Journal of Theoretical and Applied Information Technology
31st May 2023. Vol.101. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID FEATURE SELECTION MODEL BASED ON RFE AND MRMR ON ANXIETY DISORDER DATASET

PRAJESHA T. M.[1], S. VENI[2]

[1]Research Scholar, Karpagam Academy of Higher Education, Department of Computer Science, Coimbatore 641021, India

[2]Professor and Head, Karpagam Academy of Higher Education, Department of Computer Science, Coimbatore 641021, India

E-mail: [1]tmprajesha@gmail.com, [2]venikarthik04@gmail.com

ABSTRACT

According to the reports from World Health Organization large volume of people are suffering from some kind of anxiety disorders. Among which Generalized Anxiety Disorder (GAD) is the most common one. It is very necessary to identify the anxiety disorders in beginning stage otherwise it leads to medication. Aim of this paper is to create a model to find the generalized anxiety disorder with high accuracy with lesser number of features. It describes a method to find the most important features to determine GAD. A hybrid feature selection approach is proposed which combines wrapper and filter feature selection methods to find out the best features to predict anxiety disorder. A hybrid model removes the biases that may exist while using single models. This hybrid feature selection approach is tested with four classification algorithms such as Random Tree, REP Tree, JRip, LogitBoost. As compared to the previous works this model makes the prediction more accurately with seven attributes. The dataset used for the analysis is collected from an online survey among persons above eighteen years. Analysis shows that the performance has improved after feature selection and among the classification algorithms LogitBoost gave better performance measures. It also found that the questionnaires in Hospital and Anxiety Depression Scale is well suited than other measures for finding the generalized anxiety disorder.

Keywords: *Generalized Anxiety Disorder, Recursive Feature Elimination, Feature selection, Minimum Redundancy and Maximum Relevance, Classification*

1. INTRODUCTION

Mental health is the basics of all the emotions and thinking of individuals. It is the key to relationships and contributing to the society. Mental illness may affect anyone without discriminate gender, age, sex, job, etc. Mental illness can be determined by monitoring one's thought, interaction with others and emotions. Social anxiety disorder, general anxiety disorder, panic disorder, etc are some of the common mental disorders. This study concentrates on generalized anxiety disorder, which leads to worry about day-to-day events which not depends on any specific conditions or matter.

According to World Health Organization (WHO) about 7.5 % Indians are suffering from some kind of mental disorders.[1]. There are only 0.3 psychiatrists 0.07 psychologists for every 100,000 persons and treatment gap is 70-90%. It means one third of the patients are not getting any healthcare facilities [2],[3]. This is the reason for the automated system of measuring mental disorders like anxiety. When developing such system, the most importance should be given for the features measured. An important consideration of this paper is to develop a model that make the prediction of anxiety disorder with good accuracy with minimal number of features.

Feature selection [4],[5],[6] is the approach of selecting most appropriate features for the development of a model. Feature selection can be performed using three methods: filter method wrapper method and hybrid methods. Wrapper methods provide better performance than filter method because it is based on the classifier selected but filter method has low computational cost. Utilizing the advantages of filter and wrapper methods, new hybrid methods are developed [7].

In health industry different measures are used to check the health condition of a patient. To analyze

the risk factors of several deceases, it is needed to find out the relevant features associated with that decease. Feature selection methods help to do it. One hybrid feature selection method is proposed in this paper

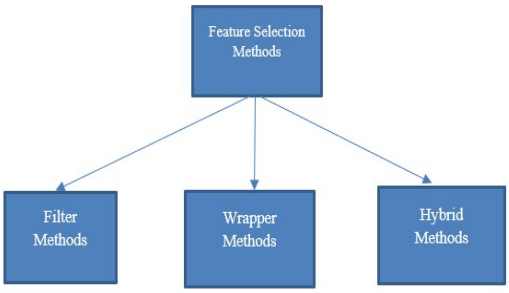

Figure 1: Feature Selection Methods

1.1 Literature Review

In paper [8] Arkaprabha and Ishita Bhakta use five machine learning algorithms ie, Catboost, Naïve Bayes, Logistic Regression, Random Forest and SVM. Catboot is a boosting algorithm here several weak learners are combined in a sequential manner to provide better results. Naive bayes perform based on Bayes theorem. Here prediction done using the concept of probability. Random forest is a bagging classification method. From the available dataset several datasets are generated and all the datasets are parallelly creates several models. By combining the results final output obtained. SVM create a linear line to separate different classes and using this hyperplane predictions are made. They performed the analysis by collecting data from seafarers. In these Catboost performed better and provided 82.6% and 84.1% accuracy and precision respectively.

