EasyChair Preprint

№ 11626

Mental Health Prediction Using Artificial Intelligence

Mrinmayee Deshpande, Pradnya Mehta, Nilesh Sable, Utkarsha Baraskar, Ishika Ingole and Vaishnavi Shinde

December 27, 2023

Mental Health Prediction Using Artificial Intelligence

Mrinmayee Deshpande[1], Pradnya Mehta[2] , Nilesh Sable[3], Utkarsha Baraskar[4], Ishika Ingole[5] ,Vaishnavi Shinde[6]

[1] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.
[2] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.
[3] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.
[4] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.
[5] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.
[6] Vishwakarma Institute of Information Technology, Kondhwa, Pune, Maharashtra, India, 411048.

Abstract

Mental Health Disorders have become a significant public health concern worldwide, necessitating accurate and timely diagnostic methods. This study aims to predict the type of Mental Disorder using Artificial Intelligence specifically, the Random Forest Algorithm which is known for its effectiveness in classification tasks. The motivation for this study is lack of a model which can accurately predict the type of mental health disorder of any person. The main objective of 'mental health prediction' is to predict the mental health of patient on the basis of symptoms only and diagnose the exact disease in order to resolve the serious issues related to mental health which are ignored by society by considering disturbed mental health as a taboo. This paper makes a survey of various mental health symptoms and problems related to it in our society which are solved using AI technologies. To test the performance of our proposed system we used several machine learning algorithms like Support Vector Machines (SVMs), Random Forest (RF) Algorithms. Here, these Algorithms are mainly used for diagnosing mental health disorders on the basis of given input (i.e. verified dataset of symptoms). The Random Forest Model achieved an overall accuracy of 95% in predicting the type of the mental disorder. Gain in the values of Precision, Recall and F1 – Score was also noted. This model is basically a chatbot which predicts accurately the type of mental disorder of a person, if any. We can expect outcomes such as early detection of any mental disorder, facilitating all self-diagnosis through this bot, free interaction of the patients with the bot, etc. through this model.

Keywords: Chatbot, Random Forest Algorithm, Support Vector Machine Algorithm

I. Introduction

Mental illness is on the verge of becoming pandemic in developing countries like India, posing difficulties among the people [1]. After the COVID 19 pandemic people from distinct age groups, distinct professions, distinct socio-economic groups, males, females and neutral genders also are reported to be suffering from various mental disorders. According to the report by World Health Organization (WHO) released on 17 June 2022, we realized the severity of the mental health issues affecting the lives of the people. WHO defines mental health issues as characterized by a combination of negative thoughts, feelings, depression, rapidly increasing competition throughout the world, increasing demands from various factors of the society and interpersonal connections. In addition the rapidly changing social scenario is affecting the mental health in a wider range. It can have a bad impact on person's daily life and relations.

One major difficulty the society faces is that people try to hide such disorders. They are not ready to accept that they have any sort of mental illness. They try to avoid approaching mental hospitals on a large scale. Even if they approach, they try to hide the facts. This problem motivated us to develop an AI based tool to overcome human interference and help people express themselves in a secure way at their leisure anywhere anytime. Further, the bot is user friendly, so the people will find it very easy to deal with.

The key objectives of this research are as follows –

1. To contribute to the scientific community's understanding of how artificial intelligence can be used to predict mental health issues.
2. To explore new machine learning algorithms or using existing ones in innovative ways.
3. To improve mental health outcomes by developing more accurate and reliable prediction models.

4. To identify mental health issues at the earliest and provide more effective treatment.
5. This research paper also helps raise awareness about the potential of artificial intelligence in mental health care. This could increase interest and investment in the field, leading to further advancements and improvements.
6. The use of AI in mental health is a relatively new field, and publishing a research paper can help advance this area of research.
7. The ultimate goal of mental health prediction using AI is to improve mental health outcomes.

AI is nowadays used to reduce human errors in medical field. This helps to improve the efficiency of diagnosis. Machine Learning is useful to test simultaneously several input variables and can test their interactions in various ways. It can test all possible combinations and hence provides us with the most accurate output[5].This article gives a general review of AI and ML and also helps in analyzing the present state and future scope of using ML in mental health prediction in India.

