Assessing Mental Health Crisis in Pandemic Situation with Computational Intelligence

Megha Rathi, Adwitiya Sinha*, Siddhant Tulsyan, Avishka Agarwal, Anushka Srivastava
Dept. of Comp. Sc. & Engineering and Information Technology Jaypee Institute of Information Technology, India
E-mail: megha.rathi@jiit.ac.in, mailtoadwitiya@gmail.com, siddhant05tulsyan@gmail.com, avishka2404@gmail.com, anushka.srivastava2398@gmail.com
*Corresponding author

Keywords: mental health crisis, healthcare information management, computational intelligence, machine learning, synthetic minority oversampling, covid-19, biomedical informatics.

Received: January 6, 2022

The coronavirus pandemic has created huge emotional distress and increased the risk of psychiatric problems. This happened owing to imposition of necessary stringent healthcare measures that infringed personal space, emotional freedom, and caused financial loss. Our physical well-being is directly associated with mental fitness and health. From analysis it has been found that feature like struggling in concentration and memory, visionary issues, and arthritis are customary symptoms in patients suffering from mental crises. Our proposed research work aims to find out the reasons behind mental illness and ways to improve mental disorders using supervised approach. The main focus is to develop a smart computationally intelligent model to assist healthcare practitioners in analysing and diagnosing severe mental illness. Our proposed model assists in analysing causes of mental disorder and aids in reducing total medicinal cost along with reduced mental illness rate. Additionally, a recommendation system is also developed for diagnosing depressive patients.

Povzetek: *Opisana je inteligentna metoda za pomoč pri mentalnih boleznih, povezanih s pandemijami.*

1 Introduction

The public health emergencies imposed during Covid-19 pandemic has caused distressed in communities at large. The mandating of sudden and unfamiliar public safety norms have caused emotional distress among people [1]. As the normal course of living was severely encroached by home confinement and social distancing, many cases of mental health crisis started to erupt. Moreover, people who suffered from recurrent ailments during the pandemic become even more vulnerable to psychiatric problems and other severe health havocs. As a result, the yearly medicinal cases for mental disorders started increasing globally, and hence it become essential to reveal the root causes for mental disorders, including anxiety, depression, and many more adverse psychosocial disorders [2]. Moreover, the total expenditure for treating patients also increased, which includes restorative cost for treatment. In order to understand these ballooning costs, several large-scale epidemiological studies are being conducted to provide information on the health of United States citizens. One such study, the Behavioural Risk Factor Surveillance System (BRFSS), conducts surveys to collect uniform data on health risk behaviours, chronic diseases, access to healthcare information, and to employ preventative medical services in the United States [3]. This survey provides valuable information on behavioural patterns which, if coupled with current big data and machine learning techniques, may help to provide valuable insights into persons at risk of mental health crises. By targeting and understanding these populations, preventative health measures could be put into place to ultimately help lower health care costs in the United

States. Adults with depression and anxiety are significantly more expected to smoke, to be obese, to be physically inactive, to binge drink, and to drink more heavily than those who do not display any symptoms of depression and anxiety. Additionally, a dose-dependent relation exists between severity of depression and the smoking intensity, obesity, and physical inactivity, in which individuals who are more depressed become prone to heavy engagements in such activities. In a study of the 2012 Behavioural Risk Factor Surveillance System (BRFSS) data, found that there are significant relationships between depression and childhood mental illness, limited usual activity, and abuse [4]. In proposed research work J48 classification tree is used to predict depression with 82% accuracy, using these predictive attributes. Our research aims to create a solid foundation with the use of machine learning in helping to predict mental crises using multiple health attributes.

2 Background study

Extensive research and case studies were conducted in assessing the acuteness of emotional distress and forecasting mental health crisis. The authors in [5] have thoroughly discussed on the major stressors caused due to quarantine and isolation measures, and different ways to reduce its impact. In [6], several tools and measures were suggested for measuring the psychological impact of the Covid-19 pandemic. Moreover, technicians often face scarcity and imbalance in healthcare data that pose a major challenge for training models and supervised learning. This has been taken forward by the author in [7] to deal with the development of classifiers from imbalanced datasets. A dataset is considered to be imbalanced, when the characterization classes are not roughly similar. Frequently certifiable informational indexes are predominately made out of ordinary precedents with just a little level of strange or intriguing models. It is additionally the situation that the expense of misclassifying an anomalous (fascinating) model as an ordinary precedent is regularly a lot higher than the expense of the invert blunder. The authors have demonstrated that their proposed technique can accomplish better classifier execution for over-examining the minority (strange) class and under inspecting the greater part (typical) class in the Receiver Operating Characteristic (ROC) space, than just under testing the larger part class. In another novel work, Synthetic Minority Oversampling Technique (SMOTE) Rough Set Theory (RST)is proposed, which is dependent on oversampling and under sampling for high imbalanced informational indexes [8]. SMOTE-RSB is a hybrid data pre-processing approach that manages imbalanced informational indexes through the development of new examples and samples, utilizing SMOT together with the use of an altering method dependent on the RST and the lower estimation of a subset. The proposed technique has been approved by a trial think about demonstrating great outcomes utilizing C4.5 as the learning calculation.

