Evolving Systems
https://doi.org/10.1007/s12530-022-09434-4

ORIGINAL PAPER

Transparent rule generator random forest (TRG-RF): an interpretable random forest

Arpita Nath Boruah[1] · **Saroj Kumar Biswas**[1] · **Sivaji Bandyopadhyay**[1]

Received: 9 September 2021 / Accepted: 15 March 2022

Abstract
Ensemble learning method exhibits a high level of performance which is very much essential and useful in various domains. Random Forest (RF) is an ensemble learning technique which creates many trees on the subset of the data and combines the output of all the trees thereby it reducing overfitting problem in decision trees and also the variance thereby improving the performance. Despite of its high-performance, RF is black box in nature which hinders the transparency and interpretability of the predictive model. A transparent system with less decision rules makes a system efficient, user convincing and manageable to a greater extent in fields like medical, business, banking etc. The expression of the decision rules into flowchart like representation makes the system transparent, explicitly understandable and closely resemblance to human reasoning. Therefore, to overcome the drawback of black box nature of the RF and to make it an efficient interpretable decision-making system, this paper proposes a Transparent Rule Generator Random Forest (TRG-RF), which extracts the significant decision rules making the RF transparent, interpretable and comprehensible. The performance of the proposed TRG-RF is compared with the performances of a simple DT as well as simple RF, Support Vector Machine (SVM) and Naïve Bayes in terms of classification accuracy, precision, recall and F1 score measures. Moreover TRG-RF is compared with RuleFit and RF + DHC which are also rule-based methods. The performance of the proposed TRG-RF is validated with 12 well known UCI datasets and the experimental results showed that the proposed TRG-RF is more efficient transparent and interpretable decision-making system.

Keywords Random forest · Transparent system · Decision making · Decision tree

1 Introduction

Now a days with the development of modern technologies and efficient computing a large volume of data is collected and analysed by various organizations. The collected data are analysed to obtained useful information which can be successfully used in various fields for proper prediction, classification or deep analysis. For analyzing data to find some interesting facts, Data Mining (DM) techniques have emerged decades before. DM (Han and Kamber 2011) is defined as the process of discovering the hidden knowledge from large amount of data by analyzing them and is widely used in almost all the sectors like the medical, business, banking, commercial (Bhambri 2011; Koh and Tan 2011; Liao et al. 2012; Mukherjee et al. 2015; Tomar and Agarwal 2013) etc. There are various data mining tasks such as association rule mining, classification, clustering (Sing and Midha 2015; Mann and Kaur 2013; Shridhar and Parmar 2017) etc. to solve different kinds of problems. Among them classification is the most common and popular (Sharma and Shani 2011). Neural Network (NN), Bayesian Classification (BC) (Kaviani and Dhotre 2017) Support Vector Machine (SVM) (Pisner and Schnyer 2020), Decision Tree (DT) (Swain and Hauska 1977; Quinlan 1986; Safavian and Landgrebe 1991; Navada et al. 2011), Random Forest (RF) (Breiman 1996, 2001) etc. are some of the well-known tools for classification. Except DT all others are black box in nature and thereby they lack the essential characteristic of a transparent predictive model to explain the cause of prediction.

✉ Arpita Nath Boruah
arpita.boruah@hotmail.com

Saroj Kumar Biswas
bissarojkum@yahoo.com

Sivaji Bandyopadhyay
director@nits.ac.in

[1] Computer Science and Engineering Department, National Institute of Technology Silchar, 788010 Assam, India

Published online: 01 April 2022

DT is a transparent data mining technique that expresses the decision into a flow chart like representation which is explicitly understandable and closely resemblance to human reasoning. RF is an ensemble learning method for classification. Ensemble learning (Polikar 2006; Sagi and Rokach 2018; Dong et al. 2020; Lughofer et al. 2007) is primarily used to improve the classification, prediction performance of a model. Ensemble learning is used to obtain a better predictive performance on a predictive modeling problem than a single predictive model. RF is one of the most popular ensemble learning techniques which trains several decision trees in parallel with bootstrapping followed by aggregation, jointly referred as bagging. Because of its high-performance RF is widely used in healthcare, business and banking sector to build a good prediction model.

But unfortunately, RF is black box in nature which hinders the transparency and comprehensibility of the predictive model; therefore, RF is not used in places where transparency of the predictive model is required. A transparent interpretable system (Rudin 2014; Doshi-Velez and Kim 2017) is explicitly understandable and interpretable for decision making and, is useful for some important tasks. For example, in case of medical science to prevent a disease like diabetes a clear explanation of the causes of the disease is required to spread awareness among the common people and to take some precautionary measures to prevent the disease in advance. Similarly, fields like automatic finance credit risk evaluation a transparent decision-making system is required to justify the explanations for why applications are granted or rejected. Thus, a rich transparent and comprehensible system with less features and constrains is required to interpret a decision which makes a system efficient, user convincing and manageable to a great extent.

To overcome the drawback of black box nature of RF and to obtain a transparent and interpretable predictive system, this paper proposes a transparent RF called Transparent Rule Generator-RF (TRG-RF) to determine the significant decision rules and to produce a rich transparent system. In TRG-RF, the RF is trained with the specific number of DTs. The trees are ranked based on their classification capacities and are selected accordingly. Subsequently the rules from ranked trees are selected on the basis of a score function and are further processed to obtain the transparent decision rule set. The performance of the proposed TRG-RF is compared with simple DT and RF, SVM and Naïve Bayes. Further two rule-based methods, RuleFit (Friedman and Popescu 2008) and RF + DHC (Mashayekhi and Gras 2017) are used to compare performance and comprehensibility of the proposed TRG-RF. RuleFit is a predictive learning algorithm which is an ensemble model where the base learners are prediction rules that are obtained from DTs. RF + DHC uses RF to generate the rules. In RF + DHC method a Downward Hill Climbing (DHC) procedure is applied for a predefined number of iterations to reduce the number of rules and to find a reduced rule set. In each iteration to select a rule to be pruned in DHC process, a fitness proportionate selection procedure with a probability is used to remove a rule by generating a random number. Both the competitors of the proposed TRG-RF are rule based which extract large collection of rules from tree ensembles, RuleFit uses a boosted tree ensemble whereas RF + DHC uses RF. The rule selection is performed by a sparse linear aggregation for RuleFit and for RF + DHC a specific number of rules are selected. However, even though with high predictive capacities, these two methods produces large set of long and complex rules. Thus to produce a more transparent and interpretable decision making system, the proposed TRG-RF is used where to generate the rules the same RF is used but the selection of rules is done by using a different approach from the previously mentioned methods. Experimental results showed that the proposed TRG-RF produces more transparent and comprehensible decision rules than the RuleFit and RF + DHC methods and also has a better performance in terms of accuracy, precision, recall and F1 score measure than simple DT, RF, SVM, Naïve Bayes, RuleFit method and RF + DHC methods.

The paper is organized as follows: Sect. 2 shows the background of RF and its application, Sect. 3 describes the proposed algorithm in details, Sect. 4 discusses experimental results of the model and finally Sect. 5 draws conclusion.