A dataset consisting of mental health-related features, collected through surveys, medical records, and self-reports, was used to train and evaluate the predictive model. The dataset included various mental disorders such as Anxiety, Bipolar Disorder, Major Depressive Disorder and Schizophrenia. Also it had labels like No Disease for the healthy individuals who had come for check- up. The data is collected primarily i.e. by visiting the hospitals in Ahmednagar locality. While providing the data the hospitals denied to disclose the identity of the hospitals as well as the patients, it being confidential. The dataset did not contain the names of the patients in order to assure their privacy. Instead they provided the data in the form of patient numbers like P1, P2, etc. So the privacy of the hospitals and the patients is secure.

The dataset we used consists of the data of 1000 patients. It consists of various attributes such as age, sex, profession, 24 symptoms and the corresponding disorders. This research does not focus on characteristics like age, sex, and career because the goal of the paper is to develop a bot that can properly predict the disorder using the symptoms.

While building the model two machine learning algorithms were implemented like Random Forest (RF) and Support Vector Machine (SVM). The implementation of both the models was compared in all respects. The accuracies of both were also compared. Various parameters like precision, recall and f1- score for both the models were calculated and compared. We tried to opt for a model which could provide maximum accuracy.

The following are the results of applying artificial intelligence to predict mental health as early as possible: -

1. We can achieve improved accuracy in predicting mental health conditions.
2. Early detection and intervention can be made easy.
3. AI algorithms may be able to create individualized treatment regimens for people with mental health disorders by examining data from a variety of sources (such as medical records, social media activity, etc.). This could improve the effectiveness of treatment and reduce the risk of negative side effects.

The remaining sections are arranged as follows –

The literature on AI applications in mental health is summarized in Section 2 along with related studies.

The data collecting, preparation, and model implementation procedures are all covered in Section 3 of the research methodology.

The predictive model's performance evaluation and findings are presented in Section 4.

Section 5 discusses the implications of the findings and explores potential directions for further research.

The paper is finally concluded in Section 6, summarizing the contributions and key takeaways of this study.

II. Literature Survey

The problem of unhealthy psychological condition is very common. It affects the performance of a person which disturbs his or her mental health. This has turned to be a serious issue and in the upcoming period it will affect the society on a large scale. To overcome such issues it is necessary to look upon them with an appropriate solution. For fulfilling this objective detecting the exact mental health disorder is first step. Thus we are looking forward to do same using artificial intelligence (AI) and machine learning (ML) techniques which are most recent and rapidly evolving.

The study of ‘mental health prediction’ assessed the detailed study of various machine learning algorithm which detection the mental health disorder on the basis of given input (dataset). We have studied the machine learning algorithms like support vector machine (SVM) and random forest algorithms. On reviewing and analyzing near about 15 research papers, we concluded that random forest machine learning algorithm has the highest accuracy.

Table 1 - Literature Survey

Author	Study Aim	Results

Oliver Higgins RN, Brooke L, Stephan K [1]	Data abstraction and synthesis	Use of AI/ML based DSS in mental health
Nor Safika Mohd Shafiee, Sofianita Mutalib [2]	SVM, KNN, Naive Bayes, Random Forest, and Artificial Neural Networks (ANN), Reinforcement learning	SVM has high accuracy between 70% to 96%
Jetli Chung and **Jason Teo** [5]	*Prisma control*, artificial neural networks, random forest, neuro-fuzzy systems, and support vector machine	Neuro-fuzzy systems has highest accuracy
Adwitiya Ray, Akansha Bh ardwaj, Yogender Kumar Malik, Shipra Singh, Rajiv Gupta [6]	Neural networks, Natural language processing	Effective role of AI in analysing and treating mental disorder
Konda Vaishnavi, U Nikhitha Kamath, B Ashwath Rao and N V Subba Reddy [7]	Stacking, random forest, KNN, decision tree algorithm	Accuracy of all algorithms is above 79%
ChangSu, Zhenxing Xu, Jyotishman Pathak & Fei Wang [14]	To review previous studies on the use of DL algorithms in studies of mental health outcomes. Used various DL algorithms like DFNN, RNN, CNN, FNN	Accuracy of CNN algorithm is 93%
Pradnya Mehta [16]	Maximum entropy classsifictaion, naïve bayes, SVM, LSA algorithm	Rating is given by LSA algorithm

These papers were of prime importance to us. From rest of the papers we concluded that the use deep learning (DL) algorithms will be beneficial for the mental health prediction.