Multi-mark learning has been turning into an inexorably dynamic region into the machine learning group since a wide variety of true issues are normally multi-named. Destroyed is an oversampling system that has been effectively connected for adjusting single-marked informational indexes, however has not been utilized in multi-name structures up until now. In this regard, authors in [9] highlighted a few methodologies are proposed and contrasted by the author all together with produce manufactured examples for adjusting informational indexes in the preparation of multi-name calculations. Results demonstrate that a right determination of seed tests for oversampling improves the grouping execution of multi-mark calculations. In yet another novel work [10], authors inspected the general social insurance costs related with sorrow depression and also, tension among essential consideration patients. Out of 2110 back to back essential consideration patients in a wellbeing support association, 12-thing Health General Questionnaire were screened with 1,962 people. 615 patients were further selected for indicative appraisal; Composite International diagnostic review performed on 373 patients and 328 were re-examining 12 months after the fact. Electronic cost records were utilized to compute absolute human services costs for the half year time frame encompassing the gauge evaluation and a comparative period encompassing the subsequent appraisal. Cost contrasts reflected higher use of general therapeutic administrations instead of higher psychological wellness treatment costs. In research [11], authors used computerized record frameworks of a vast staff model well-being up keep association (HMO) were utilized to distinguish sequential essential consideration patients with visit findings of sorrow and a correlation test of essential consideration patients with no melancholy conclusion. Comparable cost contrasts were watched for every one of the subdivisions inspected (treatment using antidepressants, treatment without antidepressants, and patients analysed at routine physical clinical visits). Drug store records showed more noteworthy perpetual medicinal sickness in the analysed discouragement gathering, however huge cost contrasts stayed after alteration ($3971 versus $2644). Two overlap cost contrasts endured for no less than a year after commencement of treatment. As an end, creators found that finding of misery is related with a summed-up increment being used of wellbeing administrations that is just halfway clarified by co grim ailments.

The authors of the paper [12], regulated a poll to 367 patients with type-1 and type-2 diabetes from the primary care clinics of two healthcare information management organizations, to get information on socioeconomics, burdensome side effects, diabetes learning, working, and diabetes self-care. Based on computerized information, we quantified therapeutic comorbidity, social insurance costs, glycosylated haemoglobin (HbA1c) levels, and oral hypoglycaemic remedy refills. Utilizing burdensome side effect seriousness tertiles (less, mid-range, or highest), they performed relapse investigations to decide the effect of burdensome indications on constancy to diabetes self-support and oral hypoglycaemic regimens, HbA1c levels, utilitarian debilitation, and human services costs. Compared with patients in the low-seriousness gloom side effect tertile, those in the medium and high-seriousness tertiles were essentially less follower to dietary suggestions. Further investigations testing the viability and cost-adequacy of upgraded models of consideration of diabetic patients with sorrow are required. In yet another contribution in the field of mental illness authors have provided information about imbalanced learning issues that hold an unlike conveyance of information tests among various classes and represent a test to any classifier as it turns out to be difficult to get familiar with the minority class tests [13]. This paper distinguishes that the majority of the current oversampling techniques may create the wrong engineered minority tests in certain situations and make learning undertakings harder. To overcome this, Majority Weighted Minority Oversampling Technique (MWMOTE) is introduced for productively handling with variant learning issues. MWMOTE first distinguishes the difficult to-learn educational minority class tests and relegates them loads as per their Euclidean separation from the closest larger part class tests. In another contribution, the authors have shown a novel Cluster Based Synthetic Oversampling (CBSO) algorithm in the proposed study [14]. CBSO receives its fundamental thought from existing manufactured oversampling methods and consolidates unsupervised clustering in its engineered information age system. One of the core machine learning algorithms that gained achievement in health analytics is Support Vector Machine (SVM). Statistics of SVM makes it suitable to handle all type of medical datasets. In numerous settings, we additionally have the choice of utilizing pool-based dynamic learning. Dynamic Learning with help vectors is examined in the study [15], i.e., a computation for picking which examples to demand straightaway. In another work, comparative

Table 1: Summarized application of machine learning techniques in mental health analysis.

