**Psychological well -being prediction and solutions using machine learning​​**

*1 Yogita Hambir*

*Assitant Professor, Department of Computer Engineering*

*Army Institute of Technology, Pune*

*2Annu*

*Student, Department of Computer Engineering*

*Army Institute of Technology, Pune*

*3Priya*

*Student, Department of Computer Engineering*

*Army Institute of Technology, Pune*

*4Puja Nayak*

*Student, Department of Computer Engineering*

*Army Institute of Technology, Pune*

*5Mahima Chauhan*

*Student, Department of Computer Engineering*

*Army Institute of Technology, Pune*

**ABSTRACT:** *Mental health is a major issue today. Any issues must be identified and fixed very away to avoid any negative consequences. With the Mental Health Tracker App, we try to achieve this. We must make the app incredibly warm and welcoming because users may be dealing with mental illness.*

*Our project's objective is to create a straightforward piece of machine learning software that monitors users' advancement while recommending steps to help them better their mental health. The application asks the user a series of questions, assesses their responses, suggests tasks, keeps tabs on their mental health, and displays the results on a dashboard. This process was completed using machine learning.*

**KEYWORDS:** Random Forest, Naïve Bayes, Support Vector Machine, Recurrent Neural Network, Sentiment Analysis, Machine Learning Algorithms, Convolutional Neural Network, Logistic Regression, Linear Regression, Mental Health, Deep Learning, Binary Classification.

**INTRODUCTION**

An individual's overall perspective and general nature can both be inferred from their level of mental health. Irregularities in cerebrum science lead to dysfunctional behaviours. It's really simple to understand how to assess mental health and provide remedies for patients who behave mentally erratically. Because of many factors, most people are susceptible to pressure and are affected by discouragement.

In 2011, a management board of the World Health Organisation (WHO) conducted a survey that found that by 2030, destitution will be the primary cause of global infection problems.

A person's mental well-being can be used to assess both their mental state and general personality. Chemical abnormalities in the brain cause mental disease. In order to comprehend and cure individuals who exhibit abnormal mental behaviour, it is imperative to examine mental health. The majority of people are susceptible to stress, whereas other people experience depression for a variety of reasons.

By 2030, depression would, in line with a 2011 prediction from a World Health Organisation (WHO) administrative panel, be the leading cause of illness worldwide. Mental health disorders must be recognised and treated as soon as possible. Those who are experiencing mental health issues can find relief through early discovery, precise diagnosis, and efficient treatment. Mental illness can have major effects on those who are afflicted, their families, and society as a whole.

Commonly utilised in general methods of mental health diagnosis are interviews, observations, or questionnaires. Although labour- and time-intensive, traditional methods are frequently employed. The medical industry has already employed mobile and wearable sensors to identify mental disorder. However, most users of these technologies have a mental illness and are subject to severe supervision. A person's mental health reflects all of their physical, psychological, emotional, and social well-being. It is established what one is thinking, experiencing, and how one reacts to situations. With healthy mental health, one may perform at their best.

A person's health depends on all phases of life, including childhood, adolescence, and adulthood. Men can develop drug addiction, personality disorders, and OCD for a variety of reasons, including stress, social anxiety, depression, obsessive compulsive disorder, substance abuse, and others. Understanding the early warning symptoms of mental illness is increasingly important to maintaining a healthy work-life balance. In order to accurately forecast the onset of mental illness, it is possible to fully leverage AI and machine learning. These apps will benefit from being implemented in real time.

**RELATED WORK**

Predictions in [1] decision trees have been made using the RF, SVM, NB, and K closest neighbour algorithms. The random forest classifier outperformed all other algorithms at predicting anxiety and depression.

