Journal of the Maharaja Sayajirao University of Baroda
ISSN:0025-0422

Design Analysis of and Prediction of Mental Health Disorders Using Machine Learning

S.Alekhya[1], Dr.K.Srinivasa Reddy[2], Y.Raju[3], P.Kavitha[4]

[1,3,4] Assistant Professor, Department of ECE, Sri Indu Institute of Engineering & Technology, Hyderabad
[2]Professor, Department of ECE, Sri Indu Institute of Engineering & Technology, Hyderabad

Abstract: ***Real-time analysis and prediction of mental health disorders using machine learning is an area of growing interest in the healthcare industry. With the increasing prevalence of mental health disorders, there is a need for accurate and timely diagnosis and treatment to improve patient outcomes. Machine learning algorithms can be trained on large datasets of patient information, including symptoms, medical history, and demographic data, to identify patterns and predict the likelihood of different mental health disorders. Mental health is the aggregation of emotional, social and psychological well-being of a person. It effects on the person's thinking, acting and feeling capability. Mental health is a measure of handling stress and decision making with every step-in life. There is so much data available that we are now able to compile data for mental health professionals by applying this approach they will benefit to clinicians the opportunity to personalize the professional & able to perform their job in better way in. Machine learning algorithms could help determine key behavioural biomarkers to aid mental health professionals in deciding if a patient is at risk of developing a particular mental health disorder. Additionally, the algorithms may assist in tracking effectiveness of a treatment plan. This paper reviews about the application of ML to mental health prediction, which includes a range of benefits across the areas of diagnosis, treatment and support, research, and clinical administration. With the majority of studies identified focusing on the detection and diagnosis of mental health conditions.***

Keywords: Machine learning, Appetite, Mental illness, Depression, Schizophrenia

1. Introduction

Anxiety, Depression and Schizophrenia are the most common type of mental disorders. Anxiety is a mental health disorder characterised by feelings of worry, fear that are strong enough to interfere with one's daily activities. Depression is characterised by persistently depressed mood or loss of interest in activities, causing significant impairment in daily life. The persistent feeling of sadness or loss of interest that characterises major depression can lead to a range of behavioural change. Schizophrenia affects a person's ability to think, feel and behave clearly. The exact cause of schizophrenia isn't known, but a combination of genetics, environment and altered brain chemistry and structure may play a role. This mental health disorder may include changes in sleep, appetite, energy level, concentration, daily behaviour or self-esteem. These can also be associated with thoughts of suicide. [6] Mental illness is an outcome of imbalances in brain chemistry. The evaluation of mental wellness is extremely critical to understand and suggest therapies for patients with a deviated mental behaviour.

This study employed five stages as follows, (1) data input process, (2) pre-processing data, (3) KNN algorithm classification process, Naive bayes and logistic regression (4) evaluation of algorithms. Aim of this thesis is to predict whether a person can have mental illness or not on the basis of their symptoms. We have collected data from real time feedback with the help of Google form. The dataset mainly consists of data of working individuals and student age between 20 to 35years. The data consists of string attributes which we later encoded to numeric attributes for better prediction. [2]It consists of 14 questions based on which our system will predict a specific mental illness that a person suffers from mainly (Anxiety, Depression or Schizophrenia). The data used here is in generally textual form. the data then converted into machine language for further process. Data cleaning was done manually. Firstly, it removed attributes with a lot of missing values. The missing value was data with an unfilled survey. Data cleaning consisted of five processes, namely, (1) removing attributes with a lot of missing values, (2) removing irrelevant attributes (3) removing identical attribute. Data pre-processing-Standardization is an important technique that is mostly performed as a pre-processing step before many Machine Learning models, to standardize the range of features of input data set. We have applied machine learning algorithm to create a model.

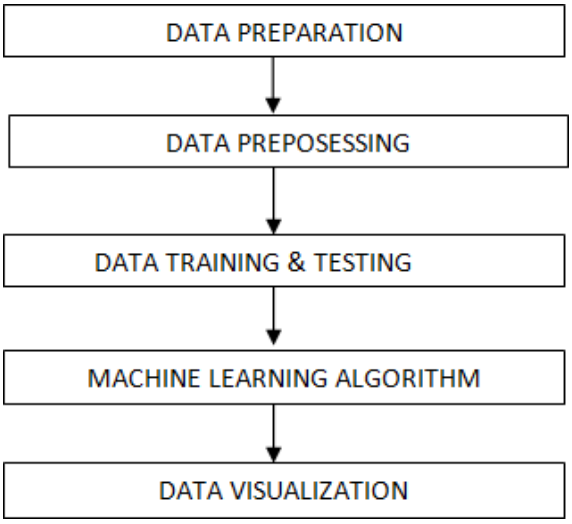

Figure 1: Shows the bock diagram of the implemented system. It illustrates the various modules and phases of the system

