Classification of Imbalanced Class Distribution using Random Forest with Multiple Weight Based Majority Voting for Credit Scoring

Ramila RajaLeximi Pannir Selvam ,Irfan Ahmed Mohammed Saleem , Ahmed Alenezi

Abstract: ***Classification is an important and most widely used technique in predicting class labels for given unlabelled instances. In this field of supervised learning, most standard classification algorithms provide better accuracy for balanced class distribution. However, in case of sensitive real world applications, especially, credit scoring and medical diagnosis containing imbalanced class samples, the standard algorithms normally produce higher misclassification rate for the minority class samples which is the field of interest of the user since, collecting the data for minority samples are equally rare and costly than majority class samples. This paper introduces an improvisation on the random forest algorithm by introducing multiple weight based majority voting that suits best for credit scoring datasets. The proposed algorithm has been evaluated and compared with other variations of random forest methods and it is proved that the proposed method improves overall performance and accuracy in predicting both majority and minority class labels.***

Index Terms: Credit Scoring; Imbalanced Dataset; Classification; Random Forest; Multiple Weight; Majority Voting.

I. INTRODUCTION

Due to the economic growth and technological evolution, the financial institutions allow their costumer in large number to invest and borrow for shaping their future. To invest or to save money, the customers look for the standard capital markets. However, in the case of lending, the customers are intensely verified to identify defaulters from non-defaulters to avoid credit risks. Normally, credit risks are identified based on assigning the score for the individual customer and is termed as a credit score. Commonly, details about the repayments of debts of a customer are stored in the repository for analyzing their responsibilities. These details are generally known as credit history which is highly significant in calculating credit risk.

Revised Manuscript Received on December 22, 2018.

Ramila RajaLeximi Pannir Selvam *.Research and Development Centre*
Bharathiar University, Coimbatore 641046, India pslexiimi@gmail.com
Irfan Ahmed Mohammed Saleem *, Department of Computer and Information Sciences,* College of Science and Arts, Taibah University,
Al Ula, Madhina , msirfan@gmail.com
Ahmed Alenezi,*Department of Computer and Information Sciences*
College of Science and Arts, Taibah University,
Al Ula, Madhina,*alenezi2015@gmail.com*

Credit score, which is calculated based on a customer's past performance and present credit status is an empirical factor for financial institutions in sanctioning loans and approving credit transactions to avoid a heavy loss. Credit scoring has become the topmost evaluation tool in financial creditability to assess and decrease the potential credit risk, to increase the cash flow rate, and to make predictions and decisions [1]. Conversely, the primary challenge in credit scoring is classifying the customers as 'good' or 'bad' and to make a decision whether to sanction the loan or not based on the trained samples [2]. Thus, it becomes the utmost important field of application for data mining, machine learning and even operational research [3].

The input dataset may be created from available internal and many external sources of data. The parameters include 1) demographic information such as Aadhar ID, age, address, time at residence, job address, time at job, postal code, etc. 2) existing relationship with bank such as time at bank, number and types of accounts, payment routine, previous claim, current credit status, and 3) credit bureau such as inquiries, trades, delinquency, public records and so forth. [4]. The collected input samples are said to be the binary class imbalanced dataset, since there are only two classes that are either 'good' (non-defaulters) or 'bad' (defaulters) in which the class 'good' has majority of instances whereas the class 'bad' has minimum number of instances [5]. Furthermore, in the field of supervised learning, the standard algorithms work well and provide better prediction accuracy on balanced dataset whereas, in case of an imbalance dataset, the algorithms are subjective towards the majority class samples which lead to higher misclassification rate for the minority class samples [8]. On the other hand, the minority class gains maximum interestingness to be learned and therefore involves a great cost for misclassification.

However, the standard classifiers achieve a good performance accuracy for imbalanced datasets, it may not accurate due to the domination of majority class samples. Owing to the skewed distribution of classes, though the classifier produces 90% of accuracy, the classification may not be accurate if the classifier predicts all the instances as majority class. This occurs based on the following reasons:

- Use of performance metrics that are biased towards the majority classes.
- Predicted classification rules for minority class are very small in number and are discarded.
- Few minority class instances are identified as noise and are discarded which even reduces their count.

Thus the significance of classification techniques for imbalanced class is gaining focus due to the sensitives of the real world applications such as credit scoring ([15], [16]), medical diagnosis ([9], [10]), fraud detection [11], face recognition ([12], [14]) and outliers in web contents [13].

In this paper, an improved random forest algorithm using multiple weight based majority voting technique (RF – MWMV) that improves the accuracy of the classifier for imbalanced credit score datasetshave been suggested.The proposed method has been analyzed based on the 4 real world credit score datasets by varying the imbalanced ratio of majority and minority class instances. The proposed method is compared with the traditional random forest and with weighted random forest algorithms. Based on the analysis, it is proved that the proposed method provides better accuracy and precision rates.

The organization of the papers is as follows. Section II describes the existing techniques based on the literature survey. Section III presents the concept of random forest. Section IV explains the proposed random forest with multiple weight based majority voting (RF-MWMV) technique. The detailed experimental analysis is presented in Section V. Finally, Section VI ends with concluding remarks.

II. LITERATURE SURVEY

Several techniques have been proposed by various researchers for classifying instances in the imbalanced datasets. Basically, minority classes are the study of interest in many imbalanced datasets and unluckily the misclassification rate will be high in minority classes due to the domination of majority class samples. The imbalanced class problem can be handled in two ways, one at the data level and then another at the algorithmic level.

The former method uses various resampling techniques in which the main focus is to balance the dataset thereby either increasing the minority class distribution by replicating the minority class samples or decreasing the majority class distribution by eliminating the majority class samples. The drawbacks of these methods are that in the case of under-sampling, hypothetically useful information may get discarded whereas under sampling leads to over fitting [18]. The most popular method is the Synthetic Minority Over-Sampling Technique (SMOTE) [17]. The algorithm generates new instances instead of replicating the instances for the minority class. This method avoids over fitting but introduces noise. Additionally, replicating or generating data for minority class in sensitive applications will not provide better predictions.

The later method introduces new algorithms or modifies existing algorithms that better fits the unbalanced dataset [19]. Ensemble classifiers are introduced for classification that uses several machine learning algorithms for categorizing or predicting the instances in imbalanced datasets using bagging, boosting and stacking techniques [20]. Most of the ensemble classifiers use the same base learning algorithm to produce homogeneous ensembles and many others employ different learning algorithm to generate heterogeneous ensembles.