Neesha Jothi, Wahidah Husain and Rashid in 2020 [9] performed feature selection by using Sharply value. It provides a more effective way of identifying individuals who are suffering some mental disorders. Sharply value is used to identify how much each feature important to make a prediction.

A compound model of Logistic regression and Decision Tree [10] was used by Silviya D'Monte and Dakshata Anchal for the proper diagnosis of anxiety disorder. While using a single model there exists several problems. In decision tree model chances of overfitting is very high. Such problems can reduce by compound models. They found out different factors that affect the mental health of Indian Patients in an economic, fast and efficient manner.

In 2015 M.Sribala [11]used artificial neural network for the prediction of anxiety disorder. Artificial neural network models work like how human brain works. It created by using input, output and hidden layers. Each layers contains large number of neurons. These neurons are activated by activation functions. For the development of better predictive models, artificial neural network using sensitivity analysis and without using sensitivity analysis was performed. From this the author found out sensitivity analysis improved the performance of the neural network.

In [12] authors adopt random forest algorithm to predict GAD among women. Here balancing of the imbalanced dataset was also performed. Imbalanced dataset means the count of one class of data points are much larger than others. So, it may not give better result. Balancing means make the count of all categories equal. It suggested that loss of interest, change in appetite, suicidal thoughts and wishes and worthlessness are the important depressive symptoms of GAD. In [13] focused on implementation of Particle Swarm Optimization algorithm (PSO) and Fuzzy Rough Set (FRS) in the classification of the anxiety disorder. Model was evaluated based on sensitivity, specificity and accuracy. Proposed hybrid model of feature selection performed better

2. MATERIALS AND METHODS

Feature selection is used to choose independent variables that highly affecting the dependent variable. When using single feature selection method there is a chance of high bias. Such biases can be overcome by using combination of multiple feature selection methods. So, a hybrid model of feature selection is proposed in this paper. Here Recursive feature elimination (RFE) and Maximum Relevance and Minimum Redundancy(mRMR) are combined to obtain better results.

2.1 Recursive Feature Elimination (RFE)

RFE is a wrapper feature selection algorithm. It selects most relevant features for the prediction of the target class. Number of features to be selected and one classification algorithm are required for finding relevant features. Features are selected based on the machine learning algorithm given. Working of RFE starting with all features in the training data set that remove the irrelevant features one by one in each iteration until reaches to the desired number.

2.2 Maximum Relevance and Minimum Redundancy (mRMR)

In this algorithm features are selected based on maximum relevance to the dependent feature and minimum redundancy to the independent features. Relevance and redundancy can be measured using MI (Mutual Information).

MI [14] can be calculated from the following equation

$$I(x,y)=\iint P(x)(y)\log\frac{P(x,y)}{P(x)P(y)} \quad (1)$$

Here x and y are any two features. Probabilities of x and y are given by P(x) and P(y). P(x,y) means the joint probabilities of x and y.

Maximum Relevance to the class label is measured by finding maximum dependency(D) of the feature to the target class. Dependency [15] is calculated as

$$D=\frac{1}{|S|}\sum_{i\in S} I(i,C) \quad (2)$$

here i is a feature, |S| is the number of features in the data set and C is the output class

When selecting a subset other aim is to reduce the redundant features. Redundancy(R)[11] can be calculated as

$$R=\frac{1}{|S|^2}\sum_{i,j\in S} I(i,j) \quad (3)$$

Here i and j are two feature sets.

Mutual Information Quotient, MIQ and Mutual Information Difference, MID are two methods to combine maximum relevancy and minimum dependency [16]. Both relevancy and dependency can be ranked by D-R in MID and D/R in MIQ.

