INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES
(Peer-reviewed, Refereed, Indexed & Open Access Journal)
DOI : 03.2021-11278686 ISSN : 2582-8568 IMPACT FACTOR : 6.865 (SJIF 2023)

Predicting for Schizophrenia by using Machine Learning Algorithms Classifiers

Nirmala S. Shinge
Assistant Professor,
Sinhgad Institute NBNCOCS,
Lonavala (Maharashtara, India)
E-mail: nima.shinge@gmail.com

Dr. B. T. Jadhav
Principal,
Yashwantrao Chavan Institute of Science,
Satara (Maharashtra, India)
E-mail: btj21875@gmail.com

DOI No. 03.2021-11278686 **DOI Link ::** https://doi-ds.org/doilink/03.2023-65465319/IRJHISIC2302027

Abstract:

Schizophrenia is a serious mental condition that places a significant clinical burden on patients. In the sphere of health care, choosing the best classification algorithm to categories and forecast disease is more significant. The importance of prediction is determined by the accuracy of the dataset and the machine learning technology used to categories the dataset. Schizophrenia may now be predicted early on because to advances in Machine Learning (ML) Algorithms. In this research paper, we used six classifiers such as Logistic Regression, Naïve Bayes, IBK, AdaBoost, Decision Table, Random Forest. For this classification Weka tool with 10 fold cross validation is used in six classifiers. The Confusion matrix is displayed, along with other relevant data and graphics.

Introduction:

Schizophrenia is a severe mental health condition in which people have an unbalanced interpretation of reality. Schizophrenia can include hallucinations, delusions, and profoundly disturbed thought and behavior that interferes with everyday functioning and can be debilitating[3]. Many Psychologist uses ICD 10 Standards for Clinical Diagnose. Schizophrenia diagnose can be done by its symptoms and various attributes. There are many questionnaire are available for severity of Schizophrenia patients.With a global age-standardized point prevalence of schizophrenia of 0.28 percent, there are 21 million cases globally.[2]Data analysis and decision making are essential components, particularly in the case of mental illness.[4] As shown in the international WHO study [5],792 million people about common mental illnesses People all over the world are suffering from

mental illness. In recent years, various studies have been conducted. Development of new and objective methods for psychiatric diagnosis based on disability-related mechanisms rather than the self-reported symptom-based diagnosis that is common today [6]. Due to the significant frequency of anxiety and depressed mood disorders in the general population, the need for novel classification techniques for mental diseases is particularly pressing.[7]

PROBLEM Statement:

Psychiatrist made decisions based on their clinical knowledge and physical examination, Choosing the best possible machine learning algorithm for disease prediction requires careful consideration.

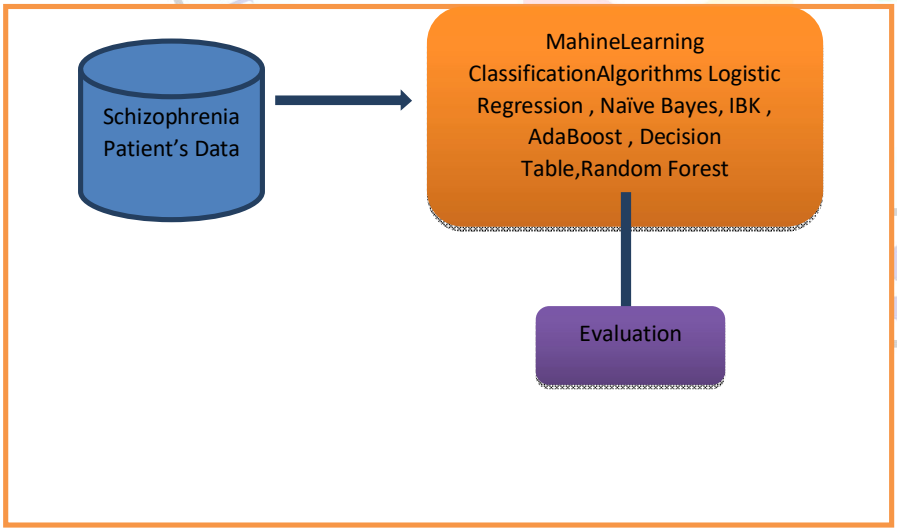
Objective of Research:

The goal of this research is to determine the best classifier that can accurately predict schizophrenia symptoms in individuals of any age while also proposing an examination of five algorithms for schizophrenia prediction using ML approaches.