Bagging refers to bootstrap aggregation that reduces the variance by averaging multiple predictions. The most widely used algorithm is Random Forest [21]. Boosting techniques boost the weak learners into strong learners by applying weights. Adaptive Boosting (Adaboost) is an example of boosting techniques [22]. Stacking, also called as super learning combines multiple heterogeneous classifications using meta-classifier in which the base classifiers are trained based on a training set, then the meta-classifier is trained on the outputs of the base classifiers as features [23].

III. RANDOM FOREST CLASSIFIER ALGORITHM

In general, random forest classifier involves building multiple random trees, in which each tree in the group is built based on the random sample drawn with replacement from the training set and predicts each instance from the test data. Furthermore, only a random subset of attributes is selected for each tree instead of employing all the attributes to further randomize the tree. Finally, the algorithm predicts the class having the majority votes [25].

Many researchers proved that the random forest outperforms than other ensembles of classifiers such as bagging, boosting and Support vector machine (SVM), particularly for disease prediction and fraud detection from an imbalanced dataset and gains more significance ([26], [27]). Reference [28] suggested a weighted vote for tree aggregation in which the method uses sensitivity and specificity as classification metrics in weight calculation. Reference [29] suggested the weighted voting for query access detection by applying weight for the trees and then embeds with probabilistic classification process for predicting the class which will be suitable for balanced datasets. The decision mechanism of Random Forests is replaced with a consensus decision making [30]. Reference [31] introduce class weights in the random forest to address the imbalanced problem in medical datasets, by assigning individual weights based on Area under the Curve (AuC) for each class instead of a single weight.

All the above methods introduce weighting technique that employs a single weight based on classification rate for the classifiers. In this proposed method, multiple weights are calculated for each classifier and for each class, based on various classification metrics that are aggregated to predict the instance of the class based on the highest vote. This method provides a better solution for the imbalanced binary class problem.

IV. RANDOM FOREST WITH MULTIPLE WEIGHT BASED MAJORITY VOTING (RF-MWMV)

The proposed method introduces multiple weight based majority voting for the random forest algorithm (RF-MWMV) to predict the class label of the test samples. Instead of using majority voting and single weighted

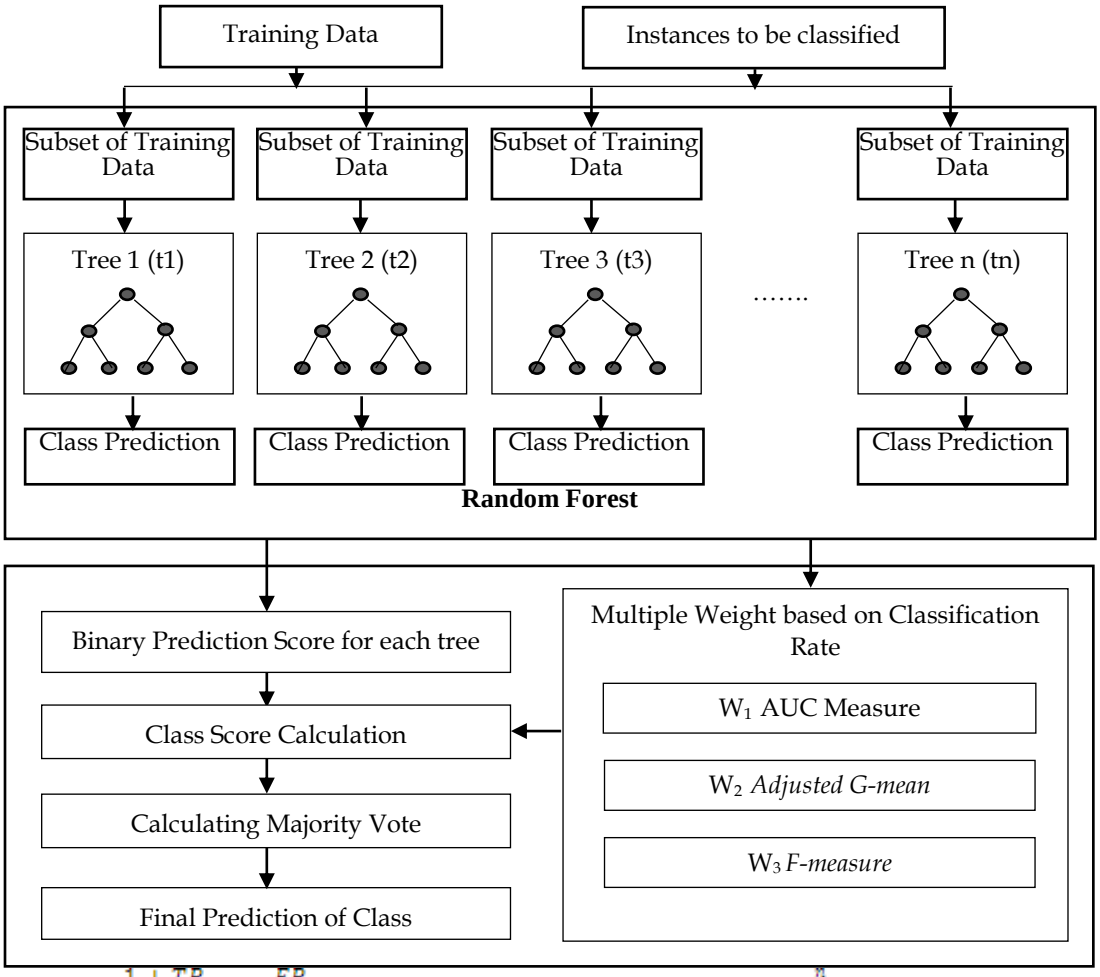
voting, this method uses multiple weight based on classification rates measured for both majority and minority class samples using three metrics which are suitable for imbalanced datasets such as AuC, Adjusted G-Mean, and F- Measure. Finally, the majority voting is applied to predict the class labels of the test instances. The procedure is depicted as a framework in Fig. 1.

Initially, the random forest is executed which generates the matrix F_{ij} where i represents the test instances ( ) and j represents the trees (t_j) with a prediction of class labels c which takes either majority class label or minority class label. The classification rates (weights) of each tree represented as AUC_t, AGM_t and FM_t is calculated using AuC measure, Adjusted G-Mean and F-Measure.

Fig. 1. *Proposed RF-MWMV Framework for Classification of Imbalanced data*

The Area under ROC Curve is calculated for all the tree classifiers in the forest using the formula given in Equation (1).