III. Design –

The architecture of our system is as follows –

1. Imported a dataset from a database.
2. The dataset was divided into training and test sets.
3. On the training set of data, the Random Forest algorithm was used.
4. The evaluation metrics were applied on the testing data.
5. Based on the training and testing data, the output was predicted.

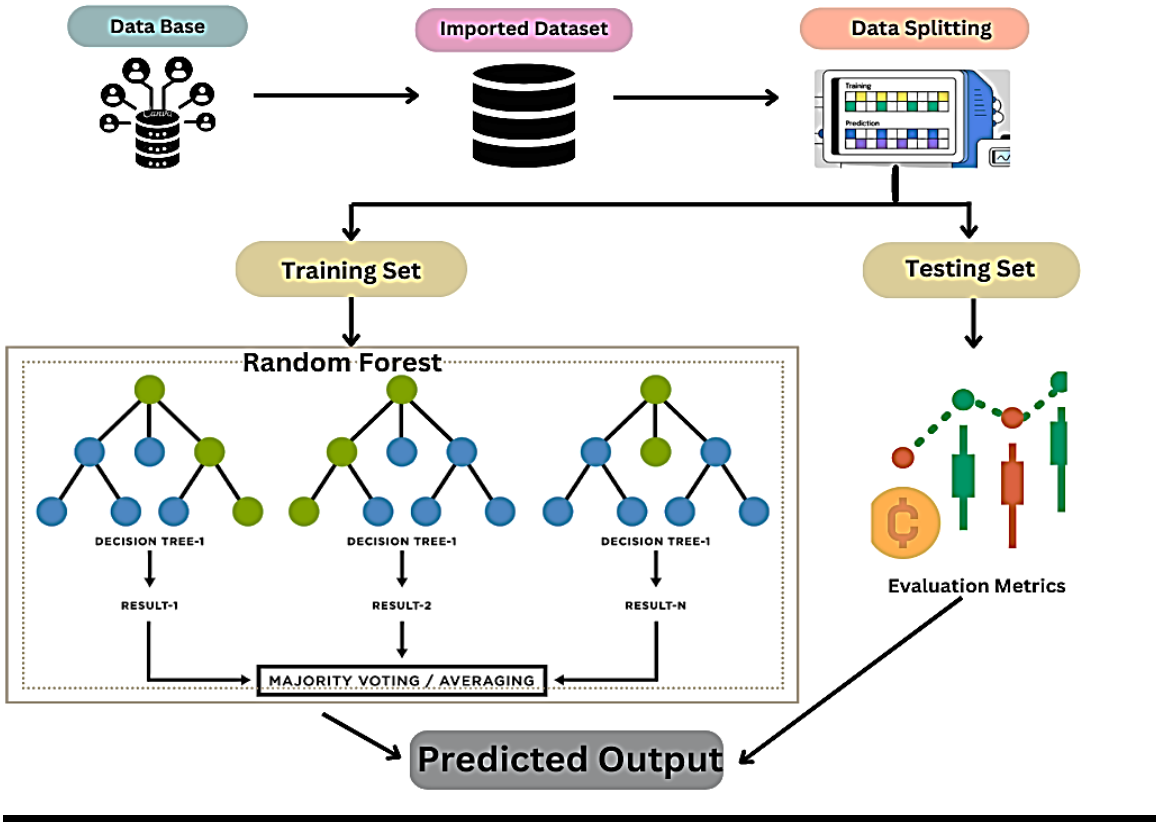


Figure. 1. System architecture

IV. Methodology

Our everyday lives depend heavily on mental health. The early detection of any type of mental disorder is crucial for our well-being. Many people don't even get to know about their diagnosis to a specific disorder in time. Also they hesitate to tell all their symptoms to the doctor. The goal of this project is to create a chatbot model that accurately predicts the type of mental condition using the random forest algorithm.