Deep learning techniques have been applied to [2] binary classification. The article was released in 2019. The conclusion of this article is that human-focused strategies will be more effective for engagement and will help to develop a good mode of representation for mental health prediction. [6] studies the covid 19 decision-making reactions. The common policies utilised for COVID 19 include data-driven policies, financial aid policies, and counselling policies. In COVID 19, data-driven policy proved to be the most effective response strategy. [14] asserts that ML works well as a stand-alone treatment for mental health problems. However, the introduction of the DL approach raises the possibility of accurately identifying and predicting a single sickness. In the group of patients with many comorbidities, changes in the frequency of active applications, the length of exercise, and GPS features occurred before depression. Every three weeks, online surveys were sent out in [21]. Participants in the study were signed up for 16 weeks. All techniques received approval from the Institutional Review Board at North Western University, and each participant's informed permission was obtained before to participation. Due to its speedy execution and excellent accuracy, Decision Tree was found to be the most efficient algorithm in [27] when employing data mining to predict mental health. For prediction, we used decision trees, RF, and Naive Bayes.

[28] predicts that in the future, online social media would dominate as the primary SID channel. It is essential to create new techniques that could link clinical mental health with automatic machine detection. On the substance misuse dataset, the random forest classification accuracy values were 87.72% and 92.15%, respectively, with and without an impute missing values learner. The effectiveness of classification was improved by 4.43% when the learner was added to impute missing data. [31] uses random forest as its only algorithm.

The SVM, KNN, ensemble classifiers provide the psychologically unwell class label an overall score of 0.95, according to [32]. The score, which is quite near to 1, indicated that the data sample was as mental distress actually belongs into that group.

Multilayer Perceptron, Multiclass Classifier, LAD Tree were found to provide outcomes for mental health difficulties in children that are more accurate, as reported in a 2016 [33] paper.

**DATA COLLECTION AND PREPROCESSING**

we have collected dataset from Kaggle “OSMI Tech Survey 2016”. The dataset contain 46234 rows and 64 columns. The data has record of past mental health of employees and their current and previous working conditions. This dataset is valuable for our project as we can train our machine learning models on this dataset to make accurate predictions for any mental health disorder. After collecting the dataset, it was cleaned and preprocessed to give better results. The null values present in dataset were replaced with the mode value of that column values.

mental\_df[col].mode() function was used to calculate the mode value.

mental\_df[col].fillna(mode, inplace=True) function fills the missing values in the dataset with the calculated mode values.

Total 15 features were manually selected to train model as they are the most relevant ones for our study.

**Table 1: list of features used to train ML model**

|  |  |
| --- | --- |
| S.No | Features |
| 1 | Are you self-employed |
| 2 | Do you work remotely |
| 3 | Have you had a mental health disorder in the past |
| 4 | Do you believe your productivity is ever affected by a mental health issue |
|  |  |
| 5 | Do you have a family history of mental illness |
| 6 | Do you feel comfortable in your working environment |
| 7 | Do you feel comfortable working with your direct supervisor |
| 8 | Do you feel that your organisation takes mental health as seriously as physical health |
| 9 | Have you observed or experienced an unsupportive or badly handled response to an issue in your current workplace |
| 10 | Are you stressed about your career |
| 11 | How willing would you be to share with friends and family about your work stress |
| 12 | Do you currently have a mental health disorder |
| 13 | Have you observed or experienced an unsupportive or badly handled response to an issue in your current workplace |
| 14 | Did you feel that your previous employers took mental health as seriously as physical health |
| 15 | Have your previous employers provided mental health benefits |

In the dataset, few columns contain employee name, age and department information which is not relevant to our study so those columns are not used for prediction and are simply dropped.

The mapping function is used to convert answers given by user into numerical values {-1,0,1} which will later be used to train the model.

**TRAINING THE MODEL**

We have trained four machine learning models: Random Forest, Support Vector Machine, Naïve Bayes, Recurrent Neural Network.

X: This variable often denotes our dataset's independent variables or input features.

y: This variable often represents our dataset's target variable, or dependent variable, which we are attempting to predict or categorise.

The dataset is split into training and testing set. Here we are using 75% dataset as training set and remaining 25% dataset as testing set.