After the calculation of weights for each tree in the forest using the above three metrics, a class score of each class is calculated as in Equation (4) based on the predictions made by the trees for each instance in the test samples where i takes the values 1,2,3.



$$W_1^t: AUC_t = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (1)$$

The metric Adjusted G-Mean [32] is calculated for each tree in the forest as given in Equation (2).

$$W_2^t: AGM_t = \frac{GM + TN_{rate}(FP + TN)}{1 + FP + TN} \quad (2)$$

where, $GM_t = \sqrt{\frac{TP}{TP+FN} * \frac{\cdot}{FP}}$

The most commonly used evaluation metric F-measure is calculated for each tree as represented in Equation (3).

$$W_3^t: FM_t = \frac{(1 + \beta^2).precision.recall}{(\beta^2.precision) + recall} \quad (3)$$

$$CS_x^{c,i} = \sum_{t=1}^{n} BPS_t^c(x) * W_i^t \quad (4)$$

where Binary Prediction Score (BPS) for each class based on the predictions made by the tree classifiers is calculated [28] as in Equation (5)

$$BPS_t^c(x) = \begin{cases} 1, if\ t\ predicts\ the\ instance\ x\ as\ class\ c \\ 0, Otherwise \end{cases} \quad (5)$$

In this stage, the prediction of the class label is made based on the highest score the class has gained for each performance metrics. Finally, the class labels are predicted based on the majority voting of each metrics as in Equation (6).

Algorithm: Pseudocode for RF-MWMV

Input: Set of trees t from Random Forest RF, instance x to classify, predicted class label p for the instance x
Output: Predicted Class PC for the given instance x

Function RF_MWMV (RF, x, p)
Begin
For each tree classifiers t from RF
//Calculate the classification rate using performance metrics suitable for imbalanced class
$Weight_1(t)$ = Area under ROC Curve(t)
$Weight_2(t)$ = Adjusted Geometric Mean(t)
$Weight_3(t)$ = F-Measure(t)
End For
For each class label c in the dataset
For each tree classifier t from RF
//Calculate the Binary Prediction Score for each tree t and class c
If (predicted class label p == the chosen class label c)
$BPS_{c,t} = 1$
Else
$BPS_{c,t} = 0$
End If
For each weight i used in calculating performance metrics
//Calculate the Class Score for each tree t and class c
$CS_{c,i} = \sum_t BPS_{c,t} * Weight_i$
End For
End For
For each weight i used in calculating performance metrics
//Identify the class having maximum score for each performance metrics
C_{wi} = Arg Max ($CS_{c,i}$)
//Identify the class having majority votes with respect to the three performance metrics
PC = mode(C_{wi})
End For
End For
Return Predicted Class PC
End Function

Fig. 2. *Pseudocode for Random Forests with Multiple Weight based Majority Voting*

V. EXPERIMENTAL ANALYSIS

Any research is valid only if it proved to be efficient and produces good results when compared with existing techniques. The experimental analysis has been carried out to analyze the performance of the proposed model. The datasets used for the experimental analysis are described in Table I and are publically available at UCI [6] and KEEL repository [7].

$$PC_x = mode\left((C^x_{w1}), (C^x_{w2}), (C^x_{w3})\right) \quad (6)$$

where, $C^x_{w_i} = Arg\ \max\left(CS^{c,i}_x\right) where\ i = 1,2,3$

The algorithm for the proposed RF-MW MV is given in Fig. 2.

Though there are several performance measures available to evaluate the classification algorithm, generally, accuracy is considered as the traditional measure to calculate the classification rate. This may not be a good measure in the case of an imbalanced class problem as it does not distinguish the correctly classified samples for both majority and minority classes. The metrics have to be chosen in such a way that it should make class distribution into consideration. Reference [8] suggest several metrics such as Area under the ROC Curve (AuC), Geometric Mean of the true rates, F-measure, Adjusted G-mean, Dominance, and Index of Balanced Accuracy. As the misclassification of minority class label is considered as the concept of interest in any field of classification containing imbalanced binary class distribution, many classifier algorithms classify the minority class label incorrectly than that of the majority class label.

The proposed algorithm RF – MWMV is compared with existing random forest algorithm (RF) [25] and single weighed random forest algorithm (WRF) [28] using several performance metrics [24]. Table II depicts the performance evaluation of the proposed and existing algorithms for all the dataset listed in Table I. The performance metrics that are suitable for imbalanced datasets are considered for the assessment of the algorithms [8].

From Table II, the proposed method improves the classification accuracy and other measures than the existing algorithms for all the four datasets. The overall classification rate of the majority class in all the experiments made using RF, WRF, and RF-MWMV are 85.1%, 85.53%, and 85.92% respectively, whereas, the classification rate of the minority class using RF, WRF and RF-MWMV are 77.37%, 78.55% and 79.29% respectively. The graph for the experimental result is depicted in Fig.3. Performance for Australian Credit Dataset, German Credit Dataset, Japanese Credit Dataset and HMEQ Dataset are shown in (a), (b), (c) and (d) of Fig.3.

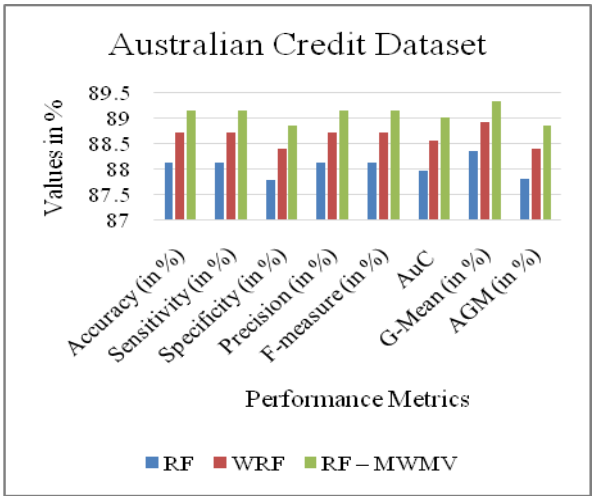
As the imbalanced dataset has the larger proportion of majority class instance than the minority class instance, the experimental analysis has been made by varying the imbalance ratio for Australian and German credit dataset.