S.No.	Author, Year	Objective	Approach	Results
1.	O. Oyebode, F. Alqahtani and R. Orji, 2020 [24].	In the recent study authors have analyzed mental health apps. They have evaluated online available 104 mental health apps and perform sentiment analysis on reviews.	Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Stochastic Gradient Descent (SGD), Logistic Regression (LR), and Random Forest (RF).	F1 Score and accuracy is compared and it is found that SGD achieved the best overall F1 score of 89.42 then followed by SVM, and LR.
2.	Ela Gore, Sheetal Rathi, 2019 [25].	In this work, authors surveyed researches done for the applicability of machine learning for mental heal analysis.	This paper surveyed numerous machine and deep learning models as SVM, K-Nearest Neighbor (KNN), Random Tree, Convolution Neural Network (CNN), Recurrent Neural Network (RNN) etc.	From the survey it is concluded that SVM with their different kernels and CNN models utilized in many of the research work. They also give better results in terms of parameters like accuracy, etc.
3.	Sabourin, A. A., Prater, J. C., & Mason, N. A., 2019 [26].	In today's competitive era students are at high mental health risk. Authors compared the mental health status of pharmacy students to other university students.	Computational techniques like SVM, Naive Bayes (NB), KNN, and Random Forest (RF) used.	RF achieves precision approximately equal to 83.33%, NB 71.42%, SVM 85.71% and KNN 55.55%.
4.	Hou, Y., Xu, J., Huang, Y., & Ma, X. ,2016 [27].	This one is another significant work done for analyzing mental health profile of students. It targets to find association between reading habits of students and depression induced due to reading	Compare algorithms like SVM, KNN, Decision Tree DT, Artificial Neural Network (ANN) , and Bayesian Classifier.	Most Accurate classifier is SVM with 82% accuracy.
5.	Gokten, E. S., & Uyulan, C. ,2021 [28].	Advanced machine learning techniques are applied to predict psychiatric disorders	Random Forest is used and applied on a record of 482 children.	Following results were obtained for kids with mental disorder: Accuracy= 72%, F1-Score=71%, Precision= 72%, and Recall= 71%.
6.	Xin, Y., Ren, X, 2022 [29].	Purpose of this research work is to forecast the psychiatric illness amongst old age people from the aspects like health profile, relationship with family, social behaviour, demographic location, and behaviour of health.	This paper used the random forest classifier to predict the depression of old age people.	The psychiatric disorder of rural old age grouped was 57.67%, and that of urban was 44.59%.
7.	Srividya, M., Mohanavalli, S., & Bhalaji, N., 2018 [30].	Application of numerous machine learning techniques to identify mental health is the main objective of this work.	Logistic Regression (LR), SVM, NB, DT, KNN, RF, and Bagging.	Highest Accuracy achieved by ensemble approach Bagging (90%) and RF (90%) followed by SVM (89%) and KNN (89%).
8.	Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R., 2020 [31].	A Machine Learning Model is developed and compared for predicting mental illness in adolescence. All techniques are explored based on statistical evaluation parameters.	RF, XGBoost, Neural Network (NN), logistic regression (LR), neural network and SVM.	Models compared using Area under Curve (AUC) and it is noticed that SVM and RF had highest AUC's equals to 0.754.

9.	Reddy, U. S., Thota, A. V., & Dharun, A., 2018 [32].	Stress patterns are analyzed in working professionals using machine learning techniques in order to highlight the factors that strongly affect the stress level.	LR, KNN, DT, Boosting, Bagging, RF.	From the results it has been concluded that embedded approach boosting achieves highest 75.13% accuracy.

analysis of various computational intelligence mathematical statistics for various infections determination, for instance, heart disease, diabetes, dengue, and hepatitis is presented [16]. Main emphasis of this review work to highlight the importance of machine learning techniques towards decision support system and diagnostics. In yet another novel work, authors highlighted the major mental issues also explored treatment coverage country wise [17]. In another survey [18] author has cited the importance and significance of smart devices for assessing anxiety, stress, and depression. Various work has been done in the area of health informatics for finding and extracting valuable insights using machine learning techniques [19-20]. From these researches it has been concluded that machine learning plays significant role in extracting and predicting health outcomes [21-23, 41-43].

In our research initiative, supervised machine learning approach is used to build a computationally efficient model to serve the mental health crisis in the society. Our proposed model ensures biomedical applicability by aiding the doctors to provide reliable healthcare service delivery to patients with mental health issues. List of related work in the domain of analyzing mental health illness is presented in Table I.

3 Proposed methodological framework

In our research, the BRFSS dataset was considered, which further required downstream analysis. This required data scrubbing and pre-processing techniques for cleaning and preparing the data for experimentation. Various machine learning algorithms were applied on the cleaned data set and respective accuracies were predicted. Recommendation system was built on the basis of this model using shiny web application [33-34].

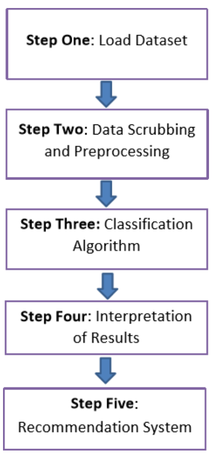

Figure 1: Structural Flow of Proposed Framework.

3.1. Data collection

The Behavioural Risk Factor Surveillance System (BRFSS) is a random annual phone-based survey which tracks health risk behaviours, chronic diseases, access to health care, and the use of preventative healthcare service management in the United States, available freely for access [4]. The most current data year (2016) was used for this project, which contained 450 attributes and 486,303 records. All questions asked in the survey (attributes) are available in [3]. Mental illness was characterized by individuals who had current anxiety and depression, life-time depression detection, and or a lifetime anxiety diagnosis and the class attribute (Mental Crises) were compiled based on these answers.

3.2. Data processing & scrubbing

Data scrubbing is the necessary action required to remove repeated, incorrect, and improperly data from the dataset [35-37]. We renamed data frame to prevent overwriting the original file, and identify the column names.