The following steps of study were followed:-

The data for this study was collected from various hospitals. On the condition of not disclosing the names, some hospitals shared their data with us. The inclusion criteria for this study were the patients diagnosed with specific type of mental disorder. Personal identifying information was anonymized and removed to ensure patient privacy. The data was converted into numeric type before processing. Various attributes of the patient were considered while processing the data like sleep disturbances, reduced appetite, delusions, hallucinations, palpitations, paranoia, fatigue, apprehension, etc.

The dataset was tested using various deep learning algorithms like Random Forest Algorithm and Support Vector Machine Algorithm. The model was divided into following 5 classes as shown in table 2.

Table 2 - Class Labels

Anxiety
Bipolar Disorder
Major Depressive Disorder
No Disease
Schizophrenia

The target variable with the type of disorders was defined. For plotting the graph of accuracy the matplotlib library in python was used. Also the module accuracy_score was imported. The data was preprocessed using the commands get_dummies and label_encoder. The X and y (i.e. Target variable) variables were defined.
Eighty percent of the dataset was utilized for training, and twenty percent was used for testing once it had been separated into these two categories. The model was trained based on training data and evaluated using performance metrics like precision, recall, F1 score and accuracy. Predictions were made based on the data. Using this trained Model, we prepared a chatbot. The users can interact freely with this chatbot in the form of their symptoms by answering specific questions asked to them by the bot. The bot in turn provides them the type of disease they might be suffering. If the user has no positive symptoms, the bot congratulates them for not having any mental disorder. This has proven very useful as the bot being robotic people feel free in answering its questions without hesitating.
The values of Precision, Recall and F1 – Score for the Random Forest Algorithm got are given in table 3.

Table 3 - Precision, Recall and F1 Score values for RF Algorithm

	Precision	**Recall**	**F1 - Score**	**Support**
Anxiety	0.87	0.95	0.91	21
Bipolar Disorder	0.91	0.91	0.91	32
Major Depressive Disorder	0.99	0.97	0.98	92
No Disease	1.00	1.00	1.00	1
Schizophrenia	0.94	0.94	0.94	54
accuracy			0.95	200
macro avg	0.94	0.95	0.95	200
weighted avg	0.95	0.95	0.95	200

Similarly, the values of Precision, Recall and F1 – Score for the SVM Algorithm got are given in table 4.

Table 4 - Precision, Recall and F1 - Score Values for SVM

	Precision	**Recall**	**F1 - Score**	**Support**
Anxiety	1.00	0.89	0.94	27
Bipolar Disorder	0.97	0.68	0.80	41
Major Depressive Disorder	0.99	0.95	0.97	111
No Disease	0.00	0.00	0.00	1

Schizophrenia	0.78	1.00	0.88	70
accuracy			0.91	250
macro avg	0.75	0.71	0.72	250
weighted avg	0.92	0.91	0.91	250