Author	Abstract	Methodology	Outcome
Emmanuel, Theodore, PE pintelas	A review of machine learning prediction method for anxiety disorder	i)Naïve bayes & Bayesian network ii)ANN iii)SVM	i)95% accuracy ii)82.35% iii)78.5%
Sandhya P, Mahekkantesaria	Prediction of mental disorder for employees in IT industry	i)Logistic regression ii)KNN iii)Decision tree iv)Random forest v)Bagging vi)Boosting vii)Neural network	79% 80.4% 80.6% 81.2% 77.7% 81.7% 81.4% Accuracy
Anu priya, shruti garg, nehaprernatigga	Predicting Anxiety, Depression and Stress in Modern Life using Mamchine Learning Algorithm	i) Decision tree ii)Random Forest ii) Naïve bayes iv)Support vector machine	Anxiety-0.733Depression-0.778Stress-0.628 Anxiety-0.714 Depression-0.798 Stress-0.723 Anxiety-0.733 Depression-0.855 Stress-0.742 Anxiety-0.678 Depression 0.803 Stress 0.667
Devakunchari Ramalingam, Vaibhav Sharma, Priyanka Zar	Study of Depression Analysis using Machine Learning Technique	Logistic Regression (SLR), Multilayer Perceptron Neural Networks (MLPNN), Support Vector Machine (SVM), and Random Forest (RF).	The system using this achieved average detection accuracy of 82.2% in case of males and 70.5 in case of females
V shrinivasanreddy, adityavivek, a dharun.	Machine learning techniques for stress prediction in working employees	Logistic Regression K-NN Decision Tree Random forest boosting	Boosting produced the highest accuracy

2. Related Work

In [13] U. S. Reddy et al. have applied various algorithms to find the most accurate one and compared the relationship between various parameters in the dataset. In [14] M. P. Dooshima et al. have used demographic, biological, psychological and environmental factors for prediction. Different mental health experts were consulted to validate the obtained parameters. In [15] M. Srividya et al. have used22. W. Zhang, X. Yang, S. Lui, Y. Meng, L. Yao, Y. Xiao, W. Deng, W. Zhang, and Q. Gong. 2015. Diagnostic prediction for social anxiety disorder via multivariate pattern analysis 25. Impact of COVID-19 and lockdown on mental health of children and adolescents: A narrative review with recommendations. In [7] S. G. Alonso et al. have conducted extensive review of different algorithms used for mental health prediction. Different techniques such as Association Rule Mining and Randomization were studied and their predictions werenoted for our project. This paper also reviewed other algorithms such as SVM, Decision tree, KNN, ANN, Naïve Bayes.

There are different kinds of systems that currently exist. Most of them use different methodologies to predict mental illness. Some of the current systems include an online survey which predicts whether the user has mental illness or not.

Machine Learning algorithms:

Brief Introduction of the classifiers:-Naive Bayes:-This classifier belongs to the probabilistic group of classifiers in the domain of machine learning. The base of this classifier is the Bayes Theorem where the features are considered to be independent of each other. It is a very popular when it comes to classification. It is a simple model where the test (unknown) instances are assigned class tags based on the trained model. K-NN:-K-nearest neighbor model can be used as classification model or regression model. For an unclassified instance as the input, we consider the k classified instances in a constraint region and accordingly the unclassified instance is given a class whose instances are most in that region. In case K=1, the unclassified instance is given the class whose neighbour is nearest to it, there is no need for count as the value of k is 1.

SVM:-A Support Vector Machine (SVM) works by finding a hyper-plane that can efficiently divide the set of objects in different classes. SVM takes a labelled training data, and outputs an optimal hyper-plane which can then be used to categorize new examples. A decision plane separates set of objects having memberships of different classes. For a 2d space, this hyper-plane or decision boundary is a straight line. In this image analytic module, SVM analyses data and recognizes image patterns. A set of training examples is provided to the algorithm and it generates a boundary in order to differentiate between the classes learning from training examples.