The attributes of original data were written in their short forms which were not easy to comprehend. These attribute names were expanded to make more sense of the data. It helped to read the data easily and connect different habits of a patient with its mental status. Since there were 450 attributes, some of these attributes were removed which were not needed Attributes that had no relevant meaning or no practical significance like telephone number, address, number of family members, etc that summed up to 60 columns, were removed. Record identification column was removed from the data base as it is unnecessary for downstream analysis. Our dataset consists of 6, 17, 07, 536, and NA values. This value was quite huge and hence was interfering in the various machine learning algorithms. Survey contained answer choices in the form of *none* (88), *do-not-know* (7), *refused* (9), etc. which were replaced to NA as it did not contribute in prediction. To normalize the data set, all the NA values were then replaced by means of their respective columns. Several attributes were explored. Count of *no* and *yes* was checked in the output column (*depressive*). This was done to check the proportion of *no* to *yes*. The ratio came out to be 1:4. Due to the less count of *no*, model prediction was not very accurate. Since data was quite huge so due to computational limitations, data set was sub sampled to 10% of the original data set. We made sure ratio of *no*to *yes* does not change in the sub sampled data, suggesting the smaller data set is representative of the whole data set. Data Scrubbing also included removing incomplete attributes

(i.e. those with >25% unanswered answers) and transforming attributes for downstream processing.
Data pre-processing is applied to transform raw data into a format that is easily understandable and upgrade the classifier performance [38]. Synthetic Minority Over-sampling Technique (SMOTE) was used to combat an imbalanced class design and to maintain the yes to no ratio in the sub sampled dataset. Fig 2 shows the comparison of number of classes (*yes* and *no*) before and after SMOTE. This strategy enables us to adjust the class configuration, wiping out any predisposition that may ruin our downstream analyses. Unbalanced classification issues cause problems to many learning calculations and algorithms. These issues are portrayed by the uneven extent of cases that are accessible for each class of the issue. SMOTE is a notable calculation to tackle this issue. Moreover, the dominant part class precedents are additionally under-examined, prompting an increasingly adjusted dataset.

The parameters *perc.over* and *perc.under* control the measure of over examining of the minority class and under-sampling of the majority classes, respectively. *perc.over* will typically be a number over 100. With this kind of qualities, for each case in the dataset having a place with the minority class, new instances of that class were made. In the event that *perc.over* is an incentive underneath 100 than a solitary case will be created for a haphazardly chosen extent (given by *perc.over*/100) of the cases having a place with the minority class on the first informational collection. The parameter *perc.under* controls the extent of instances of the dominant part class that will be arbitrarily chosen for the last *adjusted* informational index. This extent is determined as for the quantity of recently created minority class cases. The parameter k controls the manner in which the new precedents or examples are made. These precedents will be produced by utilizing the data from the k-nearest neighbours of every case of the minority class. The parameter k controls what number of these neighbours are utilized. This produces an arbitrary arrangement of minority class perceptions, utilizing bootstrapping and the datum point having k-closest neighbours. This decreased the predisposition towards the larger part class, while guaranteeing the new examples in the minority class were illustrative of the previous qualities. In this capacity k was set to be 5 and *perc.over* to be 110. The figure (2) demonstrates beginning number of *no*, which were 10000; while that of *yes* were 40000. Subsequent to applying SMOTE, number of *no* expanded to 18000 and *yes*, diminished to 12000.To further clean the dataset, Pearson correlation test was used to determine the correlation between each feature and the class attribute. Attributes with less than 10% correlation were discarded from downstream analysis.

Pearson correlation test. This is an estimate of precise association between two given variables of a system. Pearson correlation coefficient (r) is an estimate of the strength of the connection between the two variables. It has a value ranging from [-1,1]. If both variables increase and decrease together it implies positive correlation while if the value of one variable decrease with the increase in other variable value or vice-versa it indicates negative correlation.

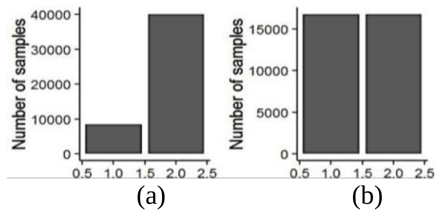

Figure 2: Comparison of number of classes (*yes*-*no*) (a) before and (b) after applying SMOTE.

3.3. Data classification

After the dataset was pre-processed and cleaned, machine learning algorithms were applied to examine its accuracy. Supervised algorithms such as k-nearest Neighbour, Random Forest, Decision Tree and SVM were applied. SVM gave an accuracy of 65% while KNN gave a precision of 70.16%. Random forest achieved an average accuracy score of 80% when *n_tree* was set to 100. Best accuracy was achieved through Decision Tree which gave 81.19% precision. A description of confusion matrix is given in Table 3. The decision tree was assembled utilizing 10-fold cross validation. The picked calculation, C4.5 or J48, was built utilizing a multistep process is presented in Table II. To start with, the single variable was discovered which best parts the information into two groups. Second, the information was separated, and the procedure was rehashed recursively until the subgroups either achieved a greatest size of 5 or no further modifications were made.

This methodology utilized a splitting criterion known as the *gain-ratio*, and was pruned utilizing a bottom up system known as *error-based* pruning. At last, precision and Area under the Curve (AUC) was surveyed to decide the reliability of the last tree and model. The Area under the Curve (AUC) of the Receiver Operating Characteristic (ROC) is a decent measure of the execution of a model. The AUC esteem can go from 0.5 (the model plays out no superior to arbitrary shot) to 1 (model suitably clarifies the reaction inside the test set).

3.4. Building recommendation model

A recommendation system was compiled to provide a user-interface program for use by doctors when their patients are in the examination room. We have developed this interface using shiny web application. This visualization helped us to give some insights on how habits like smoking, sleeping, remembering, etc can affect their mental health. All the responses of the user are recorded and scaled. We selected 6 questions according to highest gain ratio that were achieved in our decision tree model. These questions are illustrated as following.