Data Collection and Methodology:

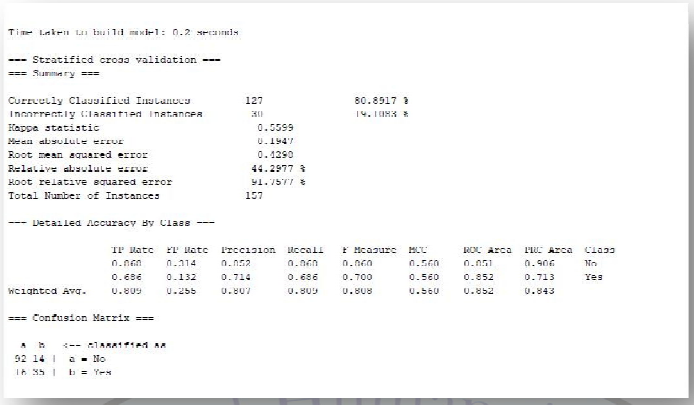
We collected raw data from Smt. Kashibai Navale Medical College and Hospital Narhe, Pune. Attributes of data were Age, Gender, Education, Professional, Marital, Any_other_Disease, Past_history, Family_history, Class, etc.

The Weka tool [8] is used for the pre-processing for this research. The performance of the six algorithms is then compared, and many key indicators are tested to see if they are a good set of guidelines for predicting disorder.



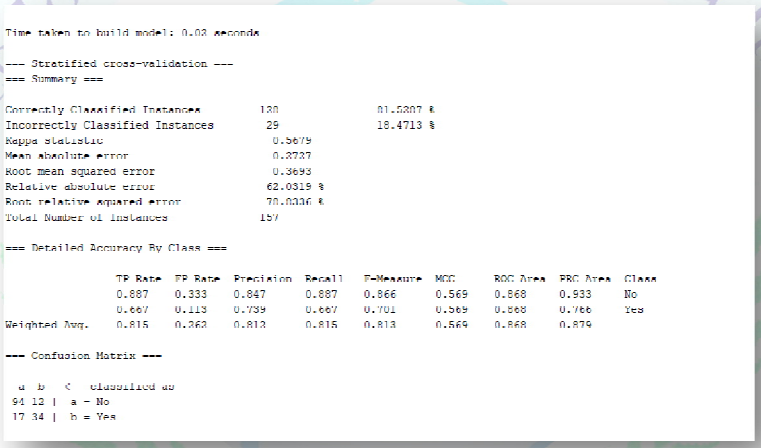
Classification Algorithms

1. Logistic Regression



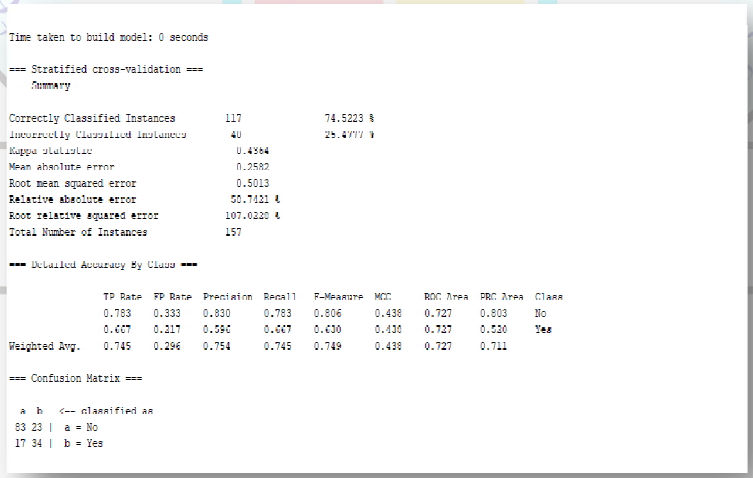
```
Time taken to build model: 0.2 seconds

=== Stratified cross validation ===
=== Summary ===

Correctly Classified Instances         127               80.8917 %
Incorrectly Classified Instances        30               19.1083 %
Kappa statistic                          0.5599
Mean absolute error                      0.1947
Root mean squared error                  0.4299
Relative absolute error                 44.2977 %
Root relative squared error             91.7577 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.860     0.314     0.852      0.860     0.860      0.560     0.851     0.906     No
                 0.686     0.132     0.714      0.686     0.700      0.560     0.852     0.713     Yes
Weighted Avg.    0.809     0.255     0.807      0.809     0.808      0.560     0.852     0.843

=== Confusion Matrix ===

  a  b   <-- classified as
 92 14 |  a = No
 16 35 |  b = Yes
```

