research papers. The analysis of the papers studied is presented in this report with features, algorithms, and datasets utilized for experimentation.

Moreover, it is observed that the systems or applications developed previously consider or rely on only one type of input like audio, video, or textual data from any of the social media platforms. So, the results predicted by existing systems are based on one type of input media. To enhance the result of accuracy in this report we are proposing a framework with a system design that will consider at least two input media. This will help the model to accurately predict the mental health condition of the user.

Keywords: Mental Health, Depression, Mental balance, Increased productivity, Psychiatrist, Psychology, Machine Learning, Natural Language Processing (NLP), Deep Learning, Text Processing.

I. INTRODUCTION

Millions of individuals all around the world suffer from depression, which is a widespread mental health issue. In order to ensure that people receive the assistance and care they need, depression must be recognized and treated. Therefore, enhancing patient outcomes requires the creation of efficient systems for depression identification and mental health surveillance. This study intends to investigate the application of methods for tracking mental health and detecting depression in clinical settings. The study will discuss numerous instruments used to evaluate depression, including physiological tests, clinical interviews, and self-reported questionnaires. The article will also discuss how people may monitor their mood, activities, and other elements that could affect their mental health by using applications that track their mental health. Around 16.6% of people will experience some form of depression at some point in their lives. Depression, also known as Major Depressive Disorder (MDD), is a mental illness that affects people all over the world. It is a major mental illness that affects one in every 15 adults annually or 6.7% of the population. Depression can be caused by a variety of factors, including excessive thinking, financial loss, unemployment, social isolation, and more. The fact that things are not developing or transpiring as anticipated is one of the primary causes. Young adults in their mid-20s and teens are particularly susceptible to depression. Depression decreases productivity, saps motivation for routine work, makes one feel unhappy and unworthy, impairs thinking and concentration, and alters appetite. These people's behavior also has an impact on the neighborhood in which they work or reside. People who are depressed hurt themselves, and in the worst situations, depression can cause them to kill themselves. In order to solve this issue, which is the main goal of this study, we propose a model that can identify depression in its early stages, at which point it may be treated and individuals can have happy lives. This suggested model may be made compatible with smartwatches, fitness bands, and mobile phones, which are standard conveniences for most individuals. The ultimate goal of this study work is to add to the expanding

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

body of literature on mental health monitoring and depression identification while also shedding light on the usefulness and possible therapeutic uses of these instruments.

II. LITERATURE REVIEW

The paralinguistic characteristics of depressed persons differ from those of non-depressed people, according to research by writers in [1]. They thus examined these traits. The initial step was data collecting and preprocessing. By using the MFCC, features were extracted from the audio file data. The classification model, which was implemented using CNN, was then given the MFCC file.

In [2], writers created an online model that enables the examination of several user characteristics with regard to two specific mental diseases. That model also offers basic outputs that can be used to build additional composite models for virtually understanding a user's mental health. The model can be used in advance to get more information from users through feedback.

In order to predict depression among social media users, the authors of the research [3] examined and analyzed a number of algorithms for analyzing users' texts and postings on social media. Natural Language Toolkit components as well as Peter Norvig's code for spelling correction were applied. A 10-fold CV technique was used to conduct the trials. The researchers came to the conclusion that Shen et al.'s dataset, which was utilized along with Eye's dataset, predicted depression more correctly and was therefore a superior option. They also found that some dataset modifications were necessary to lessen model overfitting.

Authors in [4] have suggested a deep learning-based automated approach for detecting Major Depressive Disorder (MDD). Convolutional Neural Networks (CNN) spectrogram images were used to put up the model. They were able to achieve a classification accuracy of 99.58 percent by employing the hold-out validation approach. This model's weakness was that it was developed using just 64 individuals as training data, of which 34 were MDD patients and 30 were healthy individuals.