TABLE I. DATASET DESCRIPTION

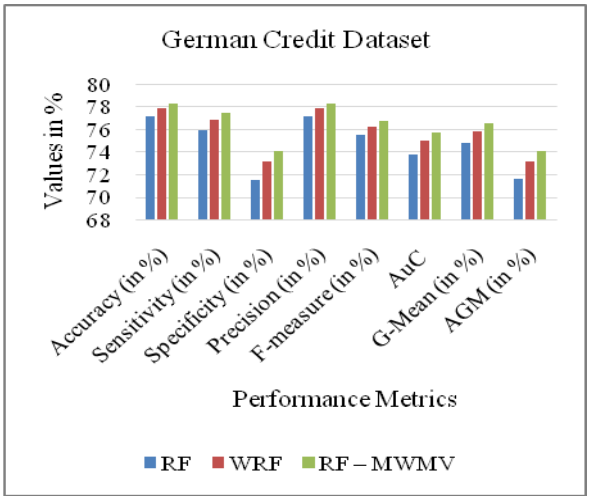
Name of the Dataset	No. of Attributes	No. of Features Selected	Total Instances	No. of Majority Class Instances	No. of Minority Class Instances
Australian Credit Dataset	14	12	690	383	307
German Credit Dataset	20	16	1000	700	300
Japanese Credit Approval	15	13	690	383	307
HMEQ	12	10	5960	4771	1189

TABLE II. COMPARISON OF ALGORITHMS USING STATISTICAL MEASURE

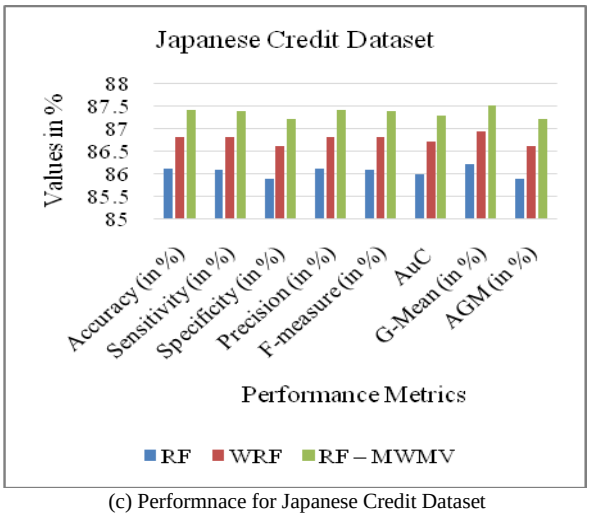
Algorithms	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)	Precision (in %)	F-measure (in %)	AuC (in %)	G-Mean (in %)	AGM (in %)
Australian Credit Dataset								
RF	88.12	88.12	87.78	88.12	88.12	87.95	88.33	87.79
WRF	88.70	88.70	88.38	88.70	88.70	88.54	88.90	88.38
RF – MWMV	**89.13**	**89.13**	**88.84**	**89.13**	**89.13**	**88.99**	**89.31**	**88.84**
German Credit Dataset								
RF	77.10	75.94	71.52	77.10	75.47	73.73	74.76	71.54
WRF	77.80	76.80	73.09	77.80	76.16	74.95	75.83	73.11
RF – MWMV	**78.30**	**77.40**	**74.03**	**78.30**	**76.72**	**75.71**	**76.52**	**74.05**
Japanese Credit Dataset								
RF	86.09	86.07	85.87	86.09	86.07	85.97	86.19	85.87
WRF	86.81	86.80	86.59	86.81	86.80	86.70	86.92	86.59
RF – MWMV	**87.39**	**87.38**	**87.19**	**87.39**	**87.38**	**87.28**	**87.49**	**87.19**
HMEQ Dataset								
RF	88.07	89.27	95.53	88.07	86.05	92.40	91.46	95.52
WRF	88.17	89.38	95.72	88.17	86.18	92.55	91.60	95.71
RF – MWMV	**88.22**	**89.45**	**95.89**	**88.22**	**86.25**	**92.67**	**91.70**	**95.88**



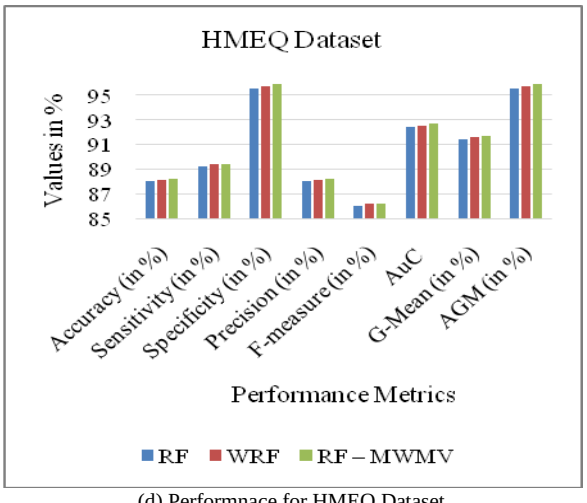
(a) Performanace for Australian Credit Dataset



(b) Performnace for German Credit Dataset



(c) Performnace for Japanese Credit Dataset



(d) Performnace for HMEQ Dataset

Fig.3. *Comparison of the algorithms using various datasets*

The imbalance ratio (IR) is described as the ratio of the number of minority class instances to the number of majority class instances. For the imbalance ratio 1:2, the imbalanced percentage is 50% which implies for 383 instances of the majority class, 192 instances of minority class (represented as Min-Class Instance Count) are considered under Australian dataset for the study. The detailed analysis has been shown in Table III for the proposed and existing algorithms by varying the imbalance ratio {1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12} with minority class instances of Australian credit dataset. In Table III, 11 experiments by varying the IR for Australian dataset has been made, out of which the proposed method provides higher accuracy rate than the traditional random forest (RF) for all the cases and better classification accuracy than single weighted random forest (WRF) for 9 cases. The highest classification accuracy values and other values are given in bold style.