The confusion matrix is shown in fig 2 of the Random Forest Algorithm–

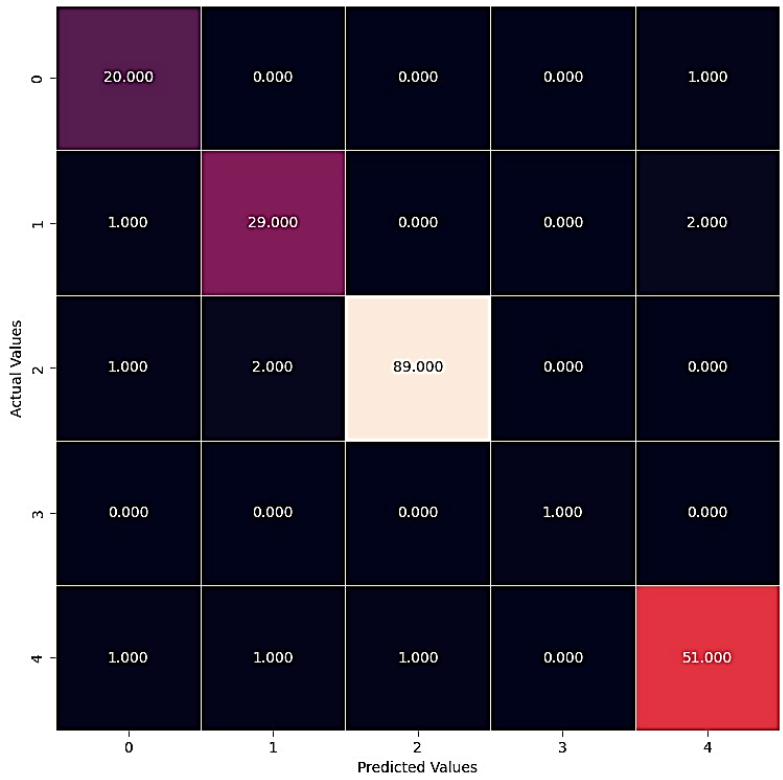


Figure. 2. Confusion Matrix

Similarly, the confusion matrix obtained for SVM algorithm is as shown in Fig 3 as follows

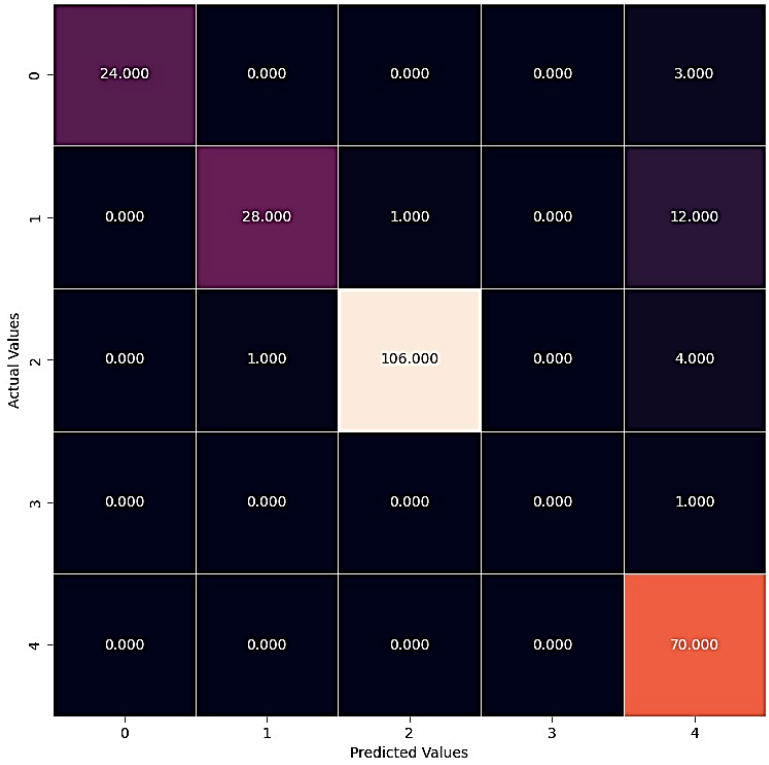


Figure. 3. Confusion Matrix

Both the algorithms (RF and SVM) were applied to the dataset and the results were compared. The bot in turn asked the questions to the user and predicted the type of mental disorder accurately. The comparative points between the two algorithms observed are mentioned in table 5 as follows –

Table 5 - Comparative study of advantages and disadvantages of both the algorithms

Algorithm	Advantages	Disadvantages
Random Forest	1. High Accuracy – It provided a high accuracy for our model i.e. 95.0%. 2. Reduced Overfitting – It reduced the overfitting and provided more reliable results for our model. 3. Handled Large Datasets – It could handle the large dataset effectively. The dataset had 1000 rows and 26 columns. Thus provided with accurate results. 4. Faster Training – Our model could be trained quickly. 5. Dealing with Outliers – It was not easily affected by strange or incorrect data points. So it gave stable predictions even if some data was unusual.	1. Lack of extrapolation – It failed to predict accurately the values outside the range of training data. It is necessary that test data should fall within the range of train data. 2. Memory Consumption – It became memory-intensive due to the formation of multiple decision trees. 3. Computationally intensive – It required more time for the formation of multiple decision trees.
Support Vector	1. Well- working with non- linear data – SVM handled the highly non- linear data (i.e. dataset with	1. Accuracy – For our model SVM proved to be less accurate than the