2 Literature survey

Considering the recent trends of using ensemble learning, many interesting models have been proposed using RF. (Wu et al. 2019) have given a comparative study of four classification model which includes Random Forest (RF), Artificial Neural Network (ANN), Naive Bayes, Logistic Regression to classify fatty liver. The study concludes that RF has the higher classification performance in comparison to the other classifier. (Ganggayah et al. 2019) have used machine learning techniques to build models for detecting and visualising significant prognostic indicators of breast cancer survival rate and have found that RF has the accuracy and calibration measure in comparison to the other machine learning techniques. The study also has determined the important variables influencing survival rate of breast cancer patients that can be employed in clinical practice. (Shaikhina et al. 2019) have used DT and RF models to identified the key risk factors associated with antibody incompatible kidney

transplantation. (Zhou and Hooker 2016) have proposed a method to build a DT that approximates the performance of complex machine learning models. Thus, the single DT could be used to interpret and simplify the predicting pattern of RF and other models. They also introduced an improved splitting method designed to stabilize tree structure and also concluded that this particular tree structure is very relevant for medical questionnaires. (Mollas et al. 2020) have proposed a method where the classical unsupervised learning techniques and an enhanced similarity metric were used to stroll among transparent trees inside a forest to interpret the RF. They termed this process of RF interpretation as "LionForests". (Mashayekhi and Gras 2017) proposed three algorithms to extract important rules from ensemble decision tree. All the three algorithms viz. RF + DHC, RF + SGL and RF + MSGL use random forest to generate the decision rules. But the selection of the significant rule-set is different. RF + DHC uses downhill climbing algorithm for rule selection whereas RF + SGL uses sparse group lasso and RF + MSGL uses multi class sparse group lasso to extract the significant rules. (Wang et al. 2020) have proposed a rule extraction method named Improved Random Forest (RF)-based Rule Extraction (IRFRE) method to derive accurate and interpretable classification rules from a decision tree ensemble for breast cancer diagnosis. An improved multi-objective evolutionary algorithm (MOEA) is employed to seek for an optimal rule predictor where the constituent rule set is the best trade-off between accuracy and interpretability. The proposed method is evaluated on three breast cancer data sets, i.e., the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, Wisconsin Original Breast Cancer (WOBC) dataset, and Surveillance, Epidemiology and End Results (SEER) breast cancer dataset. (Angelov et al. 2007) have proposed an evolving structure of fuzzy rules for classification which is transparent and interpretable and can be applicable to both fully unsupervised and partially supervised learning. (Phung et al. 2015) have stated an algorithm to extract an interpretable classification rule set from a random forest for a multi-class data classification task. The algorithm is termed as ExtractingRuleRF, which follows a greedy approach and consisting of 2 phases, rule refinement and rule extraction. (Benard et al. 2021a, b) have proposed the Interpretable Random Forests for classification problems in which they extract the rules from a random forest, based on their probability of occurrence in a random tree, and then stop the growing of the forest when the rule selection is converged. In addition to this they also proposed the Interpretable Random Forests for regression problems in which they extract the rules from the tree based on their frequency of appearance. The most frequent rules, which represent robust and strong patterns in the data, are ultimately linearly combined to form predictions. (Albu et al. 2019) have presented several approaches by which technology can assist medical decision-making and ANNs applied to modeling, prediction and decision-making related to medical systems. (Angelov and Filev 2004) have discussed a flexible model in the form of a neural network (NN) with evolving structure termed as eNN. The proposed eNN differ from the others because of the gradually evolving structure rather than the fixed structure models and also their learning algorithm is incremental and combines unsupervised on-line recursive clustering and supervised recursive on-line output parameter estimation. Moreover it has been tested a real air-conditioning installation data. (Lughofer and Klement 2008) proposed *FLEXFIS* which is to train Takagi–Sugeno fuzzy models in offline as well as online mode. This evolving system includes both adaptation of linear parameters in fuzzy systems appearing in the rule consequents and also sample mode adaptation of premise parameters appearing in the fuzzy sets along with a rule learning strategy. (Pratama et al. 2018) have presented a novel evolving ensemble classifier, termed parsimonious ensemble (pENsemble). pENsemble features some unique characteristics, where an evolving classifier, namely pClass, is utilized as its local expert. pENsemble offers a parsimonious working principle, which is resulted from pruning activities of inactive classifiers. It is equipped with two ensemble pruning strategies, which assess the relevance and generalization power of a local expert.

3 Proposed methodology: transparent rule generation-RF (TRG-RF)

Transparent Rule Generator-RF (TRG-RF) has 3 steps: RF Classifier, Tree Processing, Rule Pruning. The systematic architecture of TRG-RF is given in Fig. 1.

3.1 RF classifier

RF is an ensemble learning in which multiple decision trees are built and merges them together to get a more accurate and stable prediction. TRG-RF uses a 500 number of estimators for training the RF classifier. Variable importance in a DT of the forest and similarly the variable importance of the forest, are the two important terms required in RF. Variable importance of a DT in a forest is defined as the importance by comparing the prediction accuracy of a tree before and after random combination of the particular variable on the

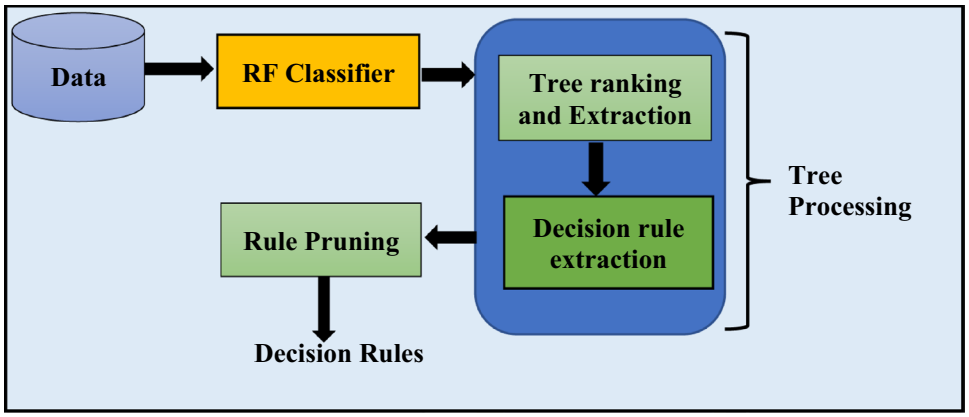

Fig. 1 TRG-RF architecture

out-of-bag (OOB) data. Similarly variable importance of the forest is defined as the average combination importance of the variable over all the trees in the forest. The variable importance of the forest is represented as Eq. 1

$$FIRF(f_i) = \frac{\sum_{t=1}^{n} FI^t(f_i)}{n} \tag{1}$$

where $FI^t(f_i)$ is the variable importance of variable f_i in a particular tree t of the RF, n is the total number of trees in the forest and $FIRF(f_i)$ gives the variable importance of f_i in the forest. The higher the value of $FIRF(f_i)$, more important is the variable in the forest. Thus, RF classifier is basically a set of DTs from a randomly selected subset of the training set and finally based on the votes from different DTs RF classifier makes the final prediction.

3.2 Tree processing

Tree Processing step is used to process the trees of the RF to obtain the decision rules. It consists of 2 sub steps: i. Tree ranking and Extraction ii. Decision rule extraction. In Tree ranking and Extraction sub step the trees of the RF are ranked based on their accuracy performance and the tree with the highest performance is extracted for further processing. The tree thus extracted is termed as ***Ranked Tree (RT)*** and thereby RT is used in the Decision rule extraction subset to generate the rules for decision making (Fig. 2).

3.3 Rule pruning

All the rules generated by RT from the previous Tree Processing step are not necessary and efficient for decision making. Thus, in the Rule Pruning step all the unnecessary and irrelevant rules are removed by the proposed Rule Pruning algorithm named **Sequential Rule Pruning Algorithm (SRPA)** to obtain an efficient set of decision rules and hence it reduces the complexity of the prediction and produces an interpretable RF. The proposed SRPA consists of 3 steps: *Score assignment, Initial Transparent Rule Set and Final Transparent Rule set*. According to Eq. 2, in the Score assignment step of SRPA, a score is assigned to all the decision rules generated by the RT of each fold.

$$SOR = \left[\left(\frac{CP - IP}{CP + IP}\right) + \left(\frac{CP}{IP + 1}\right) - \frac{IP}{CP}\right] + \frac{CP}{RL} \tag{2}$$

where, SOR = Significance of Rule, CP = Total number of correctly classified patterns by a rule, IP = Total number of incorrectly classified patterns by a rule and RL = Rule Length.