TABLE III. COMPARISON OF ALGORITHMS WITH DIFFERENT IR FOR AUSTRALIAN DATASET

Algorithms	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)	Precision (in %)	F-measure (in %)	AuC (in %)	G-Mean (in %)	AGM (in %)
Imbalanced Ratio: 1:2	*Imbalanced Percentage: 50%*	*Min-Class Instance Count :192*						
RF	88.00	87.93	85.42	88.00	87.96	86.68	87.47	85.43
WRF	88.70	88.63	86.30	88.70	88.66	87.47	88.21	86.32
RF – MWMV	**89.74**	**89.69**	**87.63**	**89.74**	**89.70**	**88.66**	**89.31**	**87.64**
Imbalanced Ratio: 1:3	*Imbalanced Percentage: 33%*	*Min-Class Instance Count :128*						
RF	88.65	88.35	83.62	88.65	88.39	85.99	86.79	83.66
WRF	89.24	88.98	84.53	89.24	89.01	86.75	87.51	84.56
RF – MWMV	**89.82**	**89.61**	**85.05**	**89.82**	**89.65**	**87.33**	**88.11**	**85.08**
Imbalanced Ratio : 1:4	*Imbalanced Percentage : 25%*	*Min-Class Instance Count :96*						
RF	90.81	90.49	83.91	90.81	90.54	87.20	87.96	83.97
WRF	91.02	90.72	84.13	91.02	90.77	87.43	88.19	84.19
RF – MWMV	**91.65**	**91.40**	**85.35**	**91.65**	**91.44**	**88.38**	**89.08**	**85.41**
Imbalanced Ratio : 1:5	*Imbalanced Percentage : 20%*	*Min-Class Instance Count :77*						
RF	90.87	90.45	79.64	90.87	90.55	85.04	85.89	79.75
WRF	91.09	90.70	79.97	91.09	90.81	85.34	86.18	80.08
RF – MWMV	**91.30**	**90.91**	**80.97**	**91.30**	**91.00**	**85.94**	**86.74**	**81.07**
Imbalanced Ratio : 1:6	*Imbalanced Percentage : 17%*	*Min-Class Instance Count :64*						
RF	91.28	90.72	77.22	91.28	90.85	83.97	84.73	77.39
WRF	91.95	91.49	79.36	91.95	91.59	85.42	86.14	79.51
RF – MWMV	**92.84**	**92.47**	**82.73**	**92.84**	**92.53**	**87.60**	**88.23**	**82.85**
Imbalanced Ratio : 1:7	*Imbalanced Percentage : 14%*	*Min-Class Instance Count :55*						
RF	92.24	91.65	77.91	92.24	91.74	84.78	85.40	78.12
WRF	92.47	91.90	79.57	92.47	91.94	85.73	86.33	79.76
RF – MWMV	**92.69**	**92.18**	**80.08**	**92.69**	**92.23**	**86.13**	**86.72**	**80.26**
Imbalanced Ratio : 1:8	*Imbalanced Percentage : 13%*	*Min-Class Instance Count :48*						
RF	92.58	91.87	77.11	92.58	91.91	84.49	85.00	77.38
WRF	93.27	92.73	80.53	93.27	92.71	86.63	87.12	80.75
RF – MWMV	**93.50**	**93.01**	**81.12**	**93.50**	**93.01**	**87.06**	**87.55**	**81.32**
Imbalanced Ratio : 1:9	*Imbalanced Percentage : 11%*	*Min-Class Instance Count :43*						
RF	92.25	91.24	73.12	92.25	91.25	82.18	82.57	73.55
WRF	**93.19**	**92.50**	**78.67**	**93.19**	**92.43**	**85.59**	**86.01**	**78.98**
RF – MWMV	**93.19**	**92.50**	**78.67**	**93.19**	**92.43**	**85.59**	**86.01**	**78.98**
Imbalanced Ratio : 1:10	*Imbalanced Percentage : 10%*	*Min-Class Instance Count :38*						
RF	93.35	92.52	74.69	93.35	92.53	83.61	83.91	75.13
WRF	94.06	93.47	79.77	94.06	93.39	86.62	86.97	80.10
RF – MWMV	**94.54**	**94.06**	**81.39**	**94.54**	**94.02**	**87.73**	**88.08**	**81.68**
Imbalanced Ratio : 1:11	*Imbalanced Percentage : 9%*	*Min-Class Instance Count :35*						
RF	93.78	92.97	75.45	93.78	92.92	84.21	84.46	75.94
WRF	94.26	93.62	77.78	94.26	93.59	85.70	85.98	78.19
RF – MWMV	**94.50**	**93.93**	**78.80**	**94.50**	**93.92**	**86.36**	**86.66**	**79.17**
Imbalanced Ratio : 1:12	*Imbalanced Percentage : 8%*	*Min-Class Instance Count :32*						
RF	93.49	92.35	68.81	93.49	92.39	80.58	80.54	69.62
WRF	**93.98**	**93.10**	**74.98**	**93.98**	**92.95**	**84.04**	**84.22**	**75.61**
RF – MWMV	**93.98**	**93.10**	**74.98**	**93.98**	**92.95**	**84.04**	**84.22**	**75.61**

Correspondingly, the analysis has been made with the German credit dataset for the proposed and existing algorithms by varying the imbalance ratio (1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12} with minority class instances. The details are presented in Table IV.