3. Implemented System

Our aim in this thesis is to predict whether a person can have mental illness or not on the basis of their symptoms. we review the non-technical and technical studies dedicated to mental illness. Mental illness is an epidemic phenomenon and is generating severe harm to people, especially adults. The thesis deals with the kinds of mental illness, several works have been done in this field and also the background studies that are important for performing the analysis. Although there is barely any work trying to predict schizophrenia taking into consideration all these features, we have tried to implement such an all-in-one model here that will take care of all these aspects of mental illness. It illustrates the methodology proposed by us. We have gathered the data from different sites and perform the process of data acquisition and feature extraction. After the pre-processing of the data, we have them labeled as 0 (no mental illness) and 1 (mental illness). The proposed method which we have used as three classification algorithm KNN, Naive bayes and Logistic regression the chapter also introduces the involve features like color histogram for data visualization.

Here we have explained the setting of various parameters that has been used for performing the experiments. We have defined the proper distribution of the data as in what proportion the modalities are used in our model. Further we have analysed our model individually for each type of modality and analysed the results. The results are also compared by using classification algorithms like Logistic regression and KNN and observed that these two gives the best results. The accuracy we have received is 96%, which is the improvement over past research.

4. Results and Analysis

Classification using Logistic regression, naive bayes and the KNN method was carried out with different values of k. it was done to obtain the optimal k value with the best accuracy. We have used three evaluation metrics: Precision, Recall, Accuracy.

Precision: It is the ratio of data elements that are correctly classified (for both the minority and majority class) to total number of classified instances‟ =TP/ (TP + FP)

Table 1: Representation of Naive bayes

Prediction	Precision	Recall	f1-score
No Mental Disease	0.91	0.94	0.93
Mental Disease	0.95	0.91	0.93

Recall: The ratio of the minority class instances that are correctly classified to the total number of actual minority class instances. R =TP/ (TP + FN).

F-Measure: Precision and Recall are used for performing the calculation of F-Measure. It is calculated by taking the harmonic mean of Precision & Recall. We can say that it is essentially an average between the two percentages. It really simplifies the comparison between the classifiers. F-measure = 2 / (1/R + 1/P).

Table 2: Representation of Logistic Regression

Prediction	Precision	Recall	f1-score
No Mental Disease	0.77	1.00	0.87
Mental Disease	1.00	0.73	0.84

Table 3: Representation of KNN

Prediction	Precision	Recall	f1-score
No Mental Disease	0.91	0.92	0.96
Mental Disease	0.93	1.00	0.96

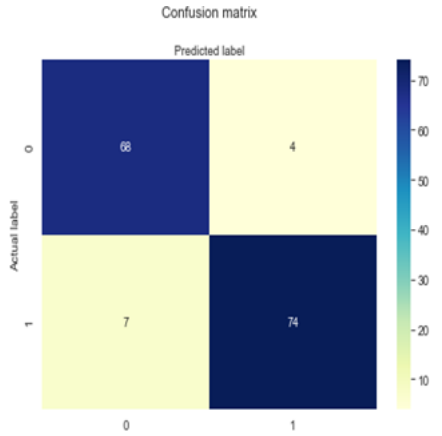

Figure 3: Confusion matrix of Logistic regression

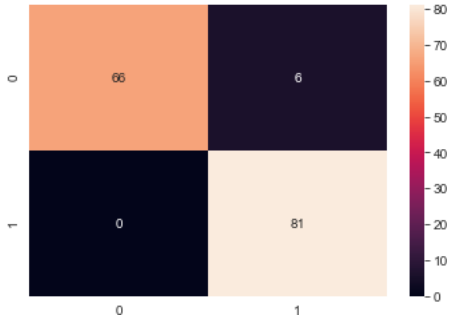

Figure 4: Confusion matrix of KNN

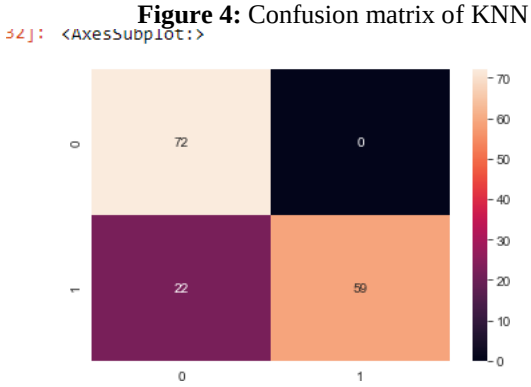

Figure 5: Confusion matrix of Naive bayes

Figures show the confusion matrix denoting the accuracy of the algorithm. It shows that 149 instances have been correctly classified as Positive while 109 have been correctly classified as negative.

$$\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN)}$$

Using the values in figures, we get the accuracy of the KNN algorithm as 0.96, Logistic regression, 0.92 and Naive bayes 0.85.