2.3 Dataset

Data collection is done by an online survey. Dataset consists of 23 features and 1000 instances. 9 questions in Patients Health Questionnaire (PHQ) [17],7 questions in HADS [18], age group, sex, student, unemployment, education, Financial Situation, Health Status are selected as the features. PHQ and HADS are measures widely used in hospitals for checking anxiety and depressions. These questionnaires help to identify the severity of anxiety. Here the person greater than 10 score in HADS are classified as a person suffering anxiety. In the pre-processing stage numeric values are assigned to each feature as shown in Table 1.

Survey is conducted among the person above eighteen years. Dataset contain observation of male and female genders. 525 observations are collected from male and 475 observations are collected from female genders. Figure 2 represents the same.

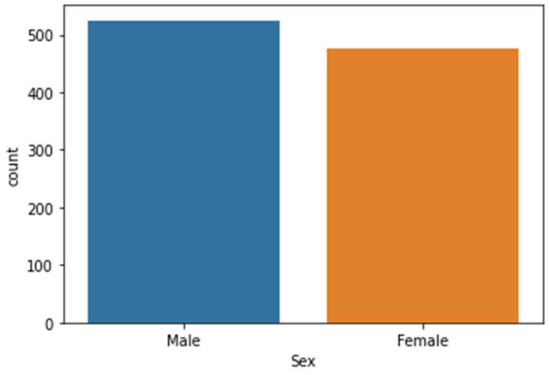

Figure 2: Count of observations based on gender

From the literature review it is observed that a greater number of females are suffering anxiety as compared to males. So, analysis is made on collected dataset. Here116 out of 475 females and 111 out of 525 males are suffering generalized anxiety disorder. More clearly 24.42 % of females and 21.14% of males are suffering anxiety in this dataset. It means chances of anxiety disorder is more in women as compared to men.

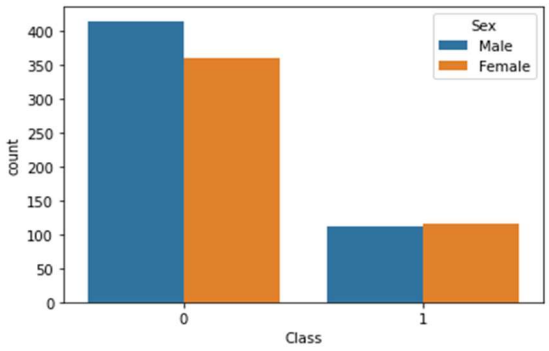

Figure 3: Analysis of anxiety disorder based on gender

Data cleaning, transformation and feature selection are main concern in data pre-processing. Here there is no chance of missing data, because while creating questionnaire all fields are kept as required fields. In the survey data are collected in the statement format, that are converted into the scoring format for the implementation purposes. If a person has total score more than 10 in HAD Scale, he is suffering from anxiety. Based on this, dataset is categorized into two classes-suffering anxiety disorder and not.

Table 1: Attributes in Dataset

No	Attributes	Assigned values
1	Sex	0-Male,1-Female
2	Age group	18-29:1 30-44:2 45-59:3 60 and bove :4
3	Education	1-6
4	Financial Situation	1-6
5	Health Status	1-3
6	Unemployed	Yes-1,No-0
7	Student	Yes-1,No-0
8	Feeling of tense or 'wound up'(GAD-1)	0-3
9	Feeling like frightened and something awful to happen (GAD-2)	0-3
10	Worrying thoughts go through my mind (GAD-3)	0-3
11	Can sit at ease and feel relaxed (GAD-4)	0-3
12	Frightened feeling (GAD-5)	0-3
13	Feel restless as I have to be on the move (GAD-6)	0-3
14	Suddenly have a feeling of panic (GAD-7)	0-3
15	Lack of interest and pleasure in doing things (PHQ-1)	0-3
16	Feeling down, hopeless or depressed (PHQ-2)	0-3
17	Trouble falling or staying asleep, or sleeping too much (PHQ-3)	0-3
18	Feeling little energy or tired (PHQ-4)	0-3
19	Over eating and poor appetite or (PHQ-5)	0-3
20	Feeling bad about yourself (PHQ-6)	0-3
21	Lack concentrating on doing things (PHQ-7)	0-3
22	Speaking or moving so slowly like other people could have noticed (PHQ-8)	0-3
23	Suicide thoughts, or of hurting yourself (PHQ-9)	0-3

2.4 Research Methodology

Collected dataset contain many fields that is in object data types. Machine learning algorithms are couldn't able to process with object data types. So first it given for preprocessing. Preprocessing done in google colab using the package pandas. Then all the observations are converted to numeric values. Then three feature selection methods such as RFE, mRMR and proposed RecursiveMR applied. Four classifiers-Random Tree, REP Tree, JRip and LogitBoost classifiers are used for the performance evaluation. Each feature selection is done on preprocessed dataset and performance evaluation is done using all the four classifiers using the tool weka.