2. NaivyBayes



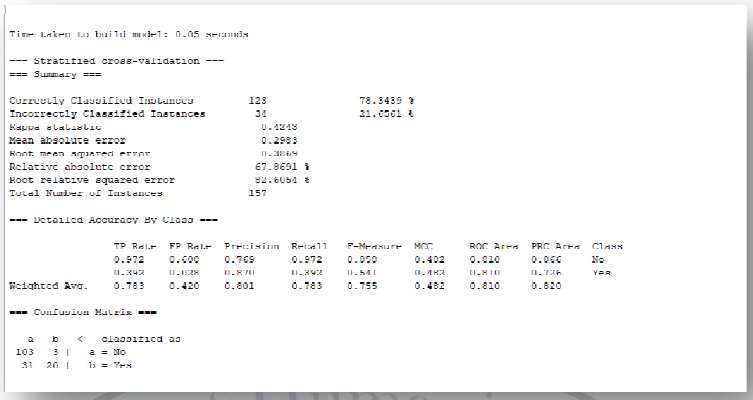
```
Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         130               81.5287 %
Incorrectly Classified Instances        29               18.4713 %
Kappa statistic                          0.5679
Mean absolute error                      0.2327
Root mean squared error                  0.3693
Relative absolute error                 52.9319 %
Root relative squared error             78.8336 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.887     0.333     0.847      0.887     0.866      0.569     0.868     0.933     No
                 0.667     0.113     0.739      0.667     0.701      0.569     0.868     0.766     Yes
Weighted Avg.    0.815     0.263     0.813      0.815     0.813      0.569     0.868     0.879

=== Confusion Matrix ===

  a  b   <-- classified as
 94 12 |  a = No
 17 34 |  b = Yes
```

3. IBK



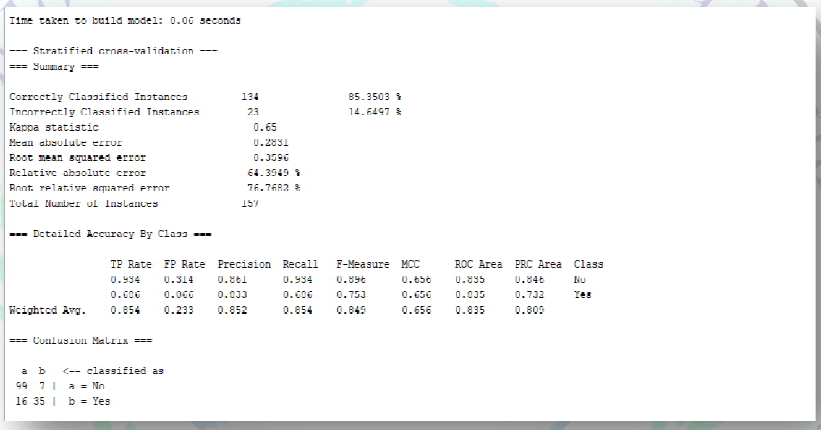
```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         117               74.5223 %
Incorrectly Classified Instances        40               25.4777 %
Kappa statistic                          0.4384
Mean absolute error                      0.2582
Root mean squared error                  0.5013
Relative absolute error                 58.7421 %
Root relative squared error            107.0228 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.783     0.333     0.830      0.783     0.806      0.438     0.727     0.803     No
                 0.667     0.217     0.590      0.667     0.630      0.438     0.727     0.530     Yes
Weighted Avg.    0.745     0.296     0.754      0.745     0.749      0.438     0.727     0.711

=== Confusion Matrix ===

  a  b   <-- classified as
 83 23 |  a = No
 17 34 |  b = Yes
```