According to a study in [5] the authors proposed the MHDeep framework. By merging information gathered from commercial warehouse management software with the knowledge-purifying capabilities of DNNs, the system enables continuous and universal diagnosis of schizoid personality, major depressive disorder, and bipolar disorder. To notify users when sizable datasets are unavailable, the concept uses a synthetic data production module. The model's architecture and strength training during training were chosen using iterative pruning and growth techniques. They tested the model using information gathered from 74 individuals. The results of the trial demonstrated the models' effectiveness in terms of computation. The MHDeep's computational complexity, efficiency, and precision make it suitable for regular diagnostics and daily monitoring.

The researchers examined online communities for people with depression and took into account how these communities differed from other online communities in the publication [6]. The contexts under investigation included affect, psycholinguistic processes, and content-specific concerns. To distinguish between depressed and controlled communities in online messages, machine learning and statistical techniques were used. It was found that there were significant differences between these two groups in every component of the study.

A novel stratified DL network has been proposed by the researchers in [7] that includes numerous linked layers to concentrate end user posting and people's recognisable depiction. Their approach used online social media data as an input for depression diagnosis, which was further divided by summarizing relevant users' social media post histories to axiomatically pick the markable user data. One benefit of the axiomatic summarization was the ability to limit the model training to the most observable data, which was helpful in reducing the Hughes phenomenon issue. Which helps the model focus on the task for which it provides inputs rather than just detecting depression by better providing contingent knowledge or information.

In [8], the authors collected data from social media for sentiment analysis and applied artificial intelligence algorithms to detect gloom and sadness. They noted that Deep Learning with Multi Class Classification demonstrated a greater precision value for exploration. To comprehend the emotions, discernment, and replications they are expressing, the gathered data was utilized. These sentiments were examined in order to forecast human behavior. Three kinds of positive, negative, and neutral data were created.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

The researchers came to the conclusion that the emojis and emoticons available in the data serve as the main indicators for forecasting users' emotions.

Through the use of social media, they provided a comprehensive view of mental health computing in the paper [9]. They methodically investigated the problems related to the two well-known mental health concerns of stress and depression in this. They created a collection of benchmark datasets from the actual world, identified various distinctive feature groupings, and put forth a number of useful detection models. They set up in-depth data analysis to expose the online behaviors that lie beneath the mental health issues.

The authors of [10] exploited the sociological behaviors and perspectives of Twitter users to make depression predictions. The Patient Health Questionnaire-9 had been completed by the user within the previous two months. The last measurement was taken with this. To enhance the performance of their model, they used the SVM-RFE and ANOVA feature selection techniques. Among the other machine learning techniques, the Random Forests technique also provided the highest level of prediction accuracy. Their model's drawback is that it only analyzed data from Twitter users who have consented to the model's dissemination.

They discussed various machine-learning methods and algorithms for depression prediction in [11]. In this study, both depressed and non-depressed people's verbal and non-verbal behaviors were taken into account. They discussed data collection and data manipulation techniques for creating a prediction model. The concept was useful for remote evaluation, assistance systems, and awareness, but it was not standalone diagnostics. It offers a lot of intuition while also acknowledging a lot of open-ended research issues.

III. EXPERIMENTAL ANALYSIS

1. Dataset Used

A survey of people was conducted by asking them a set of questions about their family environment, family history regarding mental health, employment status, workplace environment, work culture, the behavior of their coworkers, etc. We used the data collected from this survey as a dataset for our analysis.

2. Data Cleaning

Data cleaning is the process of identifying and correcting errors, such as missing values, duplicates, and outliers, in the dataset. Common techniques for data cleaning include imputation, removing duplicates, and removing outliers. The unnecessary columns like address and date won't significantly contribute to predicting the result. Hence, they were removed from the dataset. There were irregularities present in the data such as there were different entries for the field gender(male, female, maleish, trans, male, phemale). Such values have been standardized into a usable format(male, female and trans). The null entries are taken care of by replacing them with the median as the data is skewed. Also, the null values in the form of strings were replaced by the mode of that column. Further scaling and normalization of some columns were done to make the data more suitable for prediction.

3. Data Visualization

After data pre-processing the data visualization is done to understand the relationship and distribution of data in the dataset.