Machine	1000 rows and 26 columns) effectively by the formation of radial basis function (RBF) kernels. 2. Memory Efficient – When the SVM model was trained, only a subset of the training data (support vectors) was used for prediction. So it proved to be memory efficient. 3. Robust against Overfitting - It reduced the overfitting and provided more reliable results for our model. 4. Effective in high- dimensional spaces – It proved useful to implement to our dataset where more features were present than samples, on average.	RF algorithm. It provided us with an accuracy of 91.2%. 2. Proper kernel selection – The model required the proper selection of the kernel (i.e. linear, RBF or polynomial). 3. Lack of interpretability – Unlike decision trees the SVM was found to be hard to interpret, making it difficult to gain insights into the model's decision-making process. 4. Difficulty with large datasets – The training required extensive computational resources. 5. Inefficient for Multi- Clas Problems – As SVM is inherently a binary classif can be extended to handle class problems, it proved inefficient for large numb classes.

The prediction model was trained and tested using a dataset made up of factors related to mental health that were gathered through surveys, medical records, and self-reports. The dataset comprised a number of mental illnesses, including schizophrenia, bipolar disorder, major depressive disorder, and anxiety. Additionally, it contained labels reading "No Disease" for those who had come for check-ups and were in good health.

Data from 1000 patients make up the dataset we used. It includes different characteristics including age, sex, career, 24 symptoms, and the associated illnesses. The purpose of the study is to develop a bot that can reliably anticipate the illness from its symptoms. Age, sex, and career are not the main topics of this essay.

The data is mostly gathered by visiting the hospitals in the 'Ahmednagar' neighborhood. Due to the confidentiality of the information, the hospitals refused to reveal their identities or those of the patients while submitting the data. In order to protect the patients' privacy, the dataset did not include their names. Instead, they offered the information as patient numbers, such as P1, P2, etc. As a result, patients' and hospitals' privacy is protected.

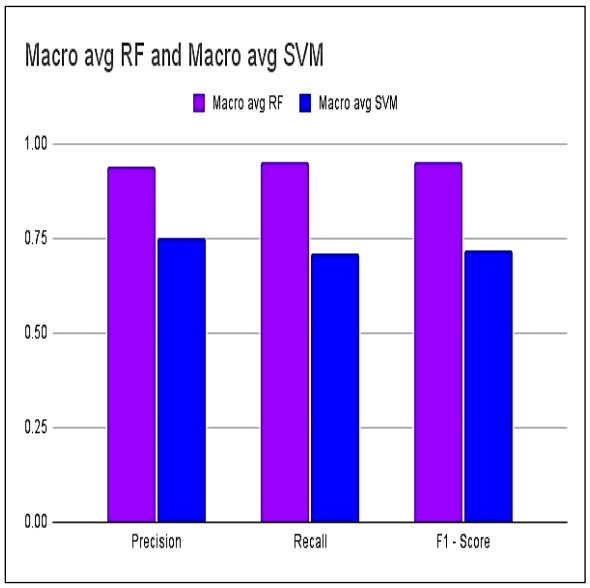
V. Results

Here, the comparison of both the algorithms (i.e. RF and SVM), in the form of strengths and weaknesses is shown in table 6.

Table 6 - comparative Study of Results

Algorithm	**Accuracy**	**F1**
Random Forest Algorithm	95.0%	0.95
Support Vector Machine Algorithm	91.2%	0.91

The comparison of the results from table 5 indicated that Random Forest Algorithm proved more effective than the SVM algorithm. The precision, recall and F1 – Score values of these models were also analyzed.