4. AdaBoost



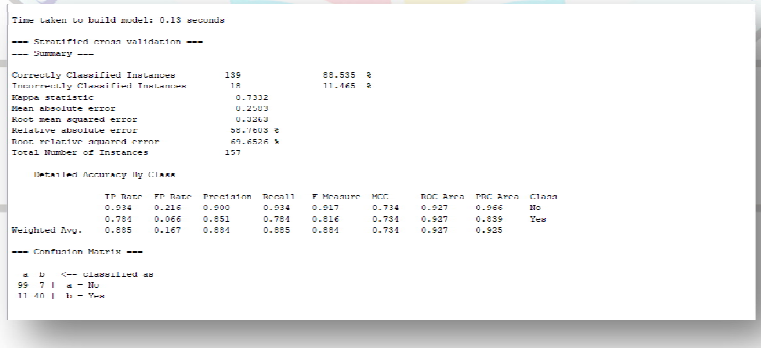
```
Time taken to build model: 0.05 seconds

--- Stratified cross-validation ---
=== Summary ===

Correctly Classified Instances         123               78.3439 %
Incorrectly Classified Instances        34               21.6561 %
Kappa statistic                          0.4248
Mean absolute error                      0.2983
Root mean squared error                  0.3869
Relative absolute error                 67.8691 %
Root relative squared error             82.6054 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.972     0.600     0.769      0.972     0.859      0.492     0.810     0.866     No
               0.400     0.028     0.870      0.242     0.541      0.492     0.810     0.774     Yes
Weighted Avg.  0.783     0.420     0.801      0.783     0.755      0.482     0.810     0.820

=== Confusion Matrix ===

   a   b   <-- classified as
 103   3 |   a = No
  31  20 |   b = Yes
```

5. Decision Table



```
Time taken to build model: 0.06 seconds

--- Stratified cross-validation ---
=== Summary ===

Correctly Classified Instances         134               85.3503 %
Incorrectly Classified Instances        23               14.6497 %
Kappa statistic                          0.65
Mean absolute error                      0.2831
Root mean squared error                  0.3596
Relative absolute error                 64.3649 %
Root relative squared error             76.7682 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.934     0.314     0.861      0.934     0.896      0.656     0.835     0.846     No
               0.686     0.066     0.833      0.686     0.753      0.656     0.835     0.732     Yes
Weighted Avg.  0.854     0.233     0.852      0.854     0.849      0.656     0.835     0.809

=== Confusion Matrix ===

  a  b   <-- classified as
 99  7 |  a = No
 16 35 |  b = Yes
```

6. Random Forest



```
Time taken to build model: 0.13 seconds

=== Stratified cross-validation ===
--- Summary ---

Correctly Classified Instances         139               88.535  %
Incorrectly Classified Instances        18               11.465  %
Kappa statistic                          0.7332
Mean absolute error                      0.2003
Root mean squared error                  0.3263
Relative absolute error                 45.7603 %
Root relative squared error             66.8526 %
Total Number of Instances              157

    Detailed Accuracy By Class

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.934     0.216     0.900      0.934     0.917      0.734     0.927     0.966     No
               0.784     0.066     0.851      0.784     0.816      0.734     0.927     0.839     Yes
Weighted Avg.  0.885     0.167     0.884      0.885     0.884      0.734     0.927     0.926

=== Confusion Matrix ===

  a  b   <-- classified as
 99  7 |  a = No
 11 40 |  b = Yes
```

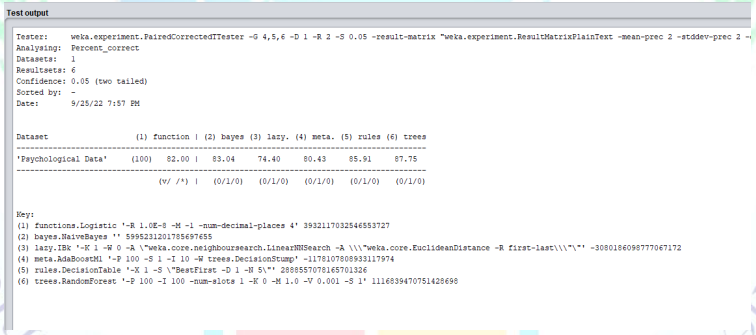
Algorithm Result and Discussion

Error Rate(in Percentage)=(Incorrectly Classified Instance/Total Number of Instance)