**RANDOM FOREST**

Using supervised classification, the random forest technique can be used. This algorithm builds a forest with many trees, as the name would imply. The look of the forest is often stronger the more trees there are in the forest. The random forest classifier works in a similar fashion, with higher accuracy results being produced by more trees in the forest. If you're familiar with the decision-tree algorithm. The question of whether or not we should be developing more decision trees, as well as how to do so, may be on your mind. due to the fact that the same dataset will be used for all nodes selection calculations. Yes. Truly, you are. You are not going to develop a forest by modelling more decision trees.

Random Forest pseudocode:

Pick "k" features at random from a total of "m" features

Use the optimum split point to determine the node "d" among the "k" features.

Use the optimum split to divide the node into daughter nodes.

Till the "l" number of nodes, repeat steps 1 to 3 as required.

To produce a "n" number of trees, repeat steps 1 to 4 a "n" number of times.

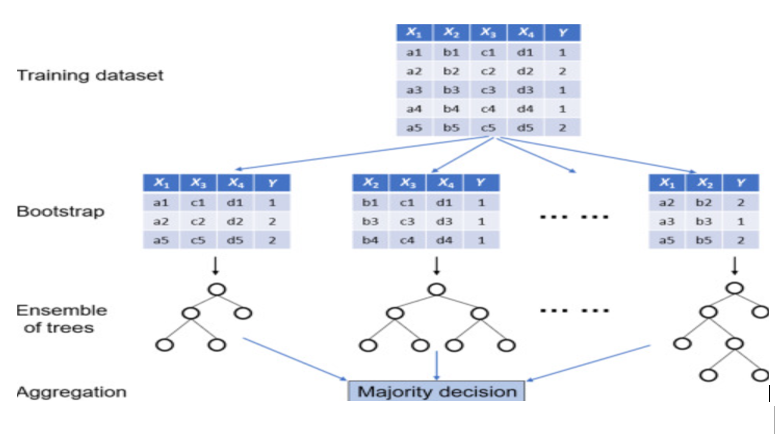


Figure 1: Random Forest model training [18]

**SUPPORT VECTOR MACHINE**

Support Vector Machine is one of the most popular methods for supervised learning, is used to address classification and regression problems. However, Machine Learning Classification problems are where it is most frequently used.

The SVM algorithm's goal is to find the optimal decision boundary or line that can categorise following data points in n-dimensional space and enable rapid assignment. This best-case decision boundary is referred to as a "hyperplane". SVM chooses the extreme points and vectors in order to construct the hyperplane. The term "Support Vector Machine" refers to a technique that uses support vectors to represent these extreme occurrences.

Data that can be separated into two groups by a single straight line are said to be linearly separable, which is a term used to describe linear SVM. Such data are classified using linear SVM classifiers.

Non-linear SVM is used for not linear separated data, which means that if dataset can’t be distinguished using a straight line, it is regarded to be non-linear data, and the classifier utilised is referred to as a Non-linear SVM classifier.

**NAÏVE BAYES**

By using sequential events, where later knowledge is learned and affects the original probability, Bayes' Theorem stands out. The terms "prior probability" and "posterior probability" are used to describe these probabilities. The initial likelihood of an event before it is contextualised in a particular scenario is known as the prior probability, also known as the marginal probability. The chance of an event following the observation of a piece of data is known as the posterior probability.

In contrast to other classifiers, nave Bayes classifiers operate under a few fundamental presumptions, hence the term "nave." It makes the supposition that a Naive Bayes model's predictors are conditionally independent of one another or unconnected to any of the other features. Additionally, it is predicated on the idea that each element influences the result equally. Although these presumptions are frequently broken in real-world situations, they make categorization problems more manageable from a computational perspective. Meaning that each variable will now only need one probability, making the computation of the model simpler. The classification algorithm performs well, especially with small sample sizes, despite this false independence assumption.