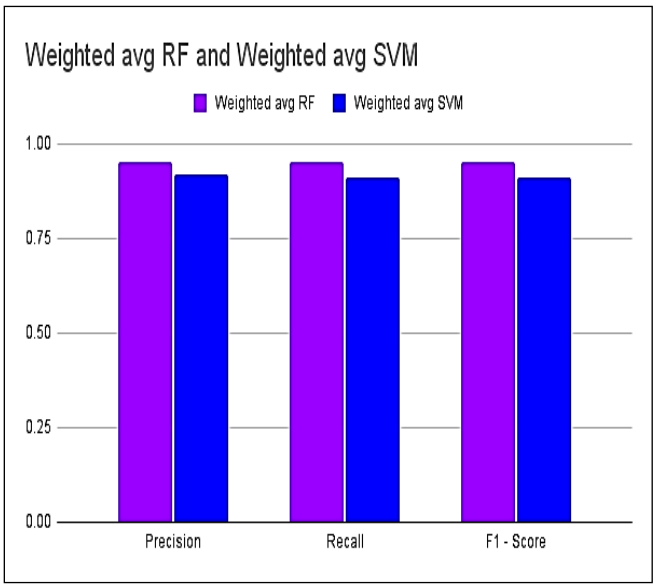


Figure 4 - Comparative Study of Macro averages

Figure 5 - Comparative Study of Weighted averages

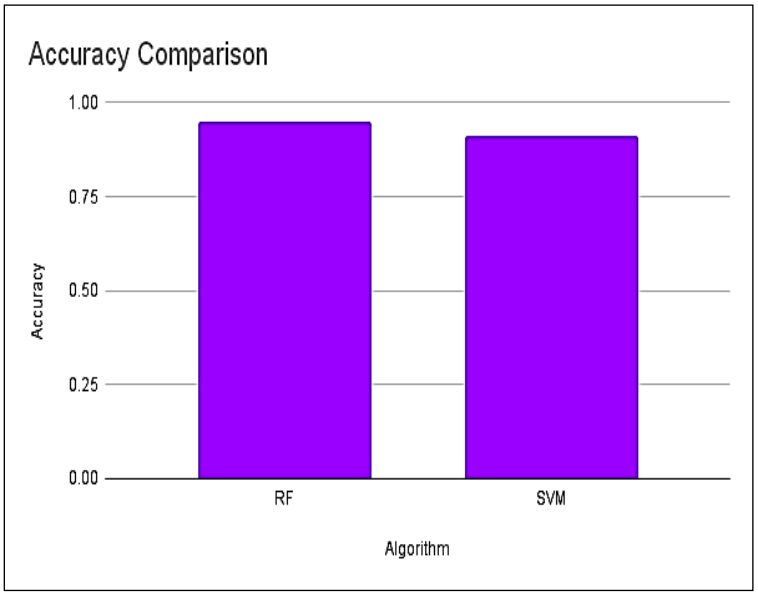


Figure 6. - Comparison of accuracies

In the figures 4, 5, 6 above the capabilities of these algorithms for the formation of the model were compared. The aforementioned statistics clearly show that the Random Forest Algorithm outperforms the Support Vector Machine Algorithm in every way. When the accuracies of both the models were compared it is observed that in our model, the Random Forest (95.0%) is superior to the Support Vector Machine (91.2%). The random forest algorithm was chosen due to its ability to handle high-dimensional data, handle nonlinear relationships, and provide interpretability.

VI. Future Directions and Limitations –

The implications of the predictive model created using the random forest algorithm for mental health disorder prediction are interpreted and discussed in this section. We present a detailed analysis of the results obtained and compare them with existing literature on AI-based mental health prediction. Additionally, we explore the generalizability of the model, address ethical considerations, and discuss its potential impact on mental health care.
We acknowledge the following limitations of our research –

1. Our project works on the basis of symptoms only. Although the dataset consisted of attributes like profession, sex, age, etc. We focused only on the symptoms and the accurate prediction of the disease since our target of research was only to build a bot which could accurately predict the type of disorder.
2. It does not predict the disorder according to the gender, range of ages, professions, etc.
3. The user should be technology friendly to make use of this model.

There is a scope to relate the disorder with age, profession, different lifestyles and gender of a person. Also there is a scope to build a model which could predict the percent intensity of which an individual is affected. Also it can tell the patient the precautions and medications which could be taken.
As our target was only to build a bot to predict accurate disorder based on the symptoms and use AI for the prediction, the paper succeeds in this motto.