In the Initial Transparent Rule Set the most significant rule of each fold is selected based on its SOR score to form a rule set called the **Initial Rule Set**. Thus the Initial Transparent Rule Set consists of the highest SOR scored rules of each fold.

This rule selection procedure of the Initial Transparent Rule Set step for each fold is given below in Algorithm 1.

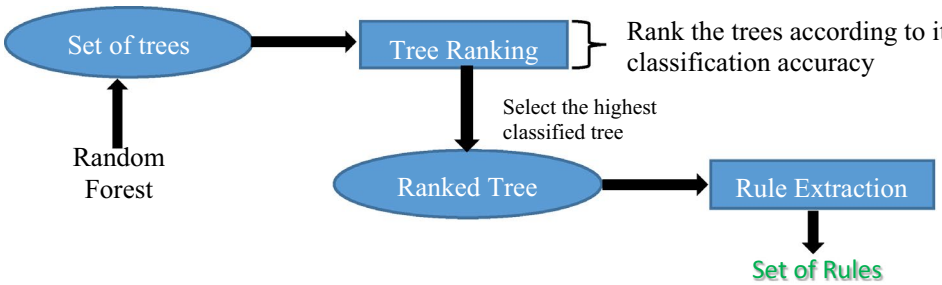

Fig. 2 Flowchart for Tree processing

Algorithm 1:
Initial Transparent Rule Set

Input: Rule-set$_{initial}$
Output: Rule$_{single}$

Notations:

S ⟶ *Rules of the fold under consideration*
n ⟶ *Number of rules in S*
i ⟶ *a variable (i=1 to n)*
R_i → *i^{th} rule in S*
A ⟶ *array of the SOR values of the rules of S*
A_i → *SOR of the rule* R_i
Max_value → Maximum value of the array A
Max_index=Array of Rules with the highest SOR in S (initialize to NULL)

Step 1: *Initialize S and n.*
Step 2: *for i=1 to n*

$$A_i = \left[\left(\frac{CP_i - IP_i}{CP_i + IP_i}\right) + \left(\frac{CP_i}{IP_i + 1}\right) - \frac{IP_i}{CP_i}\right] + \frac{CP_i}{RL_i}$$

Step 3: *Max_value =* A_1
Step 4*: for i=2 to n*
If A_i*>=Max_value*
Max_value= A_i
Step 5*: for i=1 to n*
If A_i*= Max_value*
Max_index. Append(i)
Step 6: *Return(Max_index)*

All the rules of the Initial Rule Set are not necessarily useful or important in decision making. To enhance the transparency further the Initial Rule Set is pruned. In the Final Transparent Rule Set step, the redundant and unnecessary rules of the Initial Rule Set are removed sequentially by Sequential Hill Climbing Procedure to enhance the transparency as well as the accuracy. The rules thus generated are called as the **Final Rule Set**. In the Sequential Hill Climbing Procedure for Final Rule Set, each rule of the Initial Rule set is sequentially pruned. Each time a rule is removed, the accuracy is calculated for the rule set and compared with the original set. If the accuracy increases or does not change then that pruned rule is permanently removed. This sequential process continues until no rule to be removed.

The algorithmic form of the Final Transparent Rule Set step is given Algorithm 2.

Algorithm 2:
Sequential Hill Climbing procedure for Final Rule Set

Input*: Initial Rule Set*
Output: *Final Rule Set*

Notations:
$P_{initial}$ → Initial Rule Set
T_{no} → Total number of rules in $P_{initial}$
Aac→ classification accuracy of $P_{initial}$
Flag → 0
HC→ Hill Climbing procedure
Rule-set$_{new}$ → initialize to be NULL
Aac-ruleset → classification accuracy of the Rule-set$_{new}$
R_{final} → Final Rule Set

Procedure:
Step 1: Calculate Aac
Step 2: while Flag! =1
Step 2.1: for i=1 to T_{no}
Rule-set$_{new}$= HC ($P_{initial}$, i)
Compute Aac-ruleset
If Aac-ruleset>= Aac
Aac=Aac-ruleset
$P_{initial}$=Rule-set$_{new}$
Break
Step 2.2: if i== T_{no}
Flag=1
Step 3: Return **$P_{initial}$**

Step1 of the algorithm 2 calculates the accuracy of the original rule set. In the Step 2, while the flag value is 0 proceed to step 2.1. In step 2.1, "i" is initialized to 1 and continue the loop to total number of rules in the rule set. The Hill Climbing algorithm is called to generate the new rule set, **Rule-set$_{new}$** and thereby the accuracy is calculated. Both the accuracy of the original rule set and the new rule set are compared. If the accuracy of the new rule set is more or is same that of the original accuracy value, then the original accuracy value is modified to the new accuracy and the original rule set is changed to new rule set and then break the for loop. In the step 2.2 "i" is checked with the total number of rules and if yes then change the flag value to 1 and come out of the while loop else continue the same procedure till flag is equal to 1. Finally step 3 returns the final rule set.

The Hill Climbing (HC) procedure used in Algorithm 2 is a sequential process by which a rule is selected sequentially for pruning. Its algorithmic procedure is given in Algorithm 3.

Algorithm 3: *Hill Climbing (HC) Procedure*
Input: $P_{initial}$, i **Output:** $Rule\text{-}set_{new}$
Notations: i^{th} rule: i^{th} rule of $P_{initial}$
Step 1: Rule-set$_{new}$= remove the i^{th} rule from the $P_{initial}$ Step 2: Return Rule-set$_{new}$

In the algorithm 3, the ith rule passed from the algorithm 2 is removed in step 1 to generate the new rule set, **Rule-set$_{new}$**. And Step 2 returns the Rule-set$_{new}$ to be processed in the algorithm 2.

Table 1 Characteristics of the datasets

Dataset	Number of instances	Number of features	Number of classes
Heart	270	13	2
Liver	345	6	2
Australian credit	690	14	2
Sonar	208	60	2
Ionosphere	351	33	2
Parkinson's disease	195	22	2
Swine Flu	250	12	2
Blood Transfusion Service Center	748	5	2
South German Credit	1000	21	2
Breast cancer Coimbra	116	10	2
Diabetic Retinopathy Debrecen	1151	20	2
Heart Failure Clinical records	299	13	2

4 Experimental analysis

The experimentation of the proposed model TRG-RF is performed in PYTHON 2.8 version in Windows environment. A total of 12 datasets taken from UCI (University of California, Irvine) machine learning repository are considered for classification purpose to validate the proposed model. All the experiments are done with tenfold cross validation. The proposed TRG-RF uses a total of 500 estimators and all other parameters are set as default. A brief description of the datasets is given in the Table 1 which includes the total number of samples, attributes and the number of classes of occurrences. To validate the results obtained by the proposed method TRG-RF, it is compared with simple DT as well as simple RF, SVM, Naïve Bayes, RuleFit and RF + DHC methods.

4.1 Evaluation functions

The eminent metrics accuracy, comprehensibility, precision, recall and F1 score are used to evaluate the proposed TRG-RF. Table 2 shows the confusion matrix which is used to calculate these measures. Table 3 shows the performance measures defined from the confusion matrix.