**RECURRENT NEURAL NETWORK**

For time series analysis, supervised deep learning is employed using RNN. One of the most sophisticated algorithms in the field of supervised deep learning is the recurrent neural network.

RNN and the frontal lobe:

Similar to short-term memory are RNNs. We will discover that they are able to recall recent events from the prior few observations and use that information moving ahead. One of the Frontal Lobe's roles in humans is short-term memory.

The concept of RNN

Weights are said to have Long Term Memory, often known as LTM. For instance, the weights in a classical ANN are known, thus no matter what input we give it, it will process in the same way as it did yesterday. Given that the human brain's temporal lobe controls long-term memory (LTM), the weights can be found there.

Recurrent neural networks are adjusted to create LSTM networks, which facilitate better memory retention for information. Here, the RNN's vanishing gradient issue is fixed. The classification, processing, and prediction of time series with uncertain time lags are all excellent applications for LSTM. The model is trained via back-propagation. There are three gates in an LSTM.

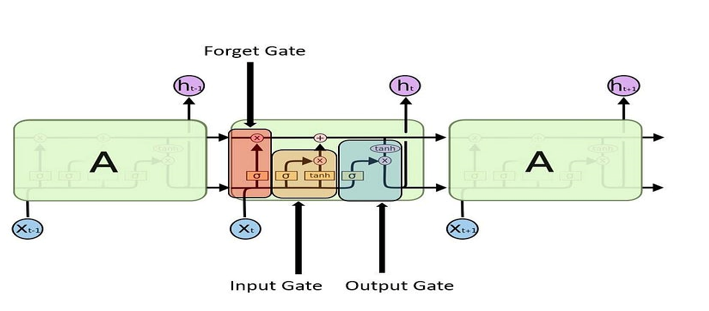


Figure 2: RNN with LSTM [13]

**TESTING THE MODEL**

Once the training of ML model is completed, the model is then tested on remaining 25% dataset. We calculate the accuracy, precision, recall, F score for the algorithms and find the optimum algorithm for our project.

After testing all four algorithms, it was found that support vector machine algorithm performed better than all other algorithms. Support Vector Machine was able to predict mental health disorders with an accuracy of 93.11%. Random forest proved to be the best after SVM in predicting psychological wellbeing of a person.