- Have you visited a doctor for routine check-ups in last 6 months?
- Do you have memory loss issues, Concentration Issues, or Trouble in finalizing decisions?
- Do you have diabetes?
- Medical history of disease like: arthritis, lupus, fibromyalgia, or gout.
- Do you have any visionary impairment?
- Details of health policies of patient. Whether person is under health cover or not?

These questions were clustered further with other six questions whose correlation coefficient came out to be more that 10%. Response to every question was grouped with these questions to give an average depressive score. If the average depressive score is more than 50% then it represents that population in this cluster is more likely to be depressive.

Table 2: Class-view & multiple attributes in mental health dataset.

S.No.	Attribute	Values	Correlation with class attribute
1.	General Health	1. Excellent; 2. Very Good; 3.Good; 4. Fair; 5. Poor	-0.295607016
2.	Multiple Healthcare Professionals	1. Only one; 2.More than one; 3. None	0.103815018
3.	Cost prohibiting seeing a doctor	1. Yes; 2. No	0.165653515
4.	Participate in physical activities or exercise in past month	1. Yes; 2. No	-0.158701513
5.	Having disease Asthma	1. Yes; 2. No	0.100175029
6.	Having disease COPD	1. Yes; 2. No	0.186787737
7.	Having disease Arthritis	1. Yes; 2. No	0.237111272
8.	Time of last visit to dentist/ dental clinic	1. within the year; 2.within past 2 years; 3. within past 5 years; 4. five or more years ago	-0.149358224
9.	Number of permanent teeth removed	1. 1-5; 2. 6 or more; All; 4. None	0.139458736
10.	Gender of Respondent	1. Male; 2. Female	-0.248546199
11.	Marital status	1. Married; 2. Divorced; 3. Widowed; 4. Separated; 5. Never Married	-0.136709092
12.	Education level	1. Never attended; 2. Elementary; 3. Some High School; 4. High School Graduate; 5. Some College/ Technical; 6. College Graduate	0.118850168
13.	Own/Rented home	1. Own; 2. Rented; 3. Other Arrangement	-0.251743782
14.	Employment status	1. Employed; 2. Self-Employed; 3. Out of work for more than one year; 4. Out of work for less than one year; 5. Home maker; 6. Student	-0.38432588
15.	Blind/difficulty in seeing	1. Yes; 2. No	0.251269243
16.	Difficulty in remembering/concentrating	1. Yes; 2. No	0.442148749
17.	Difficulty walking/climbing stairs	1. Yes; 2. No	0.215583219
18.	Difficulty dressing/bathing	1. Yes; 2. No	0.124759584
19.	Difficulty doing errands alone	1. Yes; 2. No	0.245702066
20.	Smoked at least 100 cigarettes in entire life	1. Yes; 2. No	0.109147887
21.	Frequency of days currently smoking in month	1. Every day; 2. Some days; 3. Not all days	0.133169638

22.	Have delayed getting medical care	1. Yes; 2. No	0.16235243
23.	Been without healthcare services in past 12 months	1. Yes; 2. No	0.184359727
24.	Activity has been limited due to health problems	1. Yes; 2. No	0.191887956
25.	Having health problems that require special equipment	1. Yes; 2. No	0.1665181842
26.	Been diagnosed with depressive disorder (class attribute)	1. Yes; 2. No	1

4 Results & discussions

The correlation values between various attributes and 'depressive' show common symptoms that a patient might be dealing with in mental crises during the pandemic lockdown [39]. The results for symptoms, including difficulty in concentrating or remembering, blindness and arthritis is shown in figures 5-7. These symptoms are quite common in a person suffering from a mental crisis. Correlation coefficients of these attributes were 0.442148749, 0.251269243 and 0.215583219 respectively. Further, in figure 8, the relationship between depressive and health coverage is illustrated.

Also, we have compared our results from other existing work in the same domain listed in Table II. It has been found that the discoveries of this model help the consequences of past examinations, emphatically connecting burdensome scatters and dimensions of periodontal ailment, and proposing a negative connection with tooth brushing and dental checkups to melancholy may exist. No other existing work finding out the correlation amongst attributes as we did in the proposed work which is quite effective in highlighting the positive and negative features that directly or adverse impact the outcome. Best accuracy was achieved through Decision Tree, which gave 81.19% precision. The true positives ratio came out to be 34.9 while true negatives ratio was 46.2. This low FN rate is basic in a working model, as the cost of misclassifying a mental disease is a lot higher than the expense of misclassifying a non-mental disease. The highlights with the most elevated data gain give intriguing bits of knowledge into the respondents' practices in this investigation.

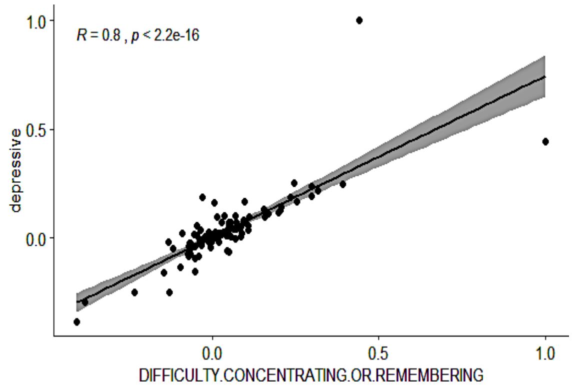

Figure 5: Relationship between depression& difficulty in concentration.