1) Categorization plot

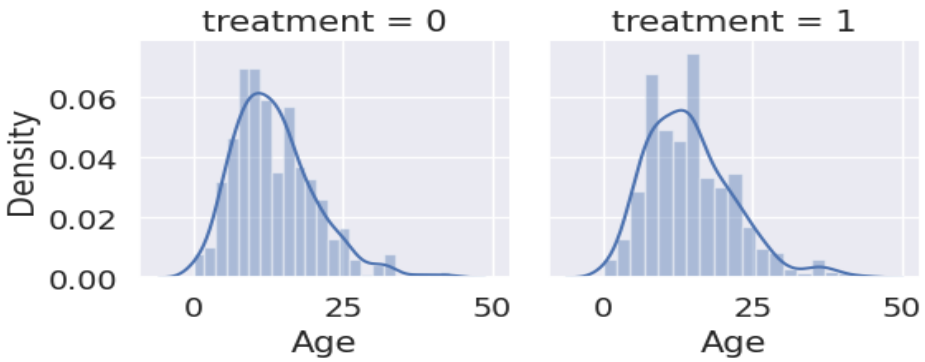

Fig 1. Categorization Plot

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

2) Heatmap

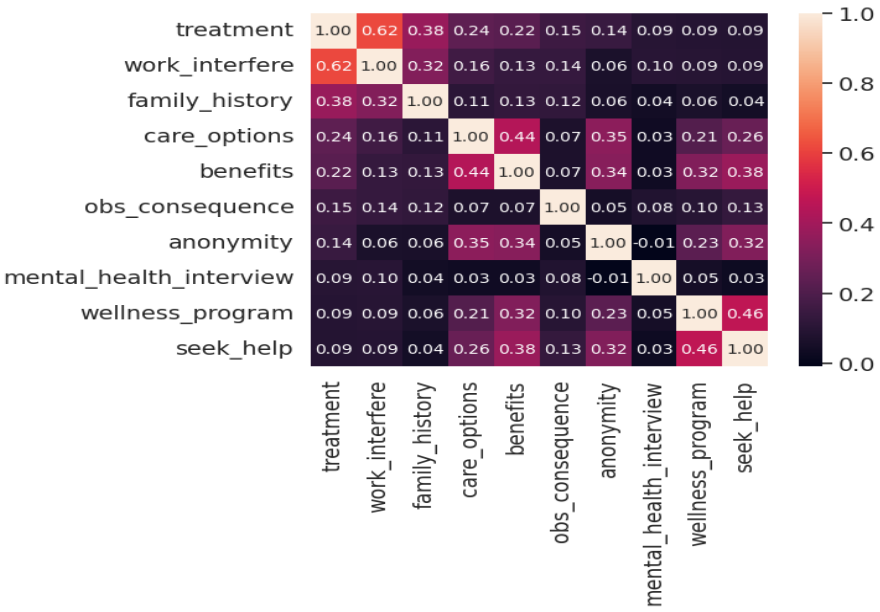

Fig 2. Heatmap

3) Distribution Plot

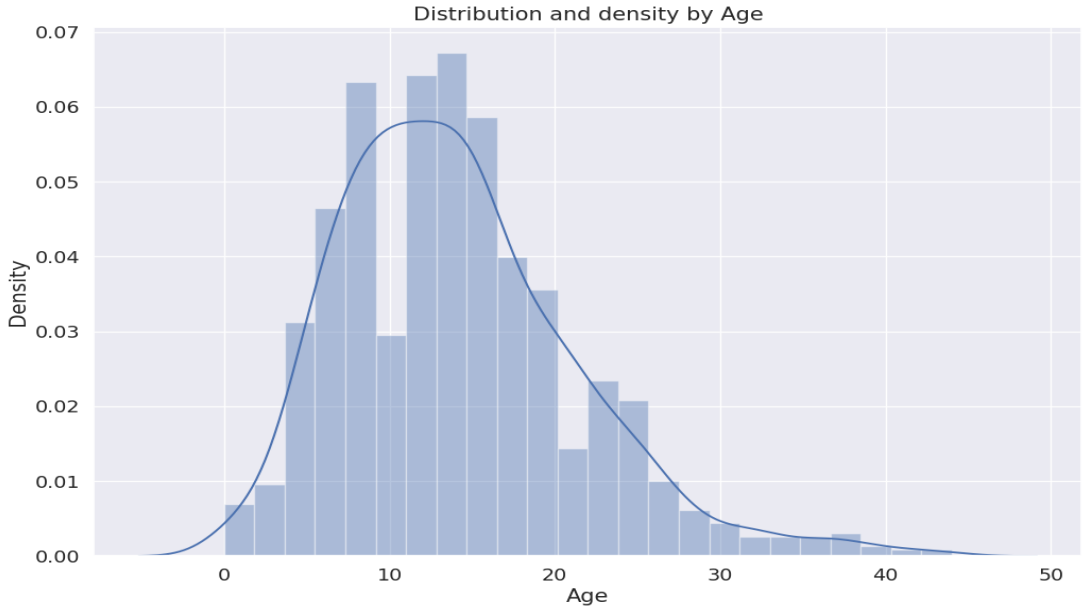

Fig 3. Distribution Plot

4. Feature selection

Feature selection is selecting the features from a given dataset that will significantly contribute to the end results. It is done to reduce the high dimensionality of the dataset which can lead to underfitting of the model. Common techniques for feature selection