Accuracy is defined as the number of correct predictions by the classifier to the total number of prediction. In other words accuracy depicts often we can expect our machine learning model will correctly predict an outcome out of the total number of times it made predictions, whereas recall is out of all positive class how much are correctly predicted and precision is out of all positive class that has been predicted how many are actually positive. Both the recall and precision scores are useful measures of the success of prediction when the classes are very imbalanced. F1-score is harmonic mean of precision and recall score and is used as

Table 2 Confusion matrix

		Predicted	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	TP

Table 3 Performance measures

Performance measures	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
f-measure	$\frac{2*precision*recall}{precision+recall}$

Table 4 Accuracy comparison of the proposed TRG-RF with simple DT and RF, SVM and Naïve Bayes

Dataset	Simple DT	SVM	Naive Bayes	Simple RF	TRG-RF
Heart	75.93	82.96	83.07	84.44	92.96
Liver	62.27	65.81	55.89	70.44	73.59
Australian credit	56.23	84.20	78.98	86.06	92.17
Sonar	67.88	53.45	58.83	82.16	90.28
Ionosphere	86.64	86.92	86.31	93.45	97.72
Parkinson's disease	78.82	81.87	67.47	86.74	94.89
Swine Flu	90.00	92.4	93.99	93.6	94.01
Blood Transfusion Service Center	72.22	74.64	75.17	77.84	80.87
South German Credit	63.49	69.19	68.89	69.1	76.51
Breast cancer Coimbra	55.38	53.33	54.77	36.67	56.33
Diabetic Retinopathy Debrecen	63.24	69.15	59.51	68.72	74.08
Heart Failure Clinical records	78.61	79.33	80.00	85.67	90.64
Average	70.89	74.44	71.91	77.91	84.50

a metrics in the scenarios where choosing either of precision or recall score can result in compromise in terms of model giving high false positives and false negatives respectively. Thus these performance measures are considered for evaluating the performances.

4.2 Results and analysis

The TRG-RF is firstly compared with Simple DT, simple RF, SVM and Naïve Bayes. Simple DT is already a transparent decision-making system which produces the decision rules that can be represented as flow chart. Whereas simple RF and SVM are both classification tools which are black box in nature. Naïve Bayes is a probabilistic classifier based on applying Bayes' theorem with strong independence assumptions between the features. Finally, the proposed TRG-RF is compared with RuleFit and RF + DHC which are rule-based method similar to the proposed method. Table 4 shows the accuracy comparison of TRG-RF, simple DT and RF along with SVM and Naïve Bayes.

From Table 4 it is observed that the proposed TRG-RF on an average has a better accuracy in comparison to the simple DT, simple RF, SVM and Naïve Bayes. A simple RF is a strong modelling technique creating diversity among the trees and much more robust than a single DT. RF aggregates many decision trees to limit overfitting as well as error due to bias and therefore yields useful results. Considering the advantage of RF, the proposed model TRG-RF enhances the performances of the RF by selecting the highest ranked tree among the forest and by using SRPA the rules generated by the RT are selected on the basis of the SOR value and further processed to enhance the predictive capacity and outperformed DT, RF, SVM and Naïve Bayes. Figure 3 shows the graphical representation of accuracy comparison of all the classification methods taken under consideration for better visualisation and understanding.

Along with the accuracy, precision, recall and F1 score are also used for the comparison. Tables 5, 6 and 7 show the precision, recall and F1 score comparison of TRG-RF with simple DT and RF, SVM, Naïve Bayes.

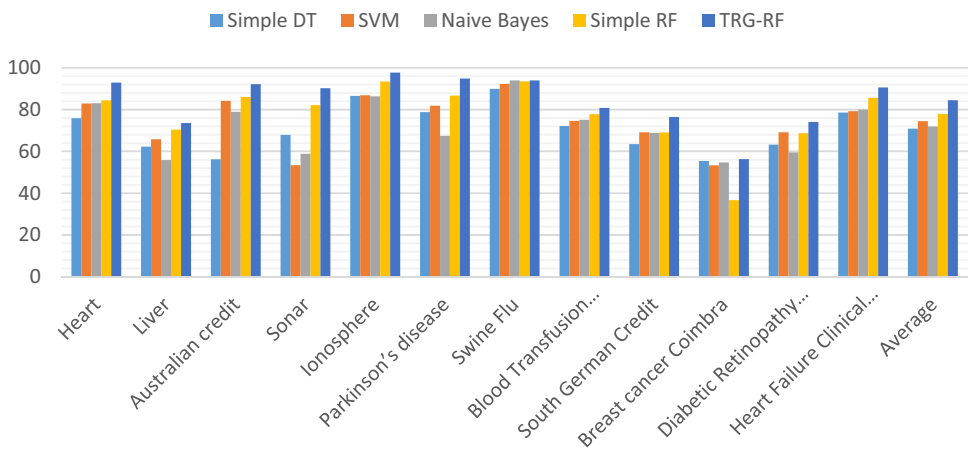

Fig. 3 Graphical representation of accuracy comparison

Table 5 Precision comparison of TRG-RF with simple DT and RF, SVM, Naïve Bayes

Dataset	Simple DT	SVM	Naive Bayes	Simple RF	TRG-RF
Heart	0.76	0.83	0.84	0.85	0.89
Liver	0.57	0.66	0.55	0.68	0.72
Australian credit	0.51	0.84	0.81	0.86	0.88
Sonar	0.52	0.49	0.47	0.73	0.80
Ionosphere	0.81	0.79	0.81	0.89	0.92
Parkinson's disease	0.65	0.74	0.61	0.71	0.79
Swine Flu	0.81	0.93	0.95	0.94	0.96
Blood Transfusion Service Center	0.35	0.39	0.49	0.50	0.59
South German Credit	0.50	0.56	0.55	0.57	0.58
Breast cancer Coimbra	0.50	0.53	0.5	0.58	0.59
Diabetic Retinopathy Debrecen	0.60	0.76	0.68	0.69	0.77
Heart Failure Clinical records	0.58	0.61	0.62	0.66	0.67
Average	0.60	0.68	0.66	0.72	0.76

Table 6 Recall comparison of TRG-RF with simple DT and RF, SVM, Naïve Bayes

Dataset	Simple DT	SVM	Naive Bayes	Simple RF	TRG-RF
Heart	0.76	0.84	0.84	0.81	0.86
Liver	0.60	0.69	0.57	0.71	0.79
Australian credit	0.51	0.84	0.77	0.86	0.90
Sonar	0.37	0.28	0.33	0.63	0.71
Ionosphere	0.77	0.77	0.78	0.88	0.91
Parkinson's disease	0.62	0.64	0.52	0.67	0.70
Swine Flu	0.82	0.92	0.95	0.94	0.95
Blood Transfusion Service Center	0.40	0.49	0.52	0.54	0.55
South German Credit	0.49	0.53	0.62	0.51	0.63
Breast cancer Coimbra	0.30	0.31	0.30	0.38	0.40
Diabetic Retinopathy Debrecen	0.57	0.75	0.62	0.69	0.76
Heart Failure Clinical records	0.50	0.59	0.65	0.64	0.66
Average	0.56	0.64	0.62	0.69	0.74

Table 7 F1 score comparison of TRG-RF with simple DT and RF, SVM, Naïve Bayes

Dataset	Simple DT	SVM	Naive Bayes	Simple RF	TRG-RF
Heart	0.76	0.83	0.84	0.83	0.87
Liver	0.58	0.67	0.56	0.69	0.75
Australian credit	0.51	0.84	0.79	0.86	0.89
Sonar	0.43	0.36	0.39	0.68	0.75
Ionosphere	0.78	0.78	0.79	0.88	0.91
Parkinson's disease	0.63	0.69	0.56	0.69	0.74
Swine Flu	0.81	0.91	0.94	0.94	0.95
Blood Transfusion Service Center	0.37	0.44	0.50	0.52	0.57
South German Credit	0.49	0.54	0.58	0.54	0.60
Breast cancer Coimbra	0.37	0.39	0.38	0.46	0.48
Diabetic Retinopathy Debrecen	0.58	0.75	0.65	0.69	0.76
Heart Failure Clinical records	0.58	0.59	0.63	0.65	0.66
Average	0.57	0.65	0.64	0.70	0.74

From Table 5, it is observed that along with accuracy TRG-RF has a better performance also in terms of precision in comparison to simple DT and RF, SVM and Naïve Bayes. Better performance in terms of precision illustrate that the proposed TRG-RF has an almost accurate classification as the number of false positives are comparatively less. From the application point of view high value of precision is very much essential for business purpose. The graphical representation of precision as depicted by Fig. 4 give the comparison visualisation of TRG-RF with simple DT, RF, SVM and Naïve Bayes for better interpretation.