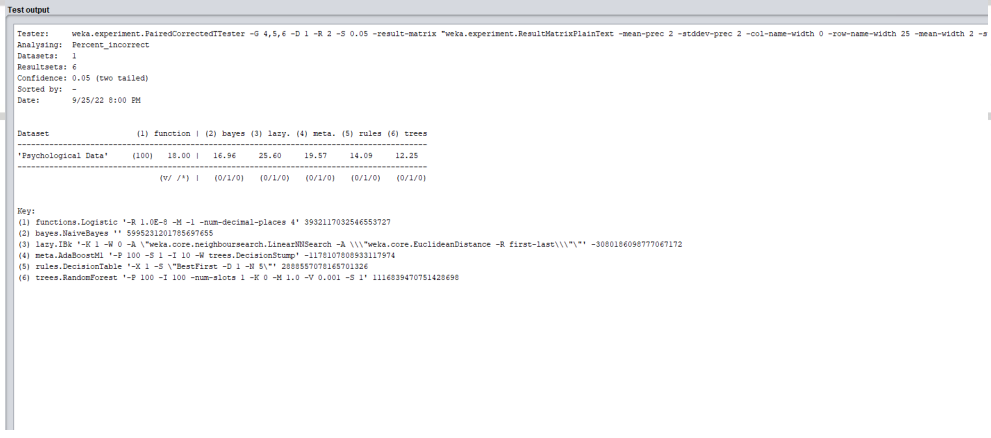
Algorithm	Correctly Instance	Incorrect Instance	Kappa Statistic	Mean Absolute	Root mean Squared Error	Relative absolute Error	Root relative squared Error	Error Rate
Logistic Regression	127	30	0.5599	0.1947	0.4298	44.29	91.75	0.1910
Naivy Bayes	128	29	0.567	0.2727	0.3693	62.03	78.83	0.1847
IBK	118	40	0.436	0.258	0.5013	58.7421	107.0228	0.2547
AdaBoost	123	34	0.42	0.29	0.3869	67.86	82.60	0.2165
Decision Table	134	23	0.65	0.28	0.35	64.39	76.78	0.1464
Random Forest	139	18	0.73	0.25	0.32	58.76	69.65	0.1146

From the evaluation of six algorithms it is clear that Random Forest performed is well as compared with other classification algorithms. By using Weka experimenter, the another evaluations such as

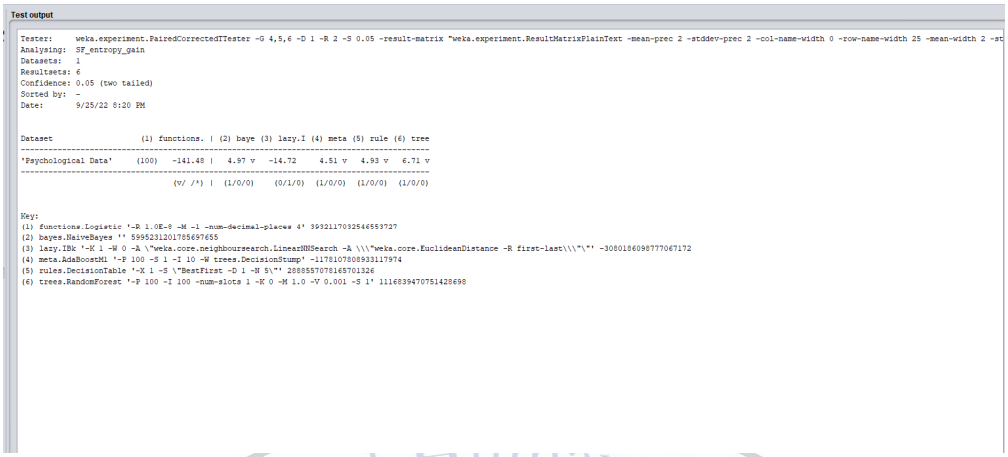
Percent correct, Percent Incorrect, entroy_gain, kappa statistics, mean absolute error.