Conclusion

In conclusion, the research paper explored the application of artificial intelligence, specifically the Random Forest algorithm and Support Vector Machine Algorithm, for predicting mental health issues. By utilizing a primary dataset, we were able to train and evaluate the model's performance in identifying patterns and making accurate predictions. The developed random forest model achieved an overall accuracy of 95.0% and the SVM model achieved an overall accuracy of 91.2% in predicting mental health disorders in the testing dataset. The RF model showed the precision, recall and f1 – score values as 0.94, 0.95, 0.95 respectively. Whereas the SVM model showed them as 0.75, 0.71, 0.72 respectively. The accuracy for the prediction of schizophrenia was highest and that of anxiety was lowest. The results demonstrated the potential of AI in mental health prediction, providing valuable insights for early intervention and targeted interventions. However, further research is also required to address limitations such as data quality. Due to continual advancements in AI and simple access to big datasets, the adoption of machine learning algorithms has the potential to fundamentally alter how mental health care is offered to people all over the world and improve outcomes.

References

1. Oliver Higgins RN, BN, BTech (CompSt), Brooke L. Short MBBS (Hon), MMed, BSc (Biochem), BMedSc (Path), FRANZCP, Stephan K. Chalup PhD, Dipl.-Math, Rhonda L. Wilson RN, BNSC, MNurs (Hons), PhD, **"**Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review", 06 February 2023
2. Nor Safika Mohd Shafiee, Sofianita Mutalib, Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia, "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning", 08 December 2020, 6-7
3. Ayako Baba, Kyosuke Bunji, "Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach", 10.5.2023
4. U. Ananthanagu, Pooja Agarwal, "A Systematic Review and Future Perspective of Mental Illness Detection Using Artificial Intelligence on Multimodal Digital Media", 01 January 2023
5. Jetli Chung, Jason Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges", 05 Jan 2022
6. Adwitiya Ray, Akansha Bhardwaj, YogenderKumar Malik, Shipra Singh, Rajiv Gupta, "Artificial intelligence and Psychiatry: An overview", Volume 70, April 2022
7. Konda Vaishnavi, U Nikhitha Kamath, B Ashwath Rao and N V Subba Reddy," Predicting Mental Health Illness using Machine Learning Algorithms", Volume 2161, 28-30 October 2021
8. Katarina Kjell, Per Johnsson and Sverker Sikstrom, "Freely Generated Word Responses Analyzed With Artificial Intelligence Predict Self-Reported Symptoms of Depression, Anxiety, and Worry", Volume 12, 04 June 2021
9. Sarah Graham, Colin Depp, Ellen E. Lee, Camille Nebeker, Xin Tu,, Ho-Cheol Kim & Dilip V. Jeste Current Psychiatry Reports , "Artificial Intelligence for Mental Health and Mental Illnesses: an Overview", Volume 21, Article 116, 07 November 2019
10. Simon D'Alfonso, "AI in mental health", Volume 36, December 2020, 112-117
11. Christopher A. Lovejoy, "Technology and mental health: The role of artificial intelligence", 01 January 2020
12. M. Srividya, S. Mohanavalli & N. Bhalaji, "Behavioral Modeling for Mental Health using Machine Learning Algorithms", 03 April 2018
13. Sunil Vasu Kalmady, Russell Greiner, Rimjhim Agrawal, Venkataram Shivakumar, Janardhanan C. Narayanaswamy, Matthew R. G. Brown, Andrew J Greenshaw, Serdar M Dursun & Ganesan Venkatasubramanian, "Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning", 18 January 2019
14. Chang Su, Zhenxing Xu, Jyotishman Pathak & Fei Wang, "Deep learning in mental health outcome research: a scoping review", Volume 10, Article 116, 22 April 2020
15. Ch.M.H. Saibaba; K V K Alekhya; K Yeshwanth; Praveen Tumuluru, "Prediction of Public Mental Health by using Machine Learning Algorithms", 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)
16. Pradnya Mehta, "Survey On Movie Rating And Review Summarization In Mobile Environment", International Journal of Engineering Research & Technology (IJERT) , Vol. 2 Issue 3, March – 2013,1,3-4