From Table 6, it is observed that along with accuracy TRG-RF has a better recall performance measure in comparison to simple DT and RF, SVM and Naïve Bayes. Better

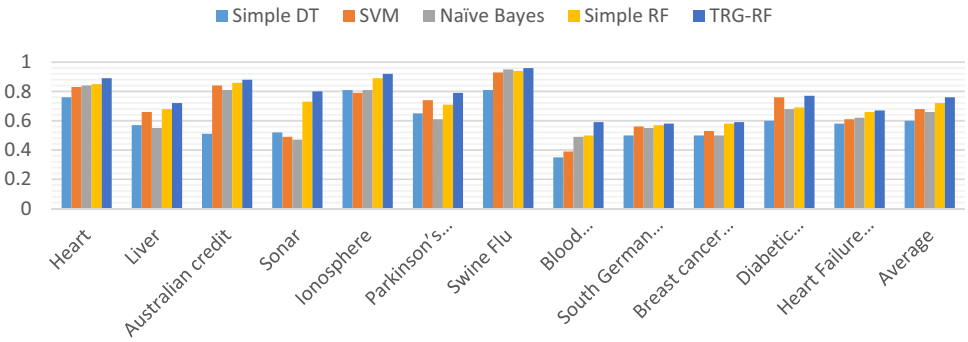

Fig. 4 Graphical representation of precision comparison

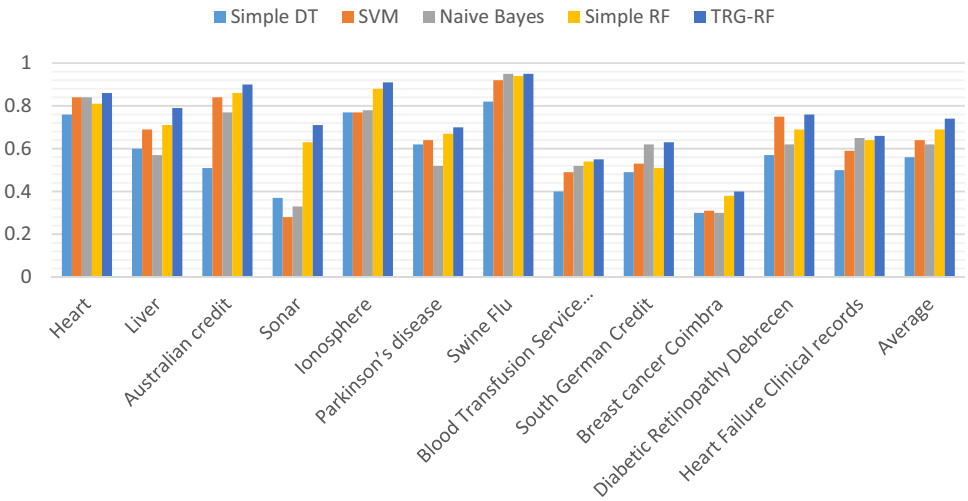

Fig. 5 Graphical representation of recall comparison

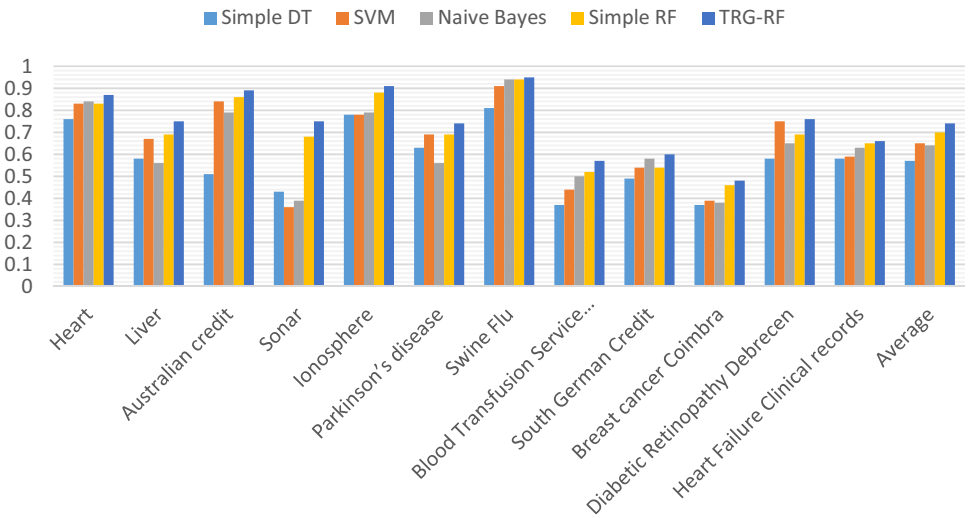

Fig. 6 Graphical representation of F1 comparison

Table 8 Accuracy comparison of RuleFit and RF + DHC with TRG-RF

Dataset	RuleFit	RF + DHC	TRG-RF
Heart	83.33	82.59	92.96
Liver	63.77	66.43	73.59
Australian credit	85.36	87.43	92.17
Sonar	57.40	61.78	90.28
Ionosphere	91.76	87.49	97.72
Parkinson's disease	79.82	79.28	94.89
Swine Flu	92.0	93.20	94.01
Blood Transfusion Service Center	77.31	79.56	80.87
South German Credit	68.30	70.14	76.51
Breast cancer Coimbra	50.23	55.89	56.33
Diabetic Retinopathy Debrecen	73.98	73.52	74.08
Heart Failure Clinical records	83.65	89.06	90.64
Average	76.91	77.5	84.50

performance in terms of recall illustrate that the proposed TRG-RF has an almost accurate classification as the number of false negatives are comparatively less. From the application point of view high value of recall is very much essential for healthcare and medical purpose. The graphical representation of recall as depicted by Fig. 5 give the comparison visualisation of TRG-RF with simple DT, RF, SVM and Naïve Bayes for better interpretation.

It is observed from Table 7 that along with accuracy TRG-RF has a better performance also in terms of f1 score in comparison to simple DT and RF, SVM and Naïve Bayes. The F1 score measure is also better for TRG-RF as the precision and recall measures are almost equal and high in comparison to simple DT, RF, SVM and Naïve Bayes. Figure 6 depicts the graphical representation of F1 score for comparison visualisation and understanding of TRG-RF with simple DT, RF, SVM and Naïve Bayes.

For the evaluation propose, RuleFit and RF + DHC are also applied on the selected datasets. All the performance measures like accuracy, precision, recall, F1 score measures and including the comprehensibility are used to compare TRG-RF with RuleFit and RF + DHC. Table 8 shows the accuracy measure whereas Table 9 shows the precision, recall, F1 score measures.