TABLE IV. COMPARISON OF ALGORITHMS WITH DIFFERENT IR FOR AUSTRALIAN DATASET

Algorithms	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)	Precision (in %)	F-measure (in %)	AuC (in %)	G-Mean (in %)	AGM (in %)
Imbalanced Ratio: 1:3	*Imbalanced Percentage: 33%*				*Min-Class Instance Count :233*			
RF	77.71	75.35	67.60	77.71	73.85	71.48	72.65	67.68
WRF	78.14	76.04	69.07	78.14	74.43	72.56	73.64	69.14
RF – MWMV	**78.67**	**76.92**	**71.02**	**78.67**	**75.10**	**73.97**	**74.92**	**71.08**
Imbalanced Ratio : 1:4	*Imbalanced Percentage : 25%*				*Min-Class Instance Count :175*			
RF	81.37	78.72	68.43	81.37	76.34	73.58	74.60	68.62
WRF	81.71	79.56	71.11	81.71	76.84	75.33	76.23	71.26
RF – MWMV	**82.06**	**80.36**	**73.66**	**82.06**	**77.34**	**77.01**	**77.76**	**73.78**
Imbalanced Ratio : 1:5	*Imbalanced Percentage : 20%*				*Min-Class Instance Count :140*			
RF	84.64	83.41	77.64	84.64	79.57	80.53	81.04	77.82
WRF	84.88	83.93	79.40	84.88	80.04	81.66	82.08	79.53
RF – MWMV	**85.12**	**84.76**	**83.04**	**85.12**	**80.35**	**83.90**	**84.07**	**83.09**
Imbalanced Ratio : 1:6	*Imbalanced Percentage : 17%*				*Min-Class Instance Count :117*			
RF	86.05	83.46	69.47	86.05	80.57	76.46	77.19	70.39
WRF	86.29	84.51	74.70	86.29	81.10	79.60	80.22	75.24
RF – MWMV	**86.54**	**85.29**	**78.32**	**86.54**	**81.62**	**81.81**	**82.29**	**78.66**
Imbalanced Ratio : 1:7	*Imbalanced Percentage : 14%*				*Min-Class Instance Count :100*			
RF	88.00	87.45	83.92	88.00	83.06	85.68	85.92	84.25
WRF	88.38	88.43	88.82	88.38	83.88	88.63	88.60	88.80
RF – MWMV	**88.63**	**88.88**	**90.62**	**88.63**	**84.40**	**89.75**	**89.62**	**90.52**
Imbalanced Ratio : 1:8	*Imbalanced Percentage : 13%*				*Min-Class Instance Count :88*			
RF	88.96	85.66	60.73	88.96	84.55	73.19	73.37	62.55
WRF	89.47	87.81	74.62	89.47	85.65	81.22	81.70	75.29
RF – MWMV	**89.72**	**88.48**	**78.38**	**89.72**	**86.17**	**83.43**	**83.87**	**78.82**
Imbalanced Ratio : 1:9	*Imbalanced Percentage : 11%*				*Min-Class Instance Count :78*			
RF	89.85	85.52	47.61	89.85	85.85	66.56	65.22	50.16
WRF	90.23	87.71	65.31	90.23	86.49	76.51	76.72	66.75
RF – MWMV	**90.62**	**89.05**	**74.55**	**90.62**	**87.30**	**81.80**	**82.20**	**75.26**
Imbalanced Ratio : 1:10	*Imbalanced Percentage : 10%*				*Min-Class Instance Count :70*			
RF	90.91	87.38	53.74	90.91	87.06	70.56	69.75	57.37
WRF	**91.30**	**89.66**	**73.25**	**91.30**	**87.93**	**81.45**	**81.77**	**74.44**
RF – MWMV	91.17	89.12	68.91	91.17	87.64	79.01	79.23	70.56
Imbalanced Ratio : 1:11	*Imbalanced Percentage : 9%*				*Min-Class Instance Count :64*			
RF	91.88	92.55	99.32	91.88	88.25	95.93	95.55	97.95
WRF	92.02	92.66	99.33	92.02	88.55	95.99	95.61	98.32
RF – MWMV	**92.15**	**92.77**	**99.34**	**92.15**	**88.84**	**96.05**	**95.68**	**98.54**
Imbalanced Ratio : 1:12	*Imbalanced Percentage : 8%*				*Min-Class Instance Count :58*			
RF	92.48	93.05	99.42	92.48	88.99	96.24	95.91	97.52
WRF	**92.74**	**93.27**	**99.44**	**92.74**	**89.61**	**96.36**	**96.04**	**98.52**
RF – MWMV	**92.74**	**93.27**	**99.44**	**92.74**	**89.61**	**96.36**	**96.04**	**98.52**

For German credit dataset, as the dataset has the minimum number of minority class samples, only 10 experimental results have been performed by varying the IR out of which the proposed method provides a higher accuracy rate than the traditional random forest (RF) for all the cases and a better classification accuracy than single weighted random forest (WRF) for 8 cases. The WRF algorithm produces a better result than the proposed for the imbalanced ratio 1:10.

From the analysis presented in Table III and Table IV, it is clear that the proposed algorithm outperforms well than the traditional random forest classifier and single weighted random forest. When the number of instances of the minority class is very low, the proposed algorithm RF – MWMV works similar to the WRF. In all other cases, the proposed method provides better results.

Also, the details of precision, recall, and f-measure for the majority class instances and the minority class instances for the conducted experiment have been presented in Table V and Table VI for Australian and German dataset respectively.

TABLE V. COMPARISON OF PRECISION, RECALL AND F-MEASURE OF MAJORITY AND MINORITY CLASS WITH DFFERENT IR FOR AUSTRALIAN DATASET

Algorithms	Precision		Recall		F-Measure	
	Majority Class	Minority Class	Majority Class	Minority Class	Majority Class	Minority Class
Imbalanced Ratio: 1:2 Imbalanced Percentage: 50% Min-Class Instance Count :192						
RF	91.64	80.73	90.46	82.89	91.05	81.79
WRF	92.17	81.77	90.98	83.96	91.57	82.85
RF – MWMV	**92.95**	**83.33**	**91.75**	**85.56**	**92.35**	**84.43**
Imbalanced Ratio: 1:3 Imbalanced Percentage: 33% Min-Class Instance Count :128						
RF	94.52	71.09	90.73	81.25	92.58	75.83
WRF	**94.78**	72.66	91.21	82.30	92.96	77.18
RF – MWMV	**94.78**	**75.00**	**91.90**	**82.76**	**93.32**	**78.69**
Imbalanced Ratio : 1:4 Imbalanced Percentage : 25% Min-Class Instance Count :96						
RF	96.08	69.79	92.70	81.71	94.36	75.28
WRF	96.08	70.83	92.93	81.93	94.48	75.98
RF – MWMV	**96.34**	**72.92**	**93.42**	**83.33**	**94.86**	**77.78**
Imbalanced Ratio : 1:5 Imbalanced Percentage : 20% Min-Class Instance Count :77						
RF	96.08	64.94	93.16	76.92	94.60	70.42
WRF	96.08	**66.23**	93.40	77.27	94.72	71.33
RF – MWMV	**96.34**	**66.23**	**93.42**	**78.46**	**94.86**	**71.83**
Imbalanced Ratio : 1:6 Imbalanced Percentage : 17% Min-Class Instance Count :64						
RF	96.61	59.38	93.43	74.51	94.99	66.09
WRF	96.87	62.50	93.92	76.92	95.37	68.97
RF – MWMV	**97.39**	**65.63**	**94.43**	**80.77**	**95.89**	**72.41**
Imbalanced Ratio : 1:7 Imbalanced Percentage : 14% Min-Class Instance Count :55						
RF	97.39	56.36	93.95	75.61	95.64	64.58
WRF	**97.65**	56.36	93.97	77.50	95.77	65.26
RF – MWMV	**97.65**	**58.18**	**94.21**	**78.05**	**95.90**	**66.67**
Imbalanced Ratio : 1:8 Imbalanced Percentage : 13% Min-Class Instance Count :48						
RF	97.91	50.00	93.98	75.00	95.91	60.00
WRF	**98.17**	54.17	94.47	78.79	96.29	64.20
RF – MWMV	**98.17**	**56.25**	**94.71**	**79.41**	**96.41**	**65.85**
Imbalanced Ratio : 1:9 Imbalanced Percentage : 11% Min-Class Instance Count :43						
RF	98.17	39.53	93.53	70.83	95.80	50.75
WRF	**98.43**	**46.51**	**94.25**	**76.92**	**96.30**	**57.97**
RF – MWMV	**98.43**	**46.51**	**94.25**	**76.92**	**96.30**	**57.97**
Imbalanced Ratio : 1:10 Imbalanced Percentage : 10% Min-Class Instance Count :38						
RF	98.43	42.11	94.49	72.73	96.42	53.33
WRF	**98.69**	47.37	94.97	78.26	96.80	59.02
RF – MWMV	**98.69**	**52.63**	**95.45**	**80.00**	**97.05**	**63.49**
Imbalanced Ratio : 1:11 Imbalanced Percentage : 9% Min-Class Instance Count :35						
RF	**98.69**	40.00	94.74	73.68	96.68	51.85
WRF	**98.69**	45.71	95.21	76.19	96.92	57.14
RF – MWMV	**98.69**	**48.57**	**95.45**	**77.27**	**97.05**	**59.65**
Imbalanced Ratio : 1:12 Imbalanced Percentage : 8% Min-Class Instance Count :32						
RF	98.69	31.25	94.50	66.67	96.55	42.55
WRF	**98.96**	**34.38**	**94.75**	**73.33**	**96.81**	**46.81**
RF – MWMV	**98.96**	**34.38**	**94.75**	**73.33**	**96.81**	**46.81**