include correlation analysis, recursive feature elimination, and principal component analysis. In our model, we performed feature selection to reduce the dimensionality of the dataset and focus on some specific features of the dataset that improve the performance of the model.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

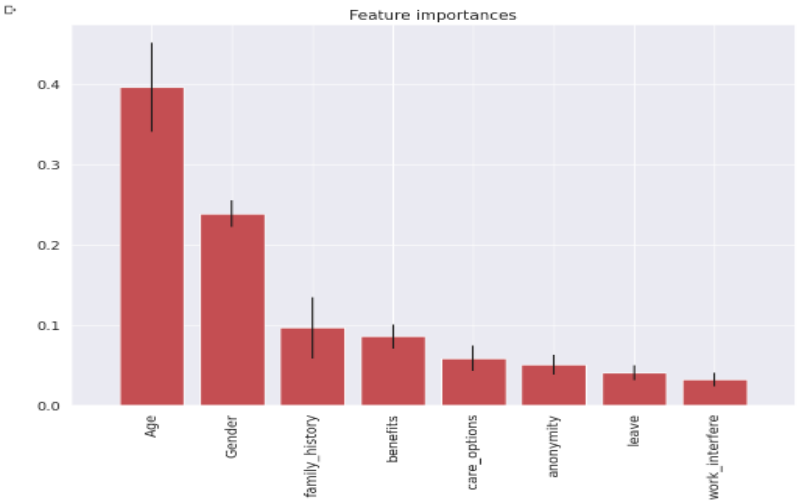

Fig 4. Feature Selection

5. Algorithms analyzed

1) Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. In our model we implement Random Forest Classifier method to calculate the best parameter to predict the output.