From Table 8 it is observed that the proposed TRG-RF gives better accuracy for all the datasets in comparison to RuleFit and RF + DHC. In RuleFit only the selected or the extracted rules are used to build the final model and hence the performance decreases and in RF + DHC rules are refined by DHC where the rule is removed by a fitness proportionate selection procedure and by repeating the process for a predefined number of iterations. Whereas in the proposed TRG-RF uses a number of DT to train the model and thereby select the best tree of each fold and merges them together to get a more accurate and stable prediction.

Figure 7 shows the graphical representation of accuracy comparison between RuleFit, RF + DHC and TRG-RF for better visualisation and understanding.

From Table 9, it is observed that TRG-RF has a better performance in terms of precision, recall and F1 score in comparison to RuleFit and RF + DHC method. This is due to an almost accurate classification as the number of false positive and false negative is comparatively less which is not so for the RuleFit and RF + DHC method. Figures 8, 9 and 10 depicts the graphical representation of precision, recall and F1 score comparison of TRG-RF with RuleFit

Table 9 Precision, Recall and F1 score comparison of TRG-RF with RuleFit and RF + DHC methods

Dataset	Precision			Recall			F1 measure		
	RuleFit	RF + DHC	TRG-RF	RuleFit	RF + DHC	TRG-RF	RuleFit	RF + DHC	TRG-RF
Heart	0.84	0.81	0.89	0.83	0.8	0.86	0.83	0.80	0.87
Liver	0.62	0.67	0.72	0.64	0.68	0.79	0.63	0.67	0.75
Australian credit	0.85	0.86	0.88	0.85	0.87	0.90	0.85	0.86	0.89
Sonar	0.5	0.56	0.80	0.31	0.54	0.71	0.38	0.55	0.75
Ionosphere	0.89	0.85	0.92	0.77	0.78	0.91	0.83	0.81	0.91
Parkinson's disease	0.65	0.63	0.79	0.66	0.64	0.70	0.65	0.63	0.74
Swine Flu	0.89	0.9	0.96	0.88	0.91	0.95	0.88	0.90	0.95
Blood Transfusion Service Center	0.54	0.57	0.59	0.49	0.50	0.55	0.51	0.53	0.57
South German Credit	0.44	0.49	0.58	0.56	0.58	0.63	0.49	0.53	0.60
Breast cancer Coimbra	0.42	0.45	0.59	0.33	0.35	0.40	0.3	0.39	0.48
Diabetic Retinopathy Debrecen	0.71	0.7	0.77	0.72	0.71	0.76	0.71	0.70	0.76
Heart Failure Clinical records	0.59	0.62	0.67	0.61	0.62	0.66	0.60	0.62	0.66
Average	0.66	0.68	0.76	0.64	0.67	0.74	0.64	0.67	0.74

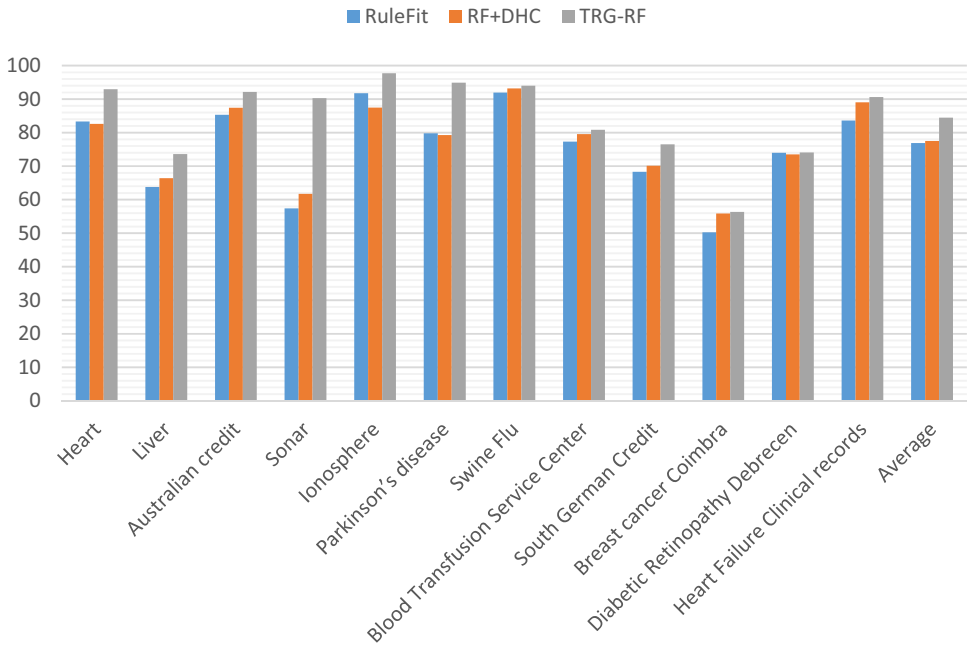

Fig. 7 Graphical representation of Accuracy comparison of RuleFit, RF + DHC with TRG-RF