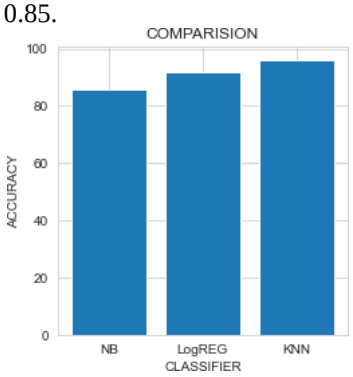

Figure 6: Comparison

5. CONCLUSION

As per the study by WHO, depression will be a major cause of mental illness in the world and people need to take more care about their mental well-being for a balanced social and professional life. People who are hesitant to approach humans for diagnosis can make use ofonline predictors for results.
To do the prediction, we have encoded the data first we have then used the KNN and logistic regression algorithm and trained a model. The accuracy we received with both the algorithm was 96% and 92%. To sum up, this research obtained 70% of training data and 30% testing data which resulted higher value of precision, recall, and accuracy compared to other data composition.

6. Future Scope

In the future, we can create a system which predict a specific mental illness formulating a scoring system for different attributes based on their importance to create a uniform scale to measure the levels of particular disease one can have, so that the person can diagnosed earlier, however extensive data collection needs to be carried out for it. Additional methods can be used to test the efficiency of the model. One can implement deep learning techniques like CNN (Convoluted Neural Networks) and verify how the model performs for the given dataset. A much more specific and vast dataset can be used as a training model since the number of responses is limited in our case. We can also customize the survey taken in order to procure responses in the right format and to increase the number of attributes as per relevance. Questionnaires from established institutions and organizations such as the World Health Organization relating to mental health can be considered for conducting a survey.

References

[1] Depression: A Global Crisis, WHO, https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf March „12

[2] U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees, " 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2018, pp. 1-4.

[3] M. P. Dooshima, E. N. Chidozie, B. J. Ademola, O.O. Sekoni, I. P. Adebayo, A Predictive Model for the Risk of Mental Illness in Nigeria

[4] Emmanuel, theodre, PE Pintelas [2018] “A review of machine learning prediction method for anxiety disorder” [DSAI 2018 conference Greece].DSAI-“software development and technologies for enhancing accessibility and fighting info-exclusion”.

[5] Vidit laijawala, Aadesh aachaliya, hardik jattta, vijaya pinjarkar [2020] “classification algorithm based mental health prediction using data mining”

[IEEE 2020].

[6] Aaditya vivek thota, A dharun [2018] “machine learning techniques for stress prediction in working employee”[IEEE 2018].

[7] Hakkum elmunysnah, risalatul, triyana, widiyaningtyas, felix [2018] “Classification of employee mental health disorder treatment with KNN algorithm”[IEEE 2018].

[8] Anu priya, shruti garg, neha prerna tigga [2019] “predicting anxiety, depression and stress in modern life using machine learning algorithm [ICCIDS 2019]” .“ ICCIDS- International conference on computational intelligence and data science ”

[9] www.towardsdatascience.com

[10] The international classification of diseases 10th edition www.WorldHealthOrganizatin.com en.wikipedia.org

[11] M. Srividya, M. Subramaniam and B. Natarajan, “Behavioral Modeling for Mental Health using Machine Learning Algorithms” “Journal of Medical Systems” Vol. 42(5):88 May 2018.

[12] DEPRESSION: A Global Crisis, WHO, https://www.who.int/mental_health/management/d epr ession/wfmh_pape r_depression_wmhd_2012

[13] U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees, " 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)

[14] Bhattacharyya, R., & Basu, S.(2018). India Inc looks to deal with rising stress in employees. Retrieved from„The Economic Times“

[15] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology.

[16] Bhakta, I and Sau, A. (2016). Prediction of Depression among Senior Citizens using Machine Learning Classifiers. International Journal of Computer Applications Vol.

[17] P. N. Tan, M. Steinbach, Vipin Kumar, “Introduction to Data Mining”, Pearson Education.

[18] V Krishnakumar, Latha Parthiban, Alzheimer‟s disease Neuroimaging Initiative, et al. 2019. A Novel Texture Extraction Technique with