Hybrid model of feature selection is proposed in this paper. It combines Maximum Relevance and Minimum Redundancy and Recursive Feature Elimination. RFE feature selection is performed using sklearn package in python. Here RFE feature selection is done on anxiety dataset by training the dataset with Logistic Regression classifier. Highest ranked 10 features are selected using RFE. The selected features are Unemployment, GAD7-1, GAD7-2, GAD7-4, GAD7-5, GAD7-6, GAD7-7, GAD-7-3, PHQ-3, PHQ-7. Then from the new dataset mRMR feature selection is performed. Attributes greater than a certain threshold value is selected. The attributes selected through this method are GAD7-1, GAD7-2, GAD7-3, GAD7-4, GAD7-

5, GAD7-6, GAD7-7. Here feature selection is performed using python and performance evaluation is done using weka. When combining the features selected from both methods it resolves biases of a particular model provides better results for various classifiers.

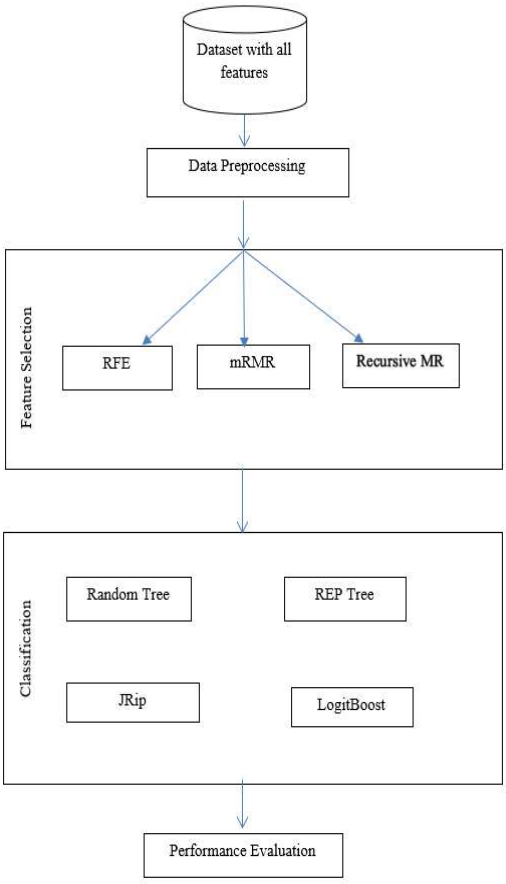

Figure 4: Methodology

2.5 Classification Algorithms.

Performance of the proposed model tested with several algorithms. Among which J48, Random Tree,REP Tree and LogitBoost provides better performance measures.So results of that four algorithms shown in this paper.

2.5.1 J48

A Decision tree classifier creates tree for performing the classification. Here nodes of the trees are the decisions and leaf of the tree represents the class labels. J48 is a popular decision tree classifier. Divide and conquer method is used in this kind of algorithms. Earlier decision tree algorithm C4.5 improved and J48 and C5.0 are developed [19]. Finding most appropriate attributes are most important concern in decision tree models. Chances of overfitting is also high in J48.

2.5.2 Random Tree

It is an ensemble classification algorithm. Ensemble Model is a combination of several weak classifiers. In random tree several decision tree models are combines together to develop a strong learner there by provides better performance measures. Here several weak decision tree models are created using different datasets created using random sampling [20].