**RESULTS:**

**Table 2: Accuracy comparison of algorithms**

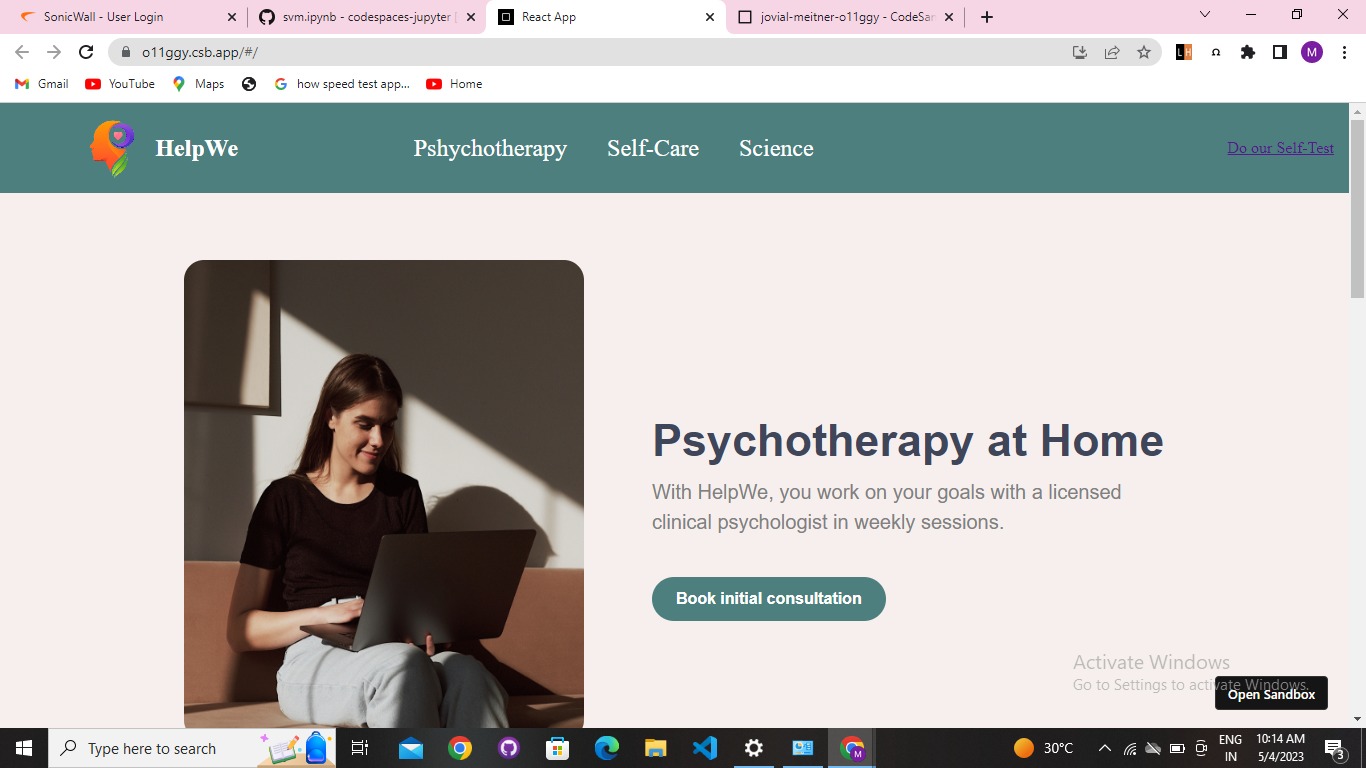
|  |  |
| --- | --- |
| Machine Learning Algorithm | Accuracy |
| Support vector machine | 93.11% |
| Recurrent neural network | 88.50% |
| Naïve bayes | 88.30% |
| Random forest | 89.13% |