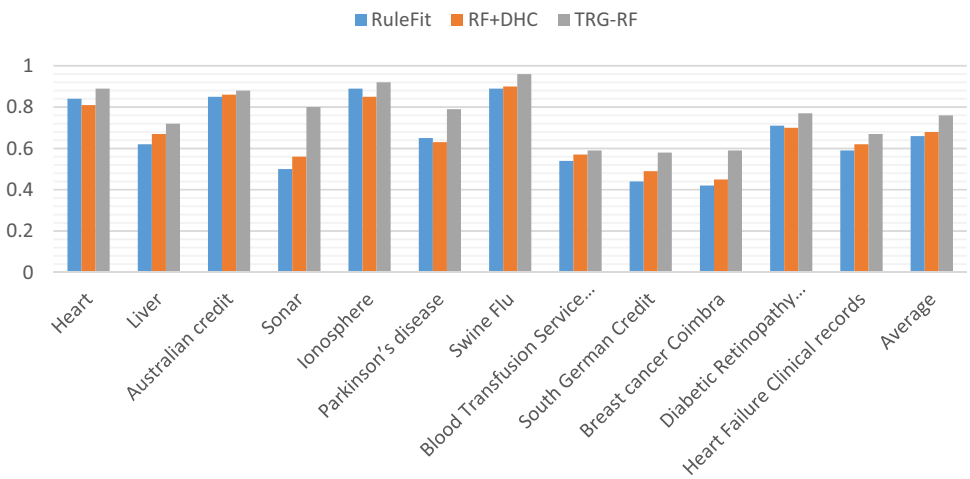

Fig. 8 Graphical representation of precision comparison

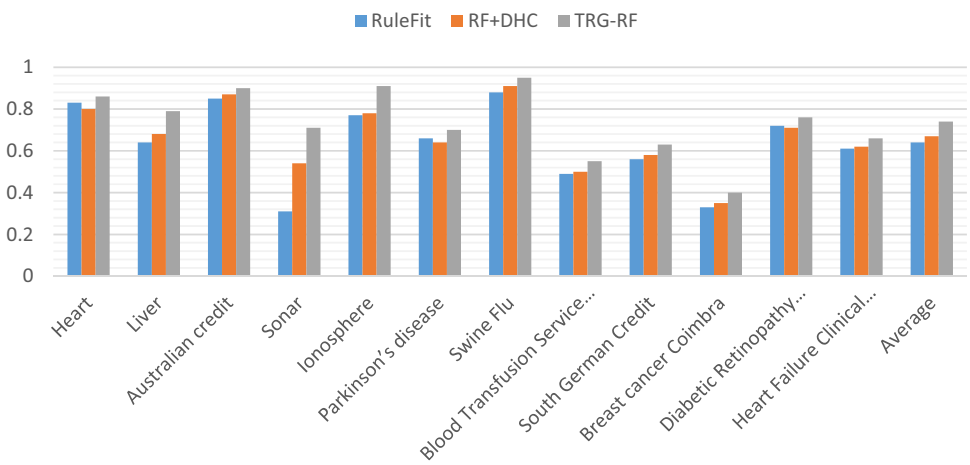

Fig. 9 Graphical representation of recall comparison

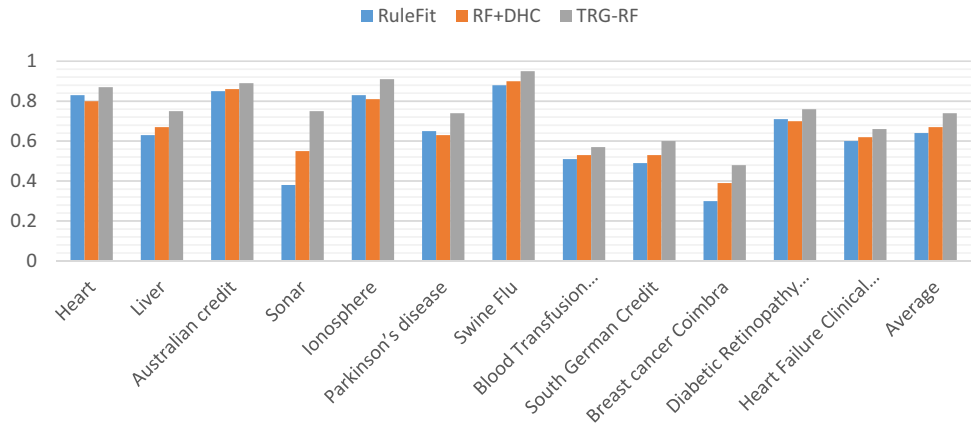

Fig. 10 Graphical representation of F1 score comparison

Table 10 Comprehensibility comparison of RuleFit and RF + DHC with TRG-RF

Dataset	Global comprehensibility			Local comprehensibility		
	RuleFit	RF + DHC	TRG-RF	RuleFit	RF + DHC	TRG-RF
Heart	14	6	5	23	25	20
Liver	15	6	3	26	20	15
Australian credit	18	10	5	35	30	21
Sonar	53	7	4	121	28	15
Ionosphere	21	8	5	35	30	9
Parkinson's disease	12	4	3	29	14	10
Swine Flu	15	8	6	23	27	20
Blood Transfusion Service Center	10	7	5	25	25	13
South German Credit	24	9	7	22	32	20
Breast cancer Coimbra	12	8	6	20	26	18
Diabetic Retinopathy Debrecen	30	13	8	69	35	21
Heart Failure Clinical records	13	7	5	25	34	11
Average	19.75	7.75	5.17	37.75	27.17	16.08

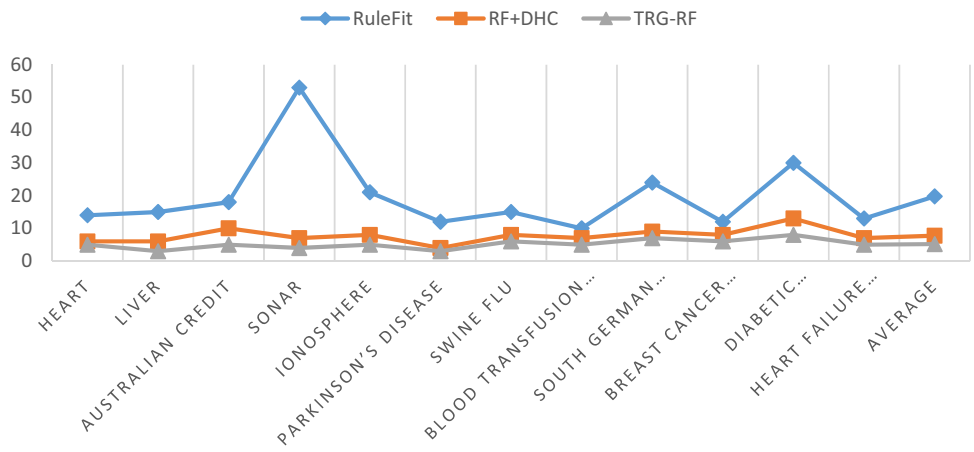

Fig. 11 Statistical representation of Global comprehensibility of TRG-RF, RuleFit and RF + DHC

and RF + DHC for a clear visualisation of difference in performance.

Comprehensibility is another important parameter in performance comparison. Comprehensibility is represented in terms of global and local. Global comprehensibility refers to the total number of decision rules of the predicting model whereas local comprehensibility is the total number of attributes/ features in the decision rules. Table 10 shows both the global and local comprehensibility comparison of TRG-RF with RuleFit and RF + DHC method.

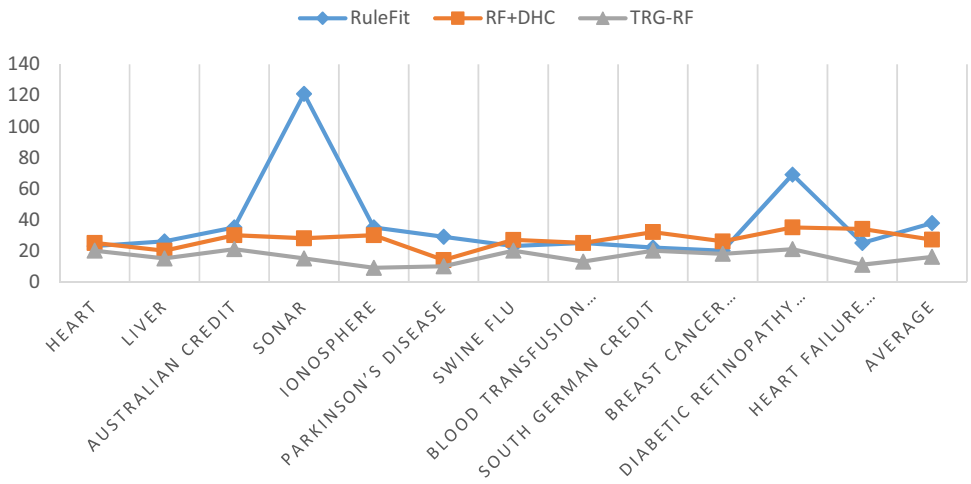

Fig. 12 Statistical representation of Local comprehensibility of TRG-RF, RuleFit and RF + DHC methods

From Table 10, it is observed that there is a clear distinction in both the global and local comprehensibility of TRG-RF with respect to RuleFit and RF + DHC methods. In case of TRG-RF the best rules selected on the basis of the score function from the ranked trees is pruned using the proposed SRPA rule pruning procedure to remove the redundant and unnecessary rules and thereby enhance the performance of the proposed method and to obtain a better interpretable system with a less number of decision rules.

Figures 11 and 12 depict the statistical representation of global and local comprehensibility for better illustration of the difference between TRG-RF, RuleFit and RF + DHC in terms of comprehensibility.

4.3 Discussion

A good predictive model needs to have a higher accuracy, precision, recall value etc. But along with these, transparency is also an important factor for any predictive model for its user understandability and interpretability. Thus this paper proposes a Transparent and interpretable RF named TRG-RF which converts RF from black box system to white-box system by extracting the generate rules to human understandable form. The proposed TRG-RF differs from the previous established models in terms of selecting and extracting the rules from the tress. In our model, the tree with the highest classification accuracy is selected and its rules are generated. The rules of the individual trees are given a score on the basis of the proposed score function, **SOR** and the rule with the highest SOR value is selected as the best rule for that particular tree and finally the best rules from all the trees are merged to form the initial rule set. Finally the initial rule set is processed and pruned using the proposed **SRPA** algorithm to get the interpretable and understandable decision rules. The proposed model TRG-RF firstly overcome the overfitting problem using the RF for generating the decision rules, then by using the Tree processing and Rule pruning procedure, a transparent set of decision rules is obtained. From the experimentation it is clearly observed that for all the considered dataset the proposed TRG-RF gives a better performance in comparison to simple DT and RF, SVM, Naïve Bayes, RuleFit and RF + DHC methods. Most importantly, in terms of transparency the proposed TRG-RF gives a better global and local comprehensibility. Thus, it can be concluded that the proposed TRG-RF is very much potential as a transparent system. For the experimentation, different range of features and sample sizes are considered and it is observed that if the number of features is less than 20 and greater than 5 then the number of rules in the final rule set is around 50% of the number of features. Additionally, around 20–40% of the rules are selected out of the average number of generated rules.