TABLE VI. COMPARISON OF PRECISION, RECALL AND F-MEASURE OF MAJORITY AND MINORITY CLASS WITH DFFERENT IR FOR GERMAN DATASET

Algorithms	Precision		Recall		F-Measure	
	Majority Class	Minority Class	Majority Class	Minority Class	Majority Class	Minority Class
Imbalanced Ratio: 1:3 Imbalanced Percentage: 33% Min-Class Instance Count :233						
RF	95.29	24.89	79.22	63.74	86.51	35.80
WRF	95.43	26.18	79.52	65.59	86.75	37.42
RF – MWMV	**95.71**	**27.47**	**79.86**	**68.09**	**87.07**	**39.14**
Imbalanced Ratio : 1:4 Imbalanced Percentage : 25% Min-Class Instance Count :175						
RF	98.00	14.86	82.16	65.00	89.38	24.19
WRF	98.14	16.00	82.37	68.29	89.57	25.93
RF – MWMV	**98.29**	**17.14**	**82.59**	**71.43**	**89.76**	**27.65**
Imbalanced Ratio : 1:5 Imbalanced Percentage : 20% Min-Class Instance Count :140						
RF	99.29	11.43	84.86	76.19	91.51	19.88
WRF	99.29	12.86	85.07	78.26	91.63	22.09
RF – MWMV	**99.43**	**13.57**	**85.19**	**82.61**	**91.76**	**23.31**
Imbalanced Ratio : 1:6 Imbalanced Percentage : 17% Min-Class Instance Count :117						
RF	**99.57**	05.13	86.26	66.67	92.44	09.52
WRF	**99.57**	06.84	86.48	72.73	92.56	12.50
RF – MWMV	**99.57**	**08.55**	**86.69**	**76.92**	**92.69**	**15.38**
Imbalanced Ratio : 1:7 Imbalanced Percentage : 14% Min-Class Instance Count :100						
RF	**99.86**	05.00	88.04	83.33	93.57	09.43
WRF	**99.86**	08.00	88.37	88.89	93.76	14.68
RF – MWMV	**99.86**	**10.00**	**88.59**	**90.91**	**93.89**	**18.02**
Imbalanced Ratio : 1:8 Imbalanced Percentage : 13% Min-Class Instance Count :88						
RF	**99.57**	04.55	89.24	57.14	94.13	08.42
WRF	**99.57**	09.09	89.70	72.73	94.38	16.16
RF – MWMV	**99.57**	**11.36**	**89.94**	**76.92**	**94.51**	**19.80**
Imbalanced Ratio : 1:9 Imbalanced Percentage : 11% Min-Class Instance Count :78						
RF	**99.43**	03.85	90.27	42.86	94.63	07.06
WRF	**99.57**	06.41	90.52	62.50	94.83	11.63
RF – MWMV	**99.57**	**10.26**	**90.87**	**72.73**	**95.02**	**17.98**
Imbalanced Ratio : 1:10 Imbalanced Percentage : 10% Min-Class Instance Count :70						
RF	**99.71**	02.86	91.12	50.00	95.23	05.41
WRF	**99.71**	**07.14**	**91.48**	**71.43**	**95.42**	**12.99**
RF – MWMV	**99.71**	05.71	91.36	66.67	95.36	10.53
Imbalanced Ratio : 1:11 Imbalanced Percentage : 9% Min-Class Instance Count :64						
RF	**100.00**	03.13	91.86	100.00	95.76	06.06
WRF	**100.00**	04.69	91.98	100.00	95.82	08.96
RF – MWMV	**100.00**	**06.25**	**92.11**	**100.00**	**95.89**	**11.76**
Imbalanced Ratio : 1:12 Imbalanced Percentage : 8% Min-Class Instance Count :58						
RF	**100.00**	01.72	92.47	100.00	96.09	03.39
WRF	**100.00**	**05.17**	**92.72**	**100.00**	**96.22**	**09.84**
RF – MWMV	**100.00**	**05.17**	**92.72**	**100.00**	**96.22**	**09.84**

In Table V for Australian dataset, though the proposed method has only less improvement in the precision rate of the majority class samples, in case of precision for minority class, the proposed method provides a far better result than the WRF in 8 cases and better recall and f-measure for 9 experiments. However, when compared with traditional RF method, the proposed method improves the precision, recall, and f-measure in all the experiments.

In Table VI for German dataset, in case of the precision rate for minority class, the proposed method provides a better result than the WRF in 8 cases and better recall and f-measure for 9 experiments out of 10 experiments made. However, when compared with traditional RF method, the proposed method improves the precision, recall and f-measure in all the experiments.

From Table V and Table VI, it is clear that the proposed method provides better precision, recall and f-measure for the minority class instances. It obviously reduces the misclassification of minority class instances than the majority class instances which is considered as the major problem in an imbalanced dataset. The experiment is conducted specifically for the credit score dataset. Thus the proposed RF – MWMV algorithm provides a better result for the credit scoring dataset.

VI. CONCLUSION

The paper presents an improved random forest algorithm with multiple weight based majority voting has been that presented for imbalanced credit score datasets. The proposed method is compared with the traditional random forest and single weight based random forest classifiers. Several investigations have been made by varying the ratio of the minority class instances through which it is noted that the proposed method works better than the existing

methods. The overall classification rate of the majority class in all the experiments made using RF, WRF, and RF-MWMV are 85.1%, 85.53%, and 85.92% respectively, whereas, the classification rate of the minority class using RF, WRF and RF-MWMV are 77.37%, 78.55%, and 79.29% respectively. Thus it is shown that the misclassification rate of minority class instances is highly reduced using the proposed method. In future, the work can be extended to improve other ensemble based classifier that better suits the imbalanced dataset.