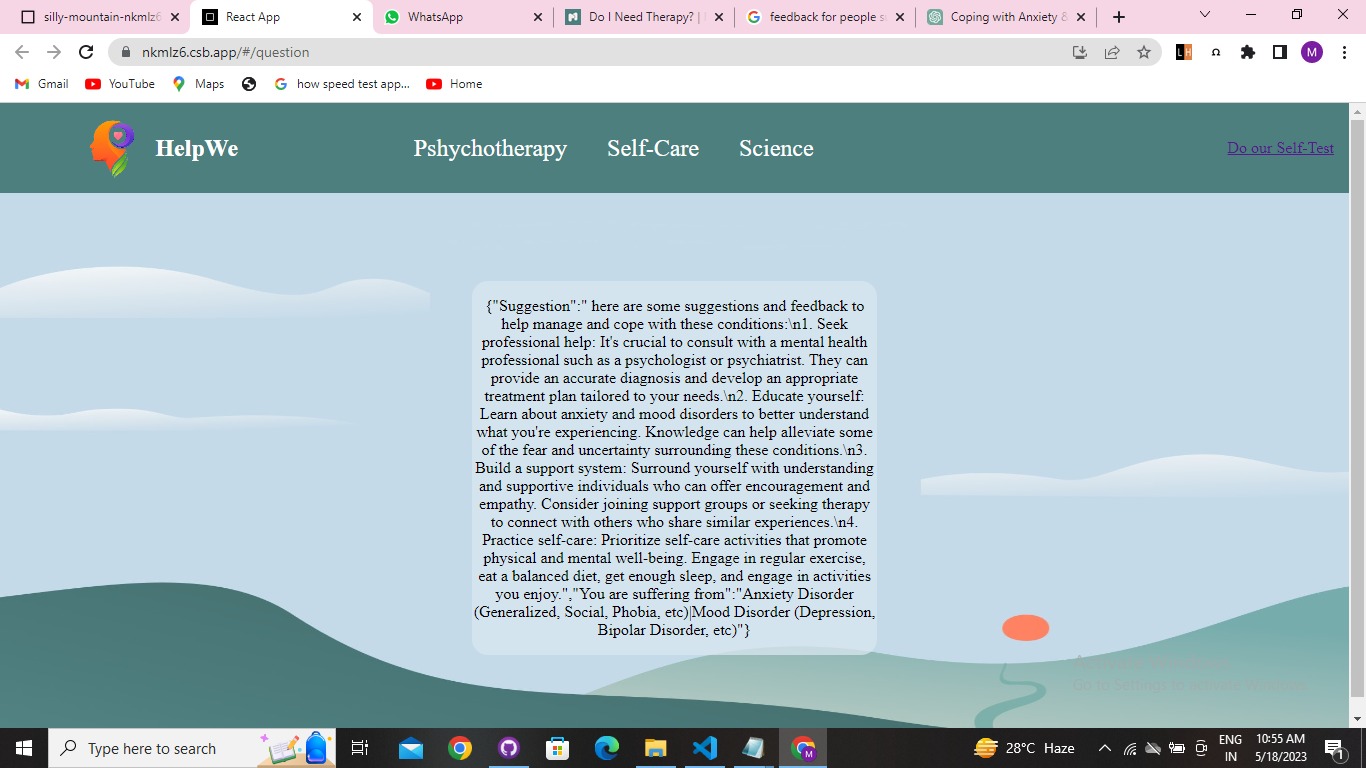
**Table 3: other comparisons**

|  |  |  |  |
| --- | --- | --- | --- |
| Machine Learning Algorithms | Precision | Recall | F-score |
| Support vector machine | 92.5 | 93.11 | 92.434 |
| Naïve bayes | 86 | 88 | 87 |
| Random forest | 86 | 89 | 87 |

RNN Loss value – 35.80

**APPLICATION IMAGES:**





**CONCLUSION**

In today’s world, for the smooth functioning of society it has become important that an individual physically and mentally healthy. Numerous people are dealing with pathological problems worldwide. With the current hospital management system in country, it is difficult to provide health care to all the people suffering from mental illness. We are implementing an app which can detect mental disorder at an early stage using machine learning algorithms. For this purpose, we are taking input for user in the form of questionnaire. We are using language translation API and natural language processing to take input in any language and in audio format. Our application will solve the problem detection of mental disorder at early stage.