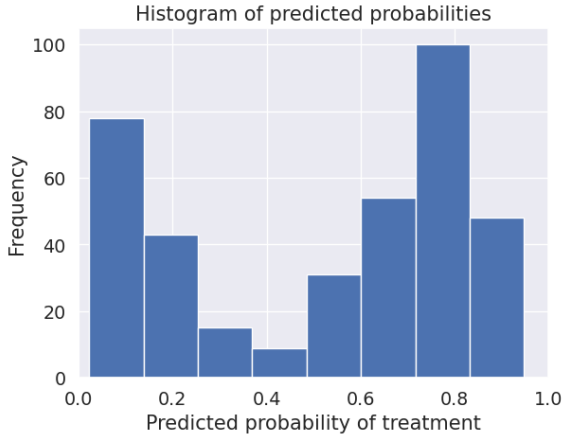

Fig 5. Predicted probabilities of Random Forest

2) Decision Tree classifier

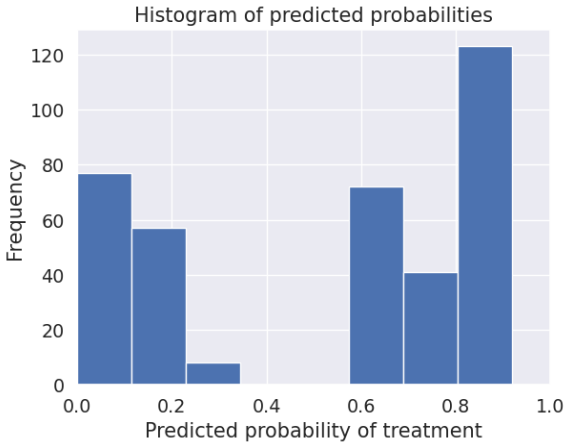

Fig 6. Predicted probabilities of Decision Tree Classifier

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

A decision tree classifier is a popular supervised learning algorithm in machine learning that is used for both classification and regression tasks. Decision tree classifiers are easy to interpret and visualize, which makes them popular in many applications. In our model we train a decision tree by using Decision Tree Classifier method. and then making predictions for testing data.

3) K-Nearest Neighbors

The k-nearest neighbors (KNN) algorithm is a type of supervised learning algorithm used for both classification and regression tasks in machine learning. The KNN algorithm is a simple yet effective algorithm that is easy to implement and works well for datasets with a small number of features or low-dimensional data. In this model we perform the KNeighbors Classifier method to predict the output.

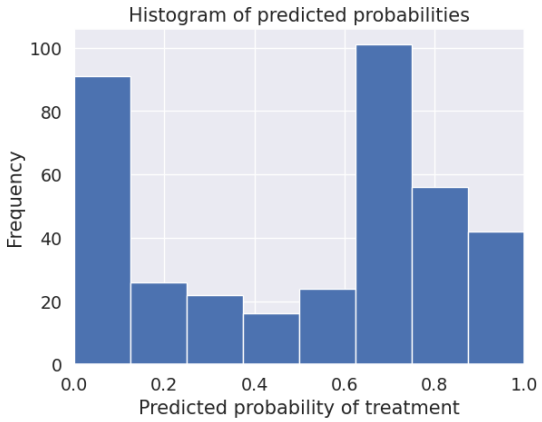

Fig 7. Predicted probabilities of K-Nearest Neighbors

4) Bagging

Bagging, short for bootstrap aggregating, is a technique in machine learning that is used to improve the accuracy and stability of a model by training multiple copies of the same algorithm on different subsets of the training data and combining their predictions through an ensemble. In this model it implement the Bagging Classifier method make a prediction for a new data point present in dataset.

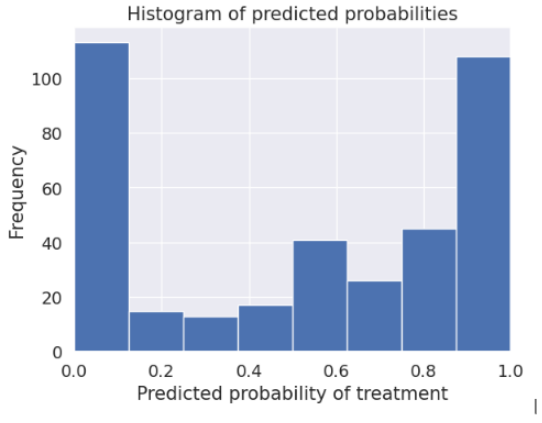

Fig 8. Predicted probabilities of Bagging

5) Boosting

Boosting is a machine learning ensemble technique that involves combining multiple weak models to create a strong model. Boosting is effective because it can reduce bias and variance in the final model, resulting in better performance. So, in our model we perform boosting by using Ada Boost Classifier method to remove bias so that we get more accurate result.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