5 Conclusion

This paper proposes TRG-RF which converts RF from black box system to white-box system by extracting the generate rules to human understandable form. The TRG-RF extracts the highest ranked tree from the forest and processes the rules using the rule pruning procedure, SRPA to obtain a highly comprehensible interpretable decision- making system. The proposed TRG-RF employs a sequential hill climbing technique for pruning the decision rules selected on the basis of the SOR value to improve the classification performance as well as the comprehensibility of the generated rules.

The proposed TRG-RF is compared with a simple DT and RF as well as SVM, Naïve Bayes in terms of their classification accuracy, precision, recall and F1 score. The proposed

TRG-RF is also compared with the two-existing models, RuleFit and RF + DHC methods. In addition to accuracy, precision, recall and F1 score both the global and local comprehensibility are also used to compare TRG-RF with RuleFit and RF + DHC methods. For evaluating the performance of the proposed TRG-RF, experiments are performed considering 12 real-life datasets from the UCI repository. It is observed from the experimental results that the efficiency of the proposed TRG-RF model in terms of accuracy, precision, recall, F1 measure along with comprehensibility is very much better. Consequently, it can be concluded that the proposed TRG-RF model can be used in many applications like medical diagnosis, banking etc. for proper explanation and interpretation of the decision-making system. This process can be further extended in future to extract a greater number of tress from the forest based on some other heuristic function and also by using other threshold value. For processing the rules of the trees some other rule pruning procedure can be applied in future.

Declarations

Conflict of interest There are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

Albu A, Precup R, Teban T (2019) Results and challenges of artificial neural networks used for decision-making and control in medical applications. Facta Universitatis Series Mech Eng. https://doi.org/10.22190/FUME190327035A

Angelov P, Filev D (2004) Flexible models with evolving structure. Int J Intell Syst 19(4):327–340

Angelov P, Zhou X, Klawonn F (2007) Evolving Fuzzy rule-based classifiers. In: 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing, pp 220–225, https://doi.org/10.1109/CIISP.2007.369172

Benard C, Biau G, Da Veiga S, Scornet E (2021a) Sirus: Stable and interpretable rule set for classification. Elect J Stat 15:427–505

Benard C, Biau G, da Veiga S, Scornet E (2021b) Interpretable random forests via rule extraction. In: Proceedings of the 24th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, pp 937–945

Bhambri V (2011) Application of data mining in banking sector. Int J Comput Sci Technol 2(2):199–202

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

Breiman L (2001) Random forest. Mach Learn 45:5–32

Dong X, Yu Z, Cao W, Shi Y, Ma Q (2020) A survey on ensemble learning. Front Comput Sci 14(2):241–258

Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608

Friedman JH, Popescu BE (2008) Predictive learning via rule ensembles. Ann Appl Stat 3(2):916–954

Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK (2019) Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Inform Decis Mak 19:48

Han J, Kamber M, Pei J (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers, San Francisco

Kaviani P, Dhotre S (2017) Short survey on Naive Bayes algorithm. Int J Adv Res Comput Sci Manag 4:607–611

Koh HC, Tan G (2011) Data mining applications in healthcare. J Healthcare Info Manage 19(2):64–72

Liao S, Chu P, Hsiao P (2012) Data mining techniques and applications – a decade review from 2000 to 2011. Expert Syst Appl 39(12):11303–11311

Lughofer E, Klement EP (2008) FLEXFIS: a variant for incremental learning of Takagi-Sugeno fuzzy systems. In: The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ '05., pp 915–920. https://doi.org/10.1109/FUZZY.2005.1452516

Lughofer E, Angelov P, Zhou X (2007) Evolving single-and multi-model fuzzy classifiers with FLEXFIS-Class. In: 2007 IEEE International Fuzzy Systems Conference, pp 1–6. https://doi.org/10.1109/FUZZY.2007.4295393

Mann AK, Kaur N (2013) Survey paper on clustering techniques. Int J Sci Eng Technol Res IJSETR 2:803–806

Mashayekhi M, Gras R (2017) Rule extraction from decision trees ensembles: new algorithms based on heuristic search and sparse group Lasso methods. Int J Info Technol Decis Making (IJITDM) 16(06):1707–1727

Mollas I, Bassiliades N, Vlahavas I, Tsoumakas G (2020). LionForests: local interpretation of random forests. arXiv:1911.08780

Mukherjee S, Shaw R, Haldar N, Changdar S (2015) A survey of data mining applications and techniques. Int J Comput Sci Info Technol 6(5):4663–4666

Navada A, Ansari AN, Patil S, Sonkamble BA (2011) Overview of use of decision tree algorithms in machine learning. IEEE Control Syst Graduate Res Colloq (ICSGRC). https://doi.org/10.1109/ICSGRC.2011.5991826

Phung LTK, Chau VTN, Phung NH (2015) ExtractingRuleRF in educational data classification: from a random forest to interpretable refined rules. In: 2015 International Conference on Advanced Computing and Applications, pp 20–27. https://doi.org/10.1109/ACOMP.2015.13

Pisner DA, Schnyer DM (2020) Support Vector Machine. Machine Learning Methods and Applications to Brain Disorders, pp 101–121

Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits Syst Mag 6(3):21–45

Pratama M, Pedrycz W, Lughofer E (2018) Evolving ensemble fuzzy classifier. IEEE Trans Fuzzy Syst 26(5):2552–2567

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106

Rudin C (2014) Algorithms for interpretable machine learning. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '14). Association for computing machinery, New York, NY, USA, 1519. https://doi.org/10.1145/2623330.2630823

Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 21(3):660–674

Sagi O, Rokach L (2018) Ensemble learning: a survey. Wiley Interdiscip Rev Data Mining Knowl Dis. https://doi.org/10.1002/widm.1249

Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova N (2019) Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. Biomed Signal Process Control 52:456–462

Sharma AK, Shani S (2011) A comparative study of classification algorithms for spam email data analysis. Int J Comput Sci Eng 3:1890–1895

Shridhar M, Parmar M (2017) Survey on association rule mining and its approaches. Int J Comput Sci Eng 5:129–135

Sing V, Midha N (2015) A survey on classification techniques in data mining. Int J Comput Sci Manag Stud 16:9–12

Swain PH, Hauska H (1977) The decision tree classifier: design and potential. IEEE Trans Geosci Electron 15(3):142–147

Tomar D, Agarwal S (2013) A survey on data mining approaches for healthcare. Int J Bio-Sci BioTechnol 5(5):241–266

Wang S, Wang Y, Wang D, Yin Y, Wang Y (2020) An improved random forest-based rule extraction method for breast cancer diagnosis. Appl Soft Comput J 86:105941

Wu C, Yeh W, Hsu W, Islam MM, Nguyen PA, Poly TN, Wang Y, Yang H, Li Y (2019) Prediction of fatty liver disease using machine learning algorithms. Comput Methods Programs Biomed 170:23–29

Zhou Y, Hooker G (2016) Interpreting models via single tree approximation. arXiv preprint arXiv:1610.09036