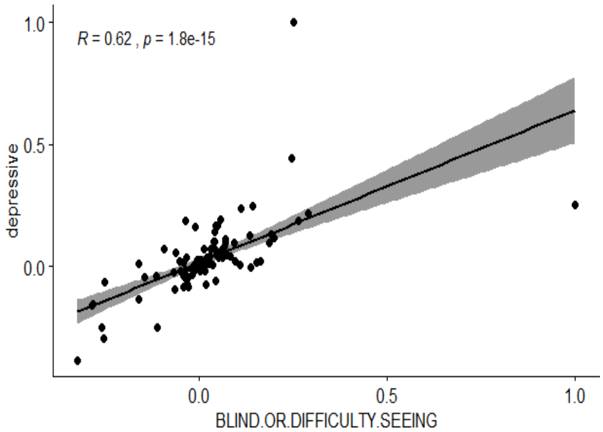

Figure 6: Relationship between depression& blindness.

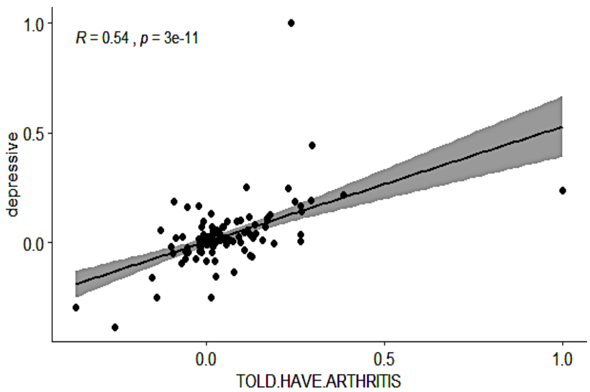

Figure 7: Relationship between depression& arthritis.

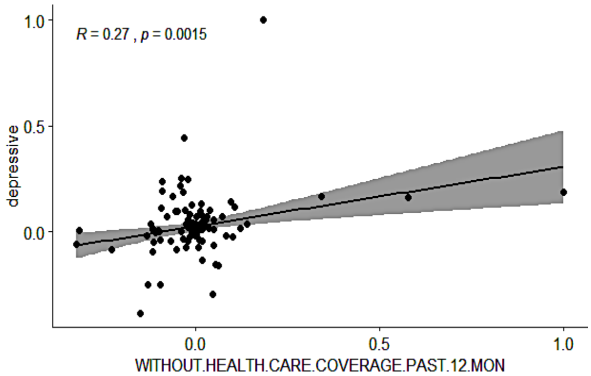

Figure 8: Relationship between depression & health coverage.

Table 3: Confusion matrix for proposed mental health model

	yes	*no*
yes	34.9	3.8
no	15.1	46.2

For our proposed model, the AUC was 0.83 as shown in figure 9. As it can be seen, the flat line initially depicts bad precision of the model. As soon as specificity value reaches a certain value of around 0.6, it escalates to a maximum value of 0.83. This shows the model has reached its maximum accuracy and hence it becomes constant thereafter. The decision tree of observed parameters is highlighted in figure 10.

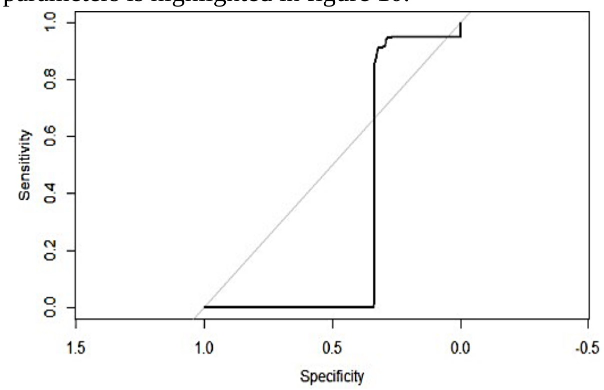

Figure 9: AUC of receiver operating characteristic.

```
DIFFICULTY.CONCENTRATING.OR.REMEMBERING = 1
| WITHOUT.HEALTH.CARE.COVERAGE.PAST.12.MON < 2
| | BLIND.OR.DIFFICULTY.SEEING < 1.95 : 1 (2472/122)
| | BLIND.OR.DIFFICULTY.SEEING >= 1.95 : 1 (2687/786)
| WITHOUT.HEALTH.CARE.COVERAGE.PAST.12.MON >= 2
| | TOLD.HAVE.ARTHRITIS < 2 : 1 (4319/452)
DIFFICULTY.CONCENTRATING.OR.REMEMBERING = 2
| LAST.VISITED.DENTIST.OR.DENTAL.CLINIC < 1
| | LENGTH.OF.TIME.SINCE.LAST.ROUTINE.CHECKU < 1 : 2 (12737/4202)
| | LENGTH.OF.TIME.SINCE.LAST.ROUTINE.CHECKU >= 1 : 2 (4219/1904)
| LAST.VISITED.DENTIST.OR.DENTAL.CLINIC >= 1
| | X.EVER.TOLD..YOU.HAVE.DIABETES = 2 : 1 (3033/742)
| | X.EVER.TOLD..YOU.HAVE.DIABETES = 1 : 2 (7907/3903)
```

Figure 10: Truncated decision tree outcome.