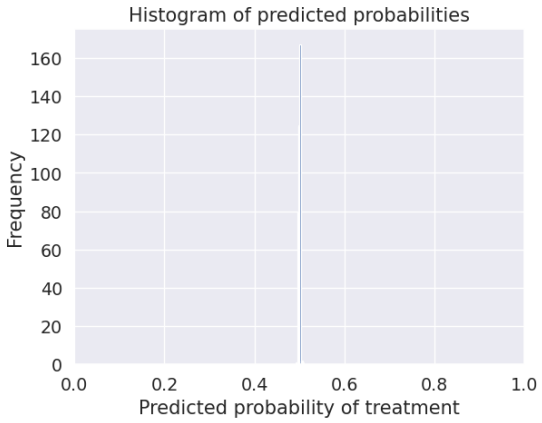

Fig 9. Predicted probabilities of Boosting

6) Stacking

Stacking is a machine learning ensemble technique that involves training a meta-model to combine the predictions of several base models. Stacking is a powerful technique because it can improve the accuracy of predictions by combining the strengths of different models. We use stacking method to improve the accuracy of our model.in which we perform building and fitting of data using K Neighbors Classifier and Random Forest Classifier then making class prediction for the testing set.

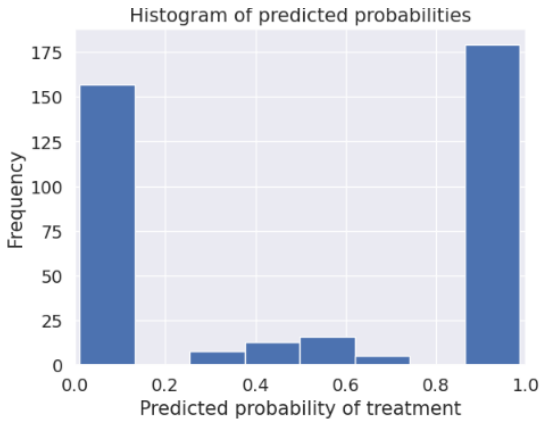

Fig 10. Predicted probabilities of Stacking

7) Neural Network

A neural network is a type of machine learning model that is inspired by the structure and function of biological neurons in the brain. During the training phase, the neural network adjusts its weights and biases to minimize the difference between its output predictions and the ground truth labels.in this model we use neural network to improve the accuracy of model. In which hidden layer perform some specific operation to make model more accurate.

IV. RESULT

After performing various algorithmic analysis on the available data we came to the conclusion that amongst all the analyzed algorithms the stacking of Random Forest, K-Nearest Neighbors and Gaussian NB with Logistic regression as stacking classifier gave the best results. .Also, we worked with various Machine Learning algorithm to get the accurate result following table shows that Accuracy of each method as below:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