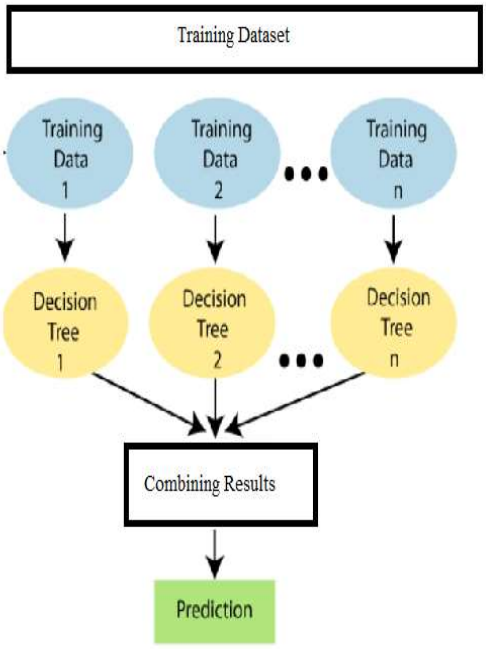

Figure 5: Random tree structure

2.5.3 REP Tree

Reduced Error Pruning Tree is a fast decision tree learner. Overfitting in decision tree models can be reduced by the technique pruning. Post pruning and repruning are the two methods used for the decision tree pruning. Here pre-pruning is applied before developing the tree. Post pruning is done after the construction of the tree. For finding best splitting attributes information gain is used in REP Tree [21].

2.5.4 Logit Boost

It is a popular classification algorithm developed using boosting technique. Boosting means a number of weak learners are sequentially combined to develop a strong learner. Logit Boost, Adaboost, XGBoost are some of the important boosting algorithms. Several datasets are generated

in this sequential process, where weightage is given to the data points the datapoints that are wrongly predicted. Final prediction is obtained by combining each individual classifier.

3. RESULTS AND DISCUSSIONS

Performance evaluation is done using four ways. Without doing any feature selection, used all the 23 features for analysis. In RFE feature selection highly ranked 7 features only considered. They are Unemployment, GAD7-1, GAD7-2, GAD7-4, GAD7-5, GAD7-6, GAD7-7. Similarly, in mRMR highly important 7 features are considered for classification which are GaAD-1, GAD7-2, GAD-3, GAD7-4, GAD7-5, GAD7-6, PHQ-3. RecursiveMR used the features GAD7-1, GAD7-2, GAD7-3, GAD7-4, GAD7-5, GAD7-6 and GAD7-7, they are obtained through the proposed model.

Accuracy is the measurement of correctly predicted instances with respect to the total number of predicted instances [22]. Accuracy evaluation using the classifiers Random Tree, REF Tree, JRip and LogitBoost are performed in this work. In this evaluation all the classifiers provide good accuracy using RecursiveMR feature selection. Among the all classifiers LogitBoost provides better classification accuracy that is 96.6%.

Table 2: Performance Evaluation Based on Accuracy

Method	Random Tree	REP Tree	JRip	Logit Boost
Without FS	91.6	90.6	94.7	96.6
RFE	95	94.3	94.4	95.3
mRMR	95.4	95.1	95.1	96.6
Recursive MR	95.9	95.1	95.4	96.6

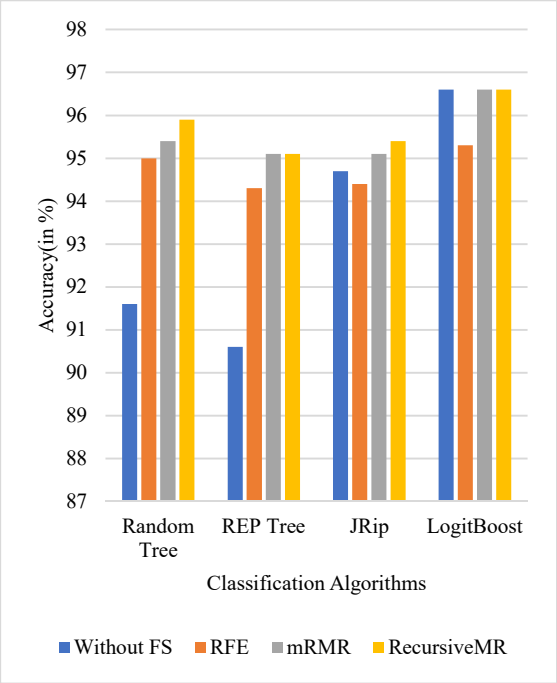

Figure 6: Comparison of performance based on Accuracy

Recall is the measure of ability of the classifier to correctly identify the positive instances [23]. Proposed feature selection model provides highest recall in Random Tree and JRip. In the performance evaluation of recall, REP Tree provides the same value (0.951) in mRMR and RecursiveMR feature selection. In the LogitBoost classification, same value (0.966) of recall obtained for all methods except RFE.