A confusion matrix summarizes the performance of the model [40]. The confusion matrix for Decision Tree is presented in Table III, and the model accuracy (calculated as (true observations/all observations)) was 81.07%. Table IV presented all the precision scores in descending order. From the table it is clear that Decision Tree outperform all other algorithm.

Table 4: Comparative analysis of precision score

Algorithm	Precision Score
Decision Tree	81.1935%
Random Forest	80.1265%
KNN	70.6312%
SVM	65.3542%

5 Conclusion & future research directions

Our research initiative addresses the ever-increasing crisis surfacing due to mental health related ailments, especially in Covid-19 pandemic situation. A set of supervised algorithms, including K-nearest Neighbor, Random Forest, Decision Tree and SVM were applied. Our proposed framework based off this model can help biomedical specialists in rapidly distinguishing in danger patients, prompting both higher rates of precaution medicinal services and early intercession, at last bringing down social insurance costs related with treating discouragement and tension in the country. Future undertakings should concentrate on expanding generally speaking exactness of the model to guarantee unwavering quality while giving specialists course with respect to their emotional well-being patients.

6 References

[1] Pfefferbaum, Betty, and Carol S. North. "Mental health and the Covid-19 pandemic." New England Journal of Medicine 383, no. 6 (2020): 510-512.

[2] Schäfer, Sarah K., M. Roxanne Sopp, Christian G. Schanz, Marlene Staginnus, Anja S. Göritz, and Tanja Michael. "Impact of COVID-19 on public mental health and the buffering effect of a sense of coherence." Psychotherapy and Psychosomatics 89, no. 6 (2020): 386-392.

[3] Bish, Connie L., Heidi MichelsBlanck, Mary K. Serdula, Michele Marcus, Harold W. Kohl III, and Laura Kettel Khan. "Diet and physical activity behaviors among Americans trying to lose weight: 2000 Behavioral Risk Factor Surveillance System." Obesity research 13, no. 3 (2005): 596-607.

[4] Centers for Disease Control and Prevention. "Behavioral risk factor surveillance system questionnaire." System 83, no. 12 (2011): 76.

[5] Brooks, Samantha K., Rebecca K. Webster, Louise E. Smith, Lisa Woodland, Simon Wessely, Neil Greenberg, and Gideon James Rubin. "The psychological impact of quarantine and how to reduce it: rapid review of the evidence." The Lancet 395, no. 10227 (2020): 912-920.

[6] Cortez, Pedro Afonso, Shijo John Joseph, Nileswar Das, Samrat Singh Bhandari, and Sheikh Shoib. "Tools to measure the psychological impact of the COVID-19 pandemic: What do we have in the platter?" Asian Journal of Psychiatry 53 (2020): 102371.

[7] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of Artificial Intelligence Research 16 (2002): 321-357.

[8] Ramentol, Enislay, Yailé Caballero, Rafael Bello, and Francisco Herrera. "SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and under sampling for high imbalanced data-sets using SMOTE and rough sets theory." Knowledge and information systems 33, no. 2 (2012): 245-265.

[9] Giraldo-Forero, Andrés Felipe, Jorge Alberto Jaramillo-Garzón, José Francisco Ruiz-Muñoz, and César Germán Castellanos-Domínguez. "Managing imbalanced data sets in multi-label problems: a case study with the SMOTE algorithm." In Iberoamerican Congress on Pattern Recognition, pp. 334-342. Springer, Berlin, Heidelberg, 2013.

[10] Simon, Gregory, Johan Ormel, Michael Von Korff, and William Barlow. "Health care costs associated with depressive and anxiety disorders in primary care." American Journal of Psychiatry 152, no. 3 (1995): 352-357.

[11] Simon, Gregory E., Michael Von Korff, and William Barlow. "Health care costs of primary care patients with recognized depression." Archives of general psychiatry 52, no. 10 (1995): 850-856.

[12] Ciechanowski, Paul S., Wayne J. Katon, and Joan E. Russo. "Depression and diabetes: impact of depressive symptoms on adherence, function, and costs." Archives of internal medicine 160, no. 21 (2000): 3278-3285.

[13] Barua, Sukarna, Md Monirul Islam, Xin Yao, and Kazuyuki Murase. "MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning." IEEE Transactions on knowledge and data engineering 26, no. 2 (2012): 405-425.

[14] Barua, Sukarna, Md Monirul Islam, and Kazuyuki Murase. "A novel synthetic minority oversampling technique for imbalanced data set learning." In International Conference on Neural Information Processing, pp. 735-744. Springer, Berlin, Heidelberg, 2011.

[15] Tong, Simon, and Daphne Koller. "Support vector machine active learning with applications to text classification." Journal of machine learning research 2, no. Nov (2001): 45-66.

[16] Fatima, Meherwar, and Maruf Pasha. "Survey of machine learning algorithms for disease diagnostic." Journal of Intelligent Learning Systems and Applications 9, no. 01 (2017): 1.

[17] T. Kolenik and M. Gams, "Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?" in IEEE Technology and Society Magazine, vol. 40, no. 1, pp. 80-86, March 2021, doi: 10.1109/MTS.2021.3056288.