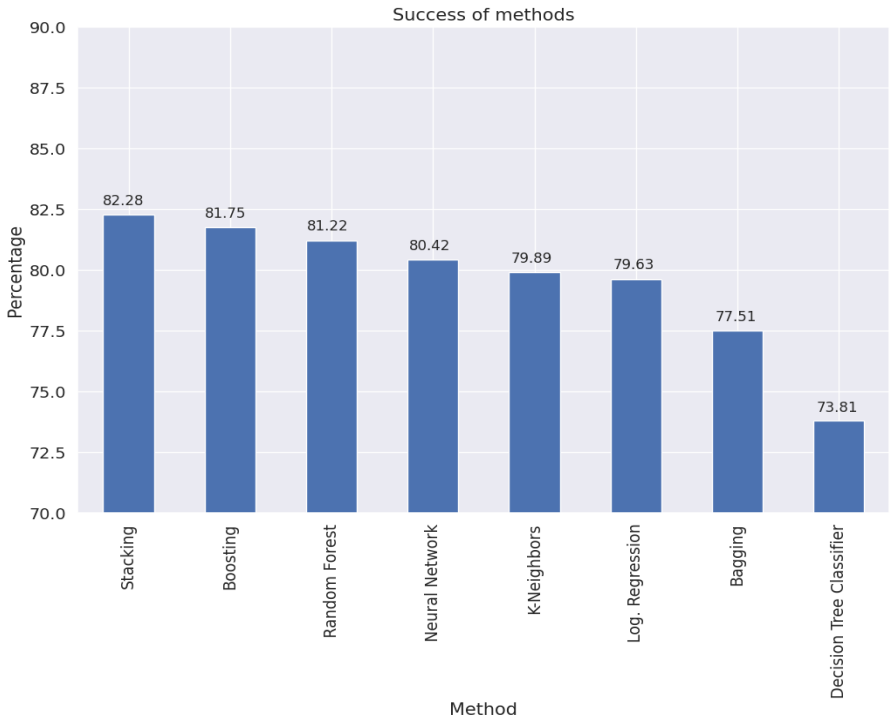

Fig 11. Graphical representation of Methods

Table 1. Accuracy of Methods

Sr No	Methods	Accuracy	Precision	AUC Score
1.	Logistic Regression	80	76.44	79.68
2	K Neighbors Classifier	79.89	74.88	79.9
3	Tree Classifier	80.69	-	-
4	Random Forest	81.22	75	81.34
5	Bagging	78.04	75.24	78.09
6	Boosting	81.74	76.10	81.85
7	Stacking	82.27	80	82.3

V. CONCLUSION

The current state of mental health is a delicate and important issue. It's essential for leading a balanced, healthy lifestyle. One's thoughts, deeds, and emotions are all influenced by their mental state. It might affect a person's effectiveness and productivity. This paper describes our method for identifying depression symptoms using machine learning technology and potentially useful algorithms. Our ML systems have been successful at identifying a user's mental state and recommending tasks to help the user recover from it.

VI. FUTURE SCOPE

By studying and analyzing all the work done in the field of depression and other mental issues detection using Machine Learning, Deep Learning and CNN we came to know that the researchers have focused on only one type of input parameter for the prediction. Two or more parameters discussed in the research mentioned above

(textual, audio, video or the EEG signals and other medical documents) can be combinedly taken as an input from the user to achieve greater accuracy. Also, the datasets for audio-visual data in this field are not available for studying which can be an area of interest for many of the researchers.

VII. REFERENCES

[1] Depression Detection in Arabic Using Speech Language Recognition Zainab Alsharif, Salma Elhag, Sulhi Alfakeh

[2] MIDAS: Mental Illness Detection and Analysis via Social Media Elvis Saravia, Chun-Hao Chang, Renaud Jollet De Lorenzo, Yi-Shin Chen

[3] A textual-based featuring approach for depression detection using machine learning classifiers and social media texts,Raymond Chiong a,*, Gregorius Satia Budhi a,b,**, Sandeep Dhakal a, Fabian Chiong c

[4] Decision Support System for major depression detection using spectrogram and convolution neural network with EEG signals,Hui Wen Loh1, Chui Ping Ooi1, Emrah Aydemir4, Turker Tuncer5, Sengul Dogan5, U Rajendra Acharya1,2,3,6,7,*

[5] Affective and Content Analysis of Online Depression Communities Thin Nguyeny, Dinh Phungy, IEEE Member, Bo Daoy, Svetha Venkateshy, IEEE Fellow, Michael Berkzy

[6] MHDeep: Mental Health Disorder Detection System based on Wearable Sensors and Artificial Neural Networks Shayan Hassantabar, Joe Zhang, Hongxu Yin, And Niraj K. Jha, Princeton University, USA

[7] DepressionNet: A Novel Summarization Boosted Deep Framework for Depression Detection on Social Media Hamad Zogan, Imran Razzak, Shoaib Jameel, Guandong Xu

[8] Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review Nirmal Varghese Babu1 · E. Grace Mary Kanaga1

[9] Mental Health Computing via Harvesting Social Media Data. Jia jia.

[10] Big data analytics on social networks for real-time depression detection Jitimon Angskun1,2, Suda Tipprasert1 and Thara Angskun1,2*

[11] On the review of image and video-based depression detection using machine learning Arselan Ashraf1 , Teddy Surya Gunawan2, Bob Subhan Riza3, Edy Victor Haryanto4, Zuriati Janin5