REFERENCES

1. C.R. Abrahams and M. Zhang. Fair Lending Compliance: Intelligence and Implications for Credit Risk Management. Wiley: Hoboken, NJ, 2008.
2. B. Benyacoub, S. EI Bernoussi, and A. Zoglat. "Building classification models for customer credit scoring. In Logistics and Operations Management (GOL)", International Conference on IEEE, 2014, June, pp. 107-111.
3. B. Baesens , C. Mues , D. Martens and J. Vanthienen J. "50 years of data mining and OR: Upcoming trends and challenges".Journal of the Operational Research Society 60(S1): 2009, pp. 816–823.
4. N. Siddiqi. Intelligent credit scoring: Building and implementing better credit risk scorecards. John Wiley & Sons, 2017.
5. V. García, A.I. Marqués, and J.S. Sánchez. "Improving risk predictions by preprocessing imbalanced credit data". In International Conference on Neural Information Processing Springer, Berlin, Heidelberg, 2012, November, pp. 68-75.
6. A. Frank and A. Asuncion, UCI machine learning repository, University of California, School of Information and Computer Science, Irvine, CA. [Online]. Available: http://archive.ics.uci.edu/ml/citation_policy.html
7. J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," J. Multiple-Valued Logic Soft Comput., vol. 17, 2011, pp. 255–287.
8. V. López, A. Fernandez, S. Garcia, V. Palade and F. Herrera, "An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics". Information Sciences 250, 2013, pp. 113-141
9. Z. Wang, V. Palade, Building interpretable fuzzy models for high dimensional data analysis in cancer diagnosis, BMC Genomics 12 ((S2):S5) ,2011.
10. A. Jain, S. Ratnoo, and D. Kumar. "Addressing class imbalance problem in medical diagnosis: A genetic algorithm approach". In Information, Communication, Instrumentation and Control (ICICIC), 2017 International Conference on IEEE, August 2017, pp. 1-8.
11. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and K. Chan, "Cost based modeling for fraud and intrusion detection: results from the jam project", In: DARPA Information Survivability Conference and Exposition, 2000, pp. 130-144.
12. N. Kwak, Feature extraction for classification problems and its application to face recognition, Pattern Recognition 41 (5), 2008, pp. 1718–1734.
13. S. Sathya Bama, M.S. Irfan Ahmed, and A. Saravanan. "A Mathematical Approach for Improving the Performance of the Search Engine through Web Content Mining". Journal of Theoretical & Applied Information Technology, 60(2), 2014
14. X. Yuan, M. Abouelenien and M. Elhoseny. A Boosting-Based Decision Fusion Method for Learning from Large, Imbalanced Face Data Set. In Quantum Computing: An Environment for Intelligent Large Scale Real Application. Springer, Cham, 2018, pp. 433-448.
15. H. He, W. Zhang and S. Zhang. A novel ensemble method for credit scoring: Adaption of different imbalance ratios. Expert Systems with Applications, 98, 2018, pp.105-117.
16. D. Tripathi, D.R. Edla and R. Cheruku. "Hybrid credit scoring model using neighborhood rough set and multi-layer ensemble classification". Journal of Intelligent & Fuzzy Systems, 34(3), 2018, pp.1543-1549.
17. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", J. Artif.Intell. Res., vol. 16, 2002, pp. 321-357.
18. O. Arbelaitz , I. Gurrutxaga , J. Muguerza , J.M. Pérez. Applying Resampling Methods for Imbalanced Datasets to Not So Imbalanced Datasets. In: Bielza C. et al. (eds) Advances in Artificial Intelligence. CAEPIA 2013. Lecture Notes in Computer Science, vol 8109. Springer, Berlin, Heidelberg, 2013.
19. W. Lin, J.J. Chen, Class-imbalanced classifiers for high-dimensional data, Briefings in Bioinformatics 14 (1) (2013)
20. Zhi-Hua Zhou, "Ensemble Methods: Foundations and Algorithms", CRC Press, 2012
21. Ho, Tin Kam. Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995.
22. Yoav Freund, Robert E. Schapire: "Experiments with a new boosting algorithm". In: Thirteenth International Conference on Machine Learning, San Francisco, 1996, pp. 148-156.
23. David H. Wolpert. Stacked generalization. Neural Networks. 5: 1992, pp. 241-259.
24. S. Sathya Bama, M. S. Irfan Ahmed, and A. Saravanan, "A Survey on Performance Evaluation Measures for information Retrieval System", International Research Journal of Engineering and Technology, Vol.2, No.2, 2015, pp.1015-1020.
25. Leo Breiman. Random Forests. Machine Learning. 45(1): 2001, pp. 5-32.
26. K. Mohammed, C. Sounak and P. Mihail, "Predicting disease risksfrom highly imbalanced data using random forest", BMC Med. Informa.Decis., Jul 2011, pp. 11-20.
27. L. Rokach, Decision forest: Twenty years of research, Information Fusion, 27 , 2016, pp. 111-125.
28. M. E. H. Daho, N. Settouti, M. E. A. Lazouni, and M. E. A. Chikh, ``Weighted vote for trees aggregation in random forest," in Proc. Int. Conf. Multimedia Comput. Syst. (ICMCS), Apr. 2014, pp. 438-443.
29. C. A. Ronao and S.-B. Cho, ``Random forests with weighted voting for anomalous query access detection in relational databases," in Artificial Intelligence and Soft Computing (Lecture Notes in Computer Science), vol. 2015. New York, NY, USA: ACM, 2015, pp. 36-48.
30. R. K. Shahzad, M. Fatima, N. Lavesson, and M. Boldt, ``Consensus decision making in random forests," in Machine Learning, Optimization, and Big Data (Lecture Notes in Computer Science). Berlin, Germany: Springer 2015, pp. 347-358.
31. M. Zhu,J. Xia,X. Jin,M. Yan,G. Cai, J. Yan and G. Ning. Class Weights Random Forest Algorithm for Processing Class Imbalanced Medical Data. IEEE Access, 6, 2018, pp.4641-4652.
32. R. Batuwita, V. Palade, adjusted geometric-mean: a novel performance measure for imbalanced bioinformatics datasets learning, Journal of Bioinformatics and Computational Biology 10 (4) (2012).