[18] Kolenik, T. (2022). Methods in digital mental health: smartphone-based assessment and intervention for stress, anxiety, and depression. In Integrating Artificial Intelligence and IoT for Advanced Health Informatics (pp. 105-128). Springer, Cham.

[19] K. Nigam, K. Godani, D. Sharma, S. Khandelwal and M. Rathi, Personalised Heart Monitoring and Reporting System. 2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH), 2020, pp. 68-73, doi: 10.1109/INBUSH46973.2020.9392184.

[20] Rathi, M., Sahu, S., Goel, A., & Gupta, P. (2022). Personalized Health Framework for Visually Impaired. Informatica, 46(1).

[21] Gautam, A., Chauhan, A. S., Srivastava, A., Jadon, C., & Rathi, M. (2019). Major Histocompatibility Complex Binding and Various Health Parameters Analysis. In Smart Healthcare Systems (pp. 151-164). CRC Press.

[22] Rathi, M., Mittal, A., & Agarwal, D. (2020, February). Prediction of Thorax Diseases Using Deep and Transfer Learning. In 2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH) (pp. 236-240). IEEE.

[23] Rathi, M., & Pareek, V. (2016). Disease prediction tool: an integrated hybrid data mining approach for healthcare. IRACST Int J Comput Sci Inf Technol Secur (IJCSITS), 6(6), 32-40.

[24] O. Oyebode, F. Alqahtani and R. Orji, "Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews," in IEEE Access, vol. 8, pp. 111141-111158, 2020, doi: 10.1109/ACCESS.2020.3002176.

[25] E. Gore and S. Rathi, "Surveying Machine Learning Algorithms On Eeg Signals Data For Mental Health Assessment," 2019 IEEE Pune Section International Conference (PuneCon), 2019, pp. 1-6, doi: 10.1109/PuneCon46936.2019.9105749.

[26] Sabourin, A. A., Prater, J. C., & Mason, N. A. (2019). Assessment of mental health in doctor of pharmacy students. Currents in Pharmacy Teaching and Learning, 11(3), 243-250.

[27] Hou, Y., Xu, J., Huang, Y., & Ma, X. (2016, November). A big data application to predict depression in the university based on the reading habits. In 2016 3rd International Conference on Systems and Informatics (ICSAI) (pp. 1085-1089). IEEE.

[28] Gokten, E. S., & Uyulan, C. (2021). Prediction of the development of depression and post-traumatic stress disorder in sexually abused children using a random forest classifier. Journal of Affective Disorders, 279, 256-265.

[29] Xin, Y., Ren, X. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. BMC Psychiatry 22, 118 (2022). https://doi.org/10.1186/s12888-022-03742-4.

[30] Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral modeling for mental health using machine learning algorithms. Journal of medical systems, 42(5), 1-12.

[31] Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. PloS one, 15(4), e0230389.

[32] Reddy, U. S., Thota, A. V., & Dharun, A. (2018). Machine learning techniques for stress prediction in working employees. In 2018 IEEE International

[33] Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-4). IEEE.

[34] Potter, G., Wong, J., Alcaraz, I., & Chi, P. (2016). Web application teaching tools for statistics using R and shiny. Technology Innovations in Statistics Education, 9(1).

[35] Conway, Jake R., Alexander Lex, and Nils Gehlenborg. "UpSetR: an R package for the visualization of intersecting sets and their properties." Bioinformatics 33, no. 18 (2017): 2938-2940.

[36] Sinha, A., & Rathi, M. (2022). Advanced Computational Techniques for Sustainable Computing. ISBN 9781003046431, Taylor & Francis, CRC Press, Chapman & Hall, pp. 1-338

[37] Adwitiya Sinha, "PSIR: A Novel Phase-wise Diffusion Model for Lockdown Analysis of COVID-19 Pandemic in India," System Assurance Engineering & Management, Springer, pp. 1-17, October 2021

[38] Ramanna, Sheela, and Lakhmi C. Jain. Emerging paradigms in machine learning. Edited by Robert J. Howlett. Heidelberg: Springer, 2013.

[39] Sinha, A., & Rathi, M. (2021). COVID-19 prediction using AI analytics for South Korea. Applied Intelligence, 51(12), 8579-8597.

[40] Sinha, A. (2021). PSIR: a novel phase-wise diffusion model for lockdown analysis of COVID-19 pandemic in India. International Journal of System Assurance Engineering and Management, Springer, 1-14.

[41] Saxena, N., Chahal, E. S., Sinha, A., & Chand, S. (2021). Coronavirus Infection Segmentation & Detection Using UNET Deep Learning Architecture. In 2021 IEEE 18th India Council International Conference (INDICON), pp. 1-6.

[42] Gjoreski, M., Mitrevski, B., Luštrek, M., & Gams, M. (2018). An inter-domain study for arousal recognition from physiological signals. Informatica, 42(1).

[43] Peng, X. (2021). Research on Emotion Recognition Based on Deep Learning for Mental Health. Informatica, 45(1).

[44] Adeniji, O. D., Adeyemi, S. O., & Ajagbe, S. A. (2022). An Improved Bagging Ensemble in Predicting Mental Disorder using Hybridized Random Forest-Artificial Neural Network Model. Informatica, 46(4).