Table 3: Performance Evaluation Based on Recall

Method	Random Tree	REP Tree	JRip	Logit Boost
Without FS	0.916	0.906	0.947	0.966
RFE	0.95	0.943	0.944	0.953
mRMR	0.954	0.951	0.951	0.966
Recursive MR	0.959	0.951	0.954	0.966

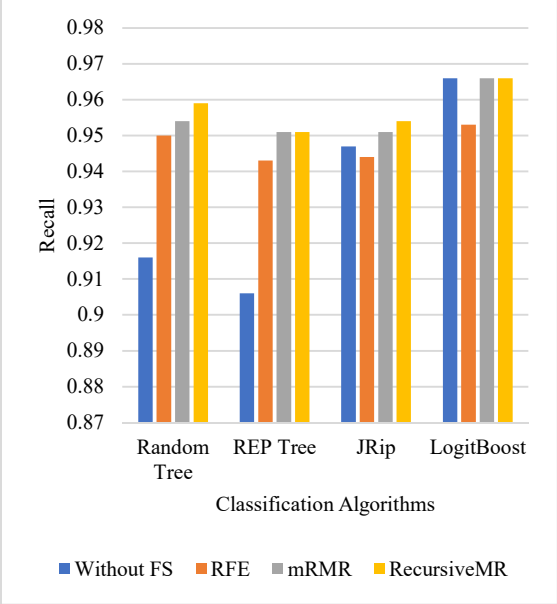

Figure 7: Comparison of performance based on Recall

A classifier's exactness is measured by precision. Proposed model shows good result in the evaluation of precision. In Random Tree and JRip, RecursiveMR give higher precision values 0.959 and 0.954 respectively. In REPTree both mRMR and RecursiveMR give the same precision that is 0.951. LogitBoost classifier provide precision of 0.966 for without FS, mRMR and RecursiveMR.

Table 4: Performance Evaluation Based on Precision

Method	Random Tree	REP Tree	JRip	Logit Boost
Without FS	0.916	0.912	0.95	0.966
RFE	0.949	0.942	0.944	0.953
mRMR	0.953	0.951	0.952	0.966
Recursive MR	0.959	0.951	0.955	0.966

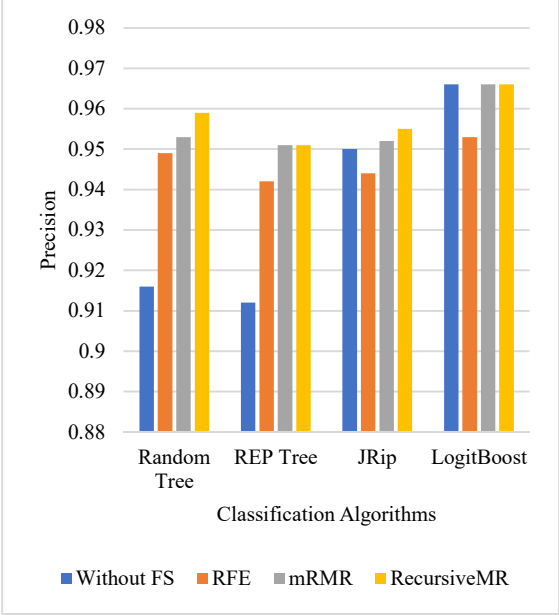

Figure 8: Comparison of performance based on Precision

Proposed work is compared with previous works on anxiety disorder dataset. All the datasets are collected on different survey. Here analysis made on which model able to find out a model with minimum number of features and providing better accuracy. In [5] feature selection based on Sharpley value is done and performance measured with classifiers Naïve Bayes, Random Forest and J48.Here 8 features are selected through feature selection. Highest accuracy obtained is 95.70% using J48 classification.

Classification of anxiety disorder dataset is done using Catboost algorithm in [4]. Here 14 features are used for the analysis. Highest accuracy obtained using 5-fold cross validation and the accuracy is 89.30%. A deep learning model is proposed in [7]. Here Sensitivity analysis is used along with neural network model. Here the author used 13 features for the analysis. Model provides a good accuracy.