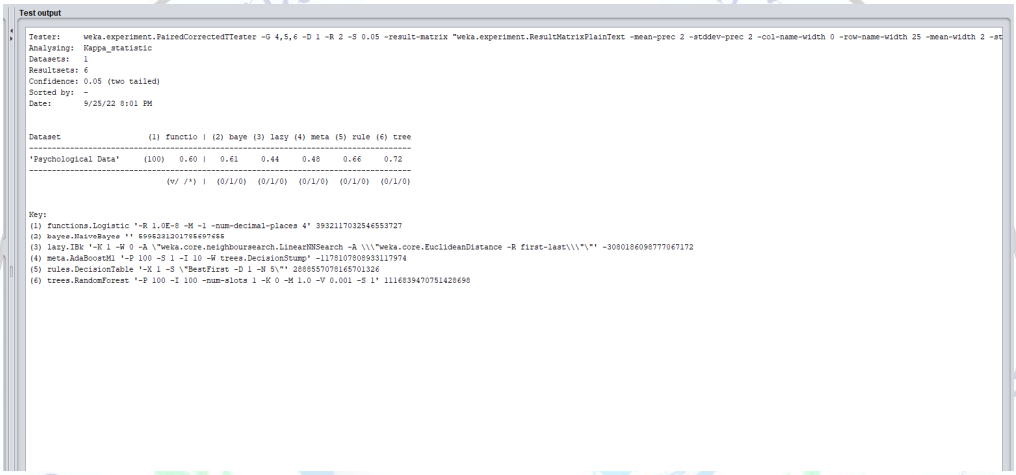
Percent correct



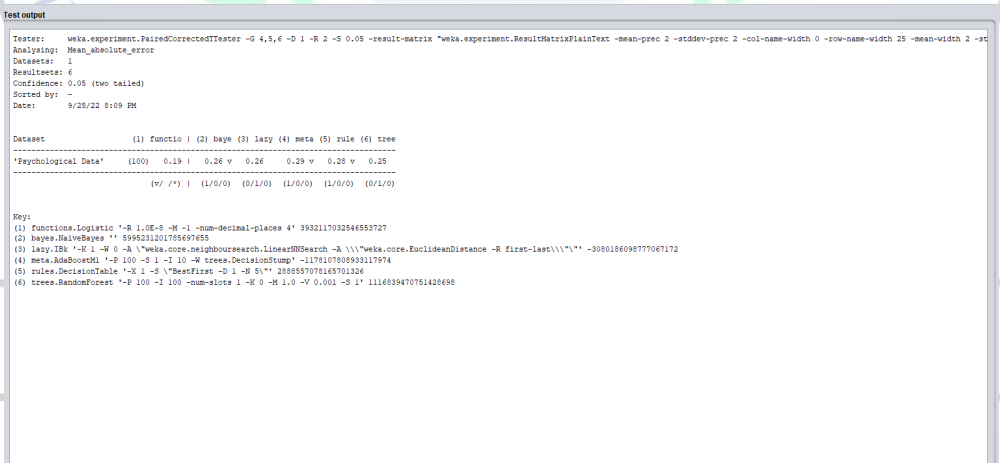
Percent Incorrect



Entroy_Gain



kappa statistics



Mean Absolute Error

Conclusion:

The aim of this paper is to evaluate the six classifiers to find the underperforming classifier to forecast schizophrenia by the data set collected from Smt. Kashibai Navale Medical college, Narhe. The dataset is handled using WEKA tool. The tool indications that Random Forest predicted the schizophrenia with 0.11% error rate. The upcoming work will be to develop this outcome and to

cultivate a software to predict the psychological disorders.

References:

1. Yassin, W. et al. Machine-learning classifcation using neuroimaging data in schizophrenia, autism, ultra-high risk and frst-episode psychosis. Transl. Psychiatry 10, 1–11 (2020).

2. Charlson, F.J.; Ferrari, A.J.; Santomauro, D.F.; Diminic, S.; Stockings, E.; Scott, J.G.; McGrath, J.; Whiteford, H.A. Global Epidemiology and Burden of Schizophrenia: Findings from the Global Burden of Disease Study 2016. Schizophr. Bull. 2018, 44, 1195–1203. [CrossRef]

3. Allende-Cid, H.; Zamora, J.; Alfaro-Faccio, P.; Alonso-Sanchez, M.F. A Machine Learning Approach for the Automatic Classification of Schizophrenic Discourse. IEEE Access 2019, 7, 45544–45553.

4. Tovar, D.; Cornejo, E.; Xanthopoulos, P.; Guarracino, M.R.; Pardalos, P.M. Data Mining in Psychiatric Research. Psychiatr. Disord. 2012, 829, 593–603

5. World Health Organization. 2017. Depression and other common mental disorders :global health estimates. Technical documents.Massachusetts Institute of Technology.24 p. pages.

6. Hofmann, S.G., Hayes, S.C., 2020. Beyond the DSM: toward a Process-Based Alternative for Diagnosis and Mental Health Treatment. Context Press/New Harbinger Publications, Oakland, CA.

7. Wu, Z., Fang, Y., 2014. Comorbidity of depressive and anxiety disorders: challenges in diagnosis and assessment. Shanghai Arch. Psychiatry 26, 227.

8. G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on 1994, pp. 357–361

9. S. Gomathi, A Deep learning of Autism Spectrum Disorder using Naïve Bayes, IBk and J48 classifiers. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019

10. Xiao, Y. et al. Support vector machine-based classifcation of frst episode drug-naïve schizophrenia patients and healthy controls using structural MRI. Schizophr. Res. 214, 11–17 (2019).