**REFERENCES**

1. Anu Priya , Shruti Garg , Neha Prerna Tigga : Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms, IEEE,2019.
2. Georgia Tech, Eric P.S Baumer : Human-Centered Machine Learning:The Case of Predicting Mental Health from Social Media,2019
3. MS. Purude Vaishali Narayanrao, Dr. P. Lalitha Surya Kumari : Analysis of Machine Learning Algorithms for Predicting Depression , IEEE ,2023.
4. Tianhua Chen, Grigoris Antoniou, Marios Adamou, Ilias Tachmazidis, Pan Su :Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning , IEEE 2021.
5. Jolly Ehiabhi and HaifengWang: Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi Channel Multi-Modal Biometric Signals, IEEE 2023
6. Shuo Yu , Member, Qing Qing, Chen Zhang , Ahsan Shehzad , Giles Oatley, and Feng Xia : Data-Driven Decision-Making in COVID-19 Response:, IEEE2021
7. Shreya Ghosh: A Deep Learning Approach Depression, Intensity Estimation via Social Media , IEEE 2021
8. Ayan Seal, Rishabh Bajpai , Jagriti Agnihotri , Anis Yazidi ,Enrique Herrera-Viedma : A Deep Convolution Neural Network Framework for Detecting Depression Using EEG , IEEE 2021
9. Jianlong Zhou , Senior Member, IEEE, Hamad Zogan, Shuiqiao Yang, Shoaib Jameel, Guandong Xu , and Fang Chen : Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia ,IEEE 2021
10. Mario Ezra Arag , Adrian Pastor Lopez-Monroy, Luis Carlos Gonzalez-Gurrola , and Manuel Montes-y-Gomez : Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression , IEEE 2023
11. Kresimir osi , Sinisa Popovi , Marko Sarlija & Ivan Kesedzi : Impact of human disasters and covid 19 pandemic on mental health :potential of digital psychiatry ,2020
12. Amna Amanat , Muhammad Rizwan , Abdul Rehman Javed , Maha Abdelhaq , Raed Alsaqour,Sharnil Pandya : Deep Learning for Depression Detection from Textual Data, 2022
13. Ashley E. TateID, Ryan C. McCabeID, Henrik Larsson1, Sebastian Lundstro , Paul Lichtenstein Predicting mental health problems in adolescence using machine learning Techniques , April 6, 2020
14. Ngumimi Karen Iyortsuun , Soo-Hyung Kim ,Min Jhon , Hyung-Jeong Yang and Sudarshan Pant : A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis , 17 January 2023
15. Siriporn Sawangarreerak , Putthiporn Thanathamathee : Random Forest with Sampling Techniques for Handling Imbalanced Prediction of University Student Depression , 5 November 2020
16. Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam : Detection of child depression using machine learning methods , December 16, 2021
17. Masanori Nagamine, Szu-Nian Yang, Yueh-Ming Tai, Chin Lin : Machine Learning Based Suicide Ideation7, JULY 2020
18. Safwan Wshah, Christian Skalka, Matthew Price : Prediction for Military Personnel ,2019
19. Predicting Posttraumatic Stress Disorder Risk - A Machine Learning Approach , 2019
20. Alina Trifan, Maryse Oliveira , José Luís Oliveira Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitation , 2019
21. Chul-Hyun Cho1,Taek Lee, Min-Gwan Kim, Hoh Peter In, Leen Kim ,Heon-Jeong LeeMood : Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study, 2019
22. Jonah Meyerhoff Tony Liu, Konrad P Kording Lyle H Unga Susan M Kaiser Chris J Karr : Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study , 2021
23. Andreas K Triantafyllidis : Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature , 2019
24. Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, X Patricia B. Munroe,Bina Joe Predicting Next Call Duration : A Future Direction to Promote Mental Health in the Age of Lockdown, March 5, 2023.
25. Adrian B. R. Shatte, Delyse M. Hutchinson and Samantha J. Teague : Machine learning in mental health: a scoping review of methods and applications ,10 January 2019
26. R. A. Rahman :Application of Machine Learning Methods in Mental Health Detection: A Systematic Review, 8, 2020
27. Vidit Laijawala ,Aadesh Aachaliya :Classification Algorithms based Mental Health Prediction using Data Mining , 2019
28. Shaoxiong Ji , Shirui Pan , Member, Xue Li, Erik Cambria , Guodong Long , and Zi Huang : Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications ,2021
29. Umme Marzia HaqueID , Enamul Kabir, Rasheda Khanam :Detection of child depression using machine learning methods , December 16, 2021
30. Predicting Generalized Anxiety Disorder Among Women Using Random Forest Approach
31. Horeya Abou-Warda1, Nahla A. Belal : A Random Forest Model for Mental Disorders Diagnostic Systems ,2017
32. M. Srividya1 & S. Mohanavalli1 & N. Bhalaj :Behavioral Modeling for Mental Health using Machine Learning Algorithms ,2018
33. Ms. Sumathi M.R : Prediction of Mental Health Problems Among Children Using Machine Learning Techniques , 2016
34. Shailesh Hinduja Mahbuba Afrin Sajib Mistry, Aneesh Krishna : Machine learning-based proactive social-sensor service for mental health monitoring using twitter data , 2022