Authors use Naïve Bayes,SVM,KNN and Decision tree for the analysis in [9]. Here best two methods are shown in the Table 5. Here first particle swarm optimization method is used for the feature selection, selected 14 features. Then a combined feature selection model of particle swarm optimization and Fuzzy rough set method is used, which select 4 features. These four features are used for the classification and found Naïve bayes and decision tree gave better accuracy that is 93.96%.

Table 5: Performance Comparison

Authors	Classification	Approach	Number of features	Accuracy (in %)
Neesha Jothi,Wahidah Husain,Nur'Aini Abdul Rashid in 2020[5]	Naïve Bayes	Sharply value	8	80.70
	Random Forest	Sharply value	8	90.50
	J48	Sharply value	8	95.70
Arkaprabha Sau,Ishita Bakta in 2019[4]	10-fold cross validation	Catboost	14	82.60
	5-fold cross validation	Catboost	14	89.30
Sribala M in 2015[7]	Neural Network	Sensitivity analysis	13	96.43
Wahid Husain,Saw Hui Yug,Nur'Aini Abdul Rashid, Neesha Jothi in 2017[9]	Naïve Bayes	PSO	14	93.41
		PSO+FRS	4	93.96
	Decision Tree	PSO	14	90.11
		PSO+FRS	4	93.96
In this work	LogitBoost	RFE	7	95.3
		mRMR	7	9.6.6
		RecursiveMR	7	96.6
	JRip	RFE	7	94.4
		mRMR	7	95.1
		RecursiveMR	7	95.4
	REP Tree	RFE	7	94.3
		mRMR	7	95.1
		RecursiveMR	7	95.1
	Random Tree	RFE	7	95
		mRMR	7	95.4
		RecursiveMR	7	95.9

In this work three feature section methods are used to selected most important 7 features. These features are tested with classifiers MRMR and RecursiveMR provided best accuracy. But as compared to other features like recall and precision it is observed thar proposed model RecursiveMR have better performance measures.

The main achievement of this study is that it is a generalized model for the anxiety prediction. Both males and females in different age group are included in this study. Most of the previous work is concentrated on a particular category of people, so model can used only for the prediction of that particular category. From the TABLE 5 it is very clear that proposed model performed well as compared to previous works.

4. CONCLUSION AND FUTURE SCOPE

Generalized Anxiety disorders are serious issue suffering by the people in the society. So, building a prediction model for the same is very essential. The proposed hybrid feature selection approach has been tested using anxiety disorder dataset. There is an improvement in accuracy, precision, recall after feature selection. The result helped to find out highly relevant features for the prediction of generalized anxiety disorder. While using RFE feature selection the result depends on underlying classification algorithm used in RFE, this bias removed by combining with mRMR. This proposed feature selection model can also apply in the other datasets for getting better results.

Presently the model applied in binary classification problem, in future it has to be test with multiclassification problems. The imbalance in the dataset is not considered in the present work. In the future scope some balancing methods can used to improve performance measures.

REFERENCES

[1] World Health Organization,South-East Asia,India ",Mental Health", Available: https://www.who.int/india/health-topics/mental-health august 2021 ,12/08/2021

[2] Our Better World,Singapore International foundation, "Mental Health in Asia: The numbers",Available: https://www.ourbetterworld.org/series/mental-health/facts/mental-health-asia-numbers?type=resource&gclid=Cj0KCQjwvaeJBhCvARIsABgTDM4VCyZufbeVgLI9osms3n0kidZL7uuTyGQz3z8L9ca5GQEmz8nC-80aAhGHEALw_wcB,12/08/2021

[3] Public Health Canada, "Mental Health - anxiety disorders".Available: https://www.canada.ca/en/health-canada/services/healthy-living/yourhealth/diseases/mental-health-anxiety-disorders.html,24/09/2021

[4] LIU, Huan and YU, Lei." Toward integrating feature selection algorithms for classification and clustering", *IEEE Transactions on knowledge and data engineering*, vol. 17, no 4, 2005, pp. 491-502

[5] Sofiane MAZA and Mohamed TOUAHRIA," Feature Selection Algorithms in Intrusion Detection System: A Survey*," KSII Transactions on Internet and Information Systems,* vol. 12, no. 10, 2018, pp. 5079-5099.

[6] ZOUACHE, Djaafar and ABDELAZIZ, Fouad Ben." A cooperative swarm intelligence algorithm based on quantum-inspired and rough sets for feature selection", *Computers & Industrial Engineering*, vol.115, ,2018 pp. 26-36

[7] Veerabhadrappa, L.Rangarajan, "Bi-level Dimensionality Reduction Methods Using Feature Selection and Feature Extraction", *International Journal of Computer Applications*, Vol. 2, ,2010,pp. 33-38

[8] Arkaprabha Sau,Ishita Bakta ,"Screening of anxiety and depression among the seafarers using machine learning technology", *Informatics in Medicine Unlocked*,Volume 16, ,2019 pp-25-36

[9] Neesha Jothi, Wahidah Husain, Nur,Aini Abdul Rashid,"Predicting generalized anxiety disorder among women using Sharpley value", *Journal of Infection and Public Health*, Vol 14,No.1, .2020,pp.103-108

[10] Silviya D'monte,Dakshata Panchal,"Data Mining Approach for Diagnose of Anxiety Disorder", *International Conference on Computing, Communication and Automation*,2015,pp.45-67

[11] M.Saribala , "An approach of artificial neural networks for prediction of generalized anxiety disorder*", International journal of research in computing and robotics*,Vol.1,Issue 3,March 2015,pp:118:124

[12] Neesha Jothi,Wahidah Husain,Nur'Aini Abdul Rashid,Lee Ker Xin," Predicting generalized anxiety disorder among women using decision tree based classification*", International journal of Business Information Systems*,Vol.29,No.1,2018,pp.67-78

[13] Wahidah Husain, Saw Hui Yng, Nur'Aini Abdul Rashid, and Neesha Jothi."Prediction of Generalized Anxiety Disorder Using Particle Swarm Optimization",*Advances in Intelligent Systems and Computing,*Springer International Publishing,2017.pp.35-56

[14] Yudong Cai et.,al "Prediction of lysine ubiquitination with mRMR feature selectionand analysis" *Amino Acids*,2018,pp.78-87

[15] Hanchuan Peng,Fuhui Long,Chris Ding," Feature Selection based on Mutual Information:Criteria of Max-Dependency,Max Relevance and Min-Redundancy",*IEEE transactions on pattern analysis and machine intelligence,* Vol. 27, NO. 8, August 2005,pp.89-97

[16] Inshik Jo,Sanbum Lee and Sejong Oh "Improved Measures of Redundancy and Relevance for mRMR Feature Selection" *Computers* Vol 8,Issue 42,2008,pp.23-45

[17] Standford Medicine,Stanford university, "Patient Health Questionnaire(PHQ-9)".Available:https://med.stanford.edu/PHQ9id.pdf, 21/10/2021

[18] Sexual Biolence Research Initiative "Hospital Anxiety and Depression Scale (HADS)", Available:https://www.svri.org/sites/default/fil

es/attachmente/2016-01-13/HAD.pdf , 11/10/2021

[19] Anil Rajput,Ramesh Prasad Aharwal,Meghns Dubey,S P Saxena,Manmohan Raghuvanshi, "J and JRIP Rules for E Govemence Data", *International Journal of Computer Science and Security(IJCSS)*,Vol 5, No 2 ,2011

[20] Sushilkumar Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News*", IJISET - International Journal of Innovative Science, Engineering & Technology*, , February 2015, Vol. 2 Issue 2,pp-45-67

[21] Dr. B. Srinivasan,P.Mekala, "Mining Social Networking Data for Classification Using Reptree", *International Journal of Advance Research in Computer Science and Management Studies*, Volume 2, Issue 10, October 2014,pp-78-90

[22] T. T. Wong, " Performance evaluation of classification algorithms by k -fold and leave-one-out cross validation ", *Pattern Recognit.*, vol. 48, no. 9, 2015pp. 2839-2846.

[23] Ming Yin,Jennifer Wortman Vaughan,Hanna Wallah,Understanding the effect of Accuracy on Trust in Machine learning models *Proceedings of the 2019 CHI conference on Human Factors in computing systems*, May 2019,pp-1-12, .