



# A novel technique: ensemble hybrid 1NN model using stacking approach

Preeti Nair<sup>1,3</sup> · Navneet Khatri<sup>2</sup> · Indu Kashyap<sup>1</sup>

Received: 10 September 2017 / Accepted: 8 February 2018 / Published online: 20 February 2018  
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2018

**Abstract** This paper proposes a novel hybrid classification model which has enhanced the performance of the standard kNN ( $k = 1$ ) classification model significantly. In this study by the means of ensemble stacking approach kNN classification model and rotation forest classification model are hybridized as base classifiers and simple logistic classifier as the meta classification model. The performance of this proposed hybrid model was assessed using Accuracy and FMeasure. The model was compared with standard kNN and nine other classification models. The results showed that the proposed hybrid model has notably high performance than the other models.

**Keywords** k nearest neighbor · Rotation forest · Simple logistic · Hybrid classification model

## 1 Introduction

The process of finding interesting patterns from enormous amounts of data is called data mining. Such rich and fascinating patterns can be valuable for large businesses and

for making smart decisions. This helps in improving customer relationship, developing marketing policies, improving sales and reduces costs.

Data mining is a multidisciplinary field which bonds statistics, machine learning, artificial intelligence and database technologies to predict future from large data repositories. The data mining methods such as association, classification and clustering can be applied on various kinds of data such as database data, transactional data and data warehouse.

The focus of this paper is on one of the techniques of data mining called classification. The process of sorting objects into similar groups is called classification. The classification process has two steps, first is the training step or the learning step where a classification model also known as a classifier finds correlations between the class labels and the features in a given dataset. In the second step, this classification model is supplied with test data to see the performance of the model. There are various application areas where classification can be useful like spam filtering, fraud detection, target marketing, customer attraction, customer retention, performance prediction, manufacturing, medical diagnosis etc. So there is a great need in the research field to improve the accuracy of the classifier. There are many classification models and much work has been carried out to improve the efficiency of these traditional models. Nearest neighbour is one of most popular classical classification model which has been chosen for this paper.

Using ensemble methods for improving the classification is one of the active research fields in machine learning. An ensemble blends a series of  $k$  trained models  $M_1, M_2, M_3, \dots, M_k$  with the aim to improved composite classification model,  $M^*$  [1]. In this paper a novel technique has

---

✉ Preeti Nair  
preeti.nair84@gmail.com

Navneet Khatri  
navneet.khatri@gmail.com

Indu Kashyap  
indu.fet@mriu.edu.in

<sup>1</sup> Manav Rachana University, Delhi Suraj Kund Road, Sector 43, Faridabad, Haryana 121004, India

<sup>2</sup> Maharshi Dayanand University, Delhi Road, University Secretariat, Rohtak, Haryana 124001, India

<sup>3</sup> T 2/1, DLF Phase 3, Gurgaon, Haryana 122002, India

been developed to improve the efficiency of Nearest Neighbor model through ensemble method.

## 2 Related work

Yu et al. [2], have proposed a paper on special dataset problems such as imbalance data, spare data, and have addressed using HBKNN and latter to tackle the noisy data problem in the high dimensional dataset, an ensemble method for random subspace with HBKNN (RS-HBKNN) is proposed, which outperforms most of the classification approaches.

Vinoth et al. [3], A text document classifier is proposed by incorporating the KNN classifier with support vector machine (SVM) classifier. In this proposed SVM-KNN, approach minimises the limitations in classification accuracy. Here the training data reduced from diverse classes and it is utilized by the SVM. The SVs from different classes provided as learning data for KNN classification algorithm where instead of Euclidean function the nearest centroid distance function is used this minimises the computation time for distance calculations.

Mittal and Gill [4], the aim of this approach is to propose a hybrid model for an efficient diabetic prediction. A hidden pattern is obtained by using feature selection on the dataset and then a two layered classification is applied on this refined data. Here in this study SVM classifier and Neural Network classifier are hybridized. This proposed system is compared with other few classifiers and has got good accuracy.

Aci et al. [5], the aim of this study is to eliminate those data which are difficult to train. The three classification algorithm k nearest neighbour, Bayesian methods and genetic algorithm is combined to produce a new dataset out of the original data. This method is tested on various dataset. The new data gives better classification accuracy than the old data.

Miloud-Aouidate and Baba-Ali [6], this paper is based on the hybridization of kNN and ant colony optimisation, using condensing approach with respect to kNN. Condensing allow to reduce the instances from the datasets significantly so that the accuracy of this training set is very close to the complete set of the training data. So the condensing kNN and optimization approach together formed a novel algorithm which outperforms the standard kNN and other condensed kNN algorithms.

## 3 Classification approaches

In this study classification models such as nearest neighbors, rotation forest, and simple logistic models are hybridised using the ensemble method called STACKING.

### 3.1 Nearest neighbour

Classifier or k-nearest neighbour classifier is the instance based learner classifier that compares an unknown instance with the training instances similar to it. When an unknown instance is given, the k-NN classifier searches for the k instances closest to the unknown instance. Here in this paper the value for k is equal to 1 i.e. only 1NN is considered. The closeness is obtained by distance metrics like Euclidean distance, Manhattan distance or Minkowski distance. Euclidean distance is considered here. To calculate Euclidian distance from two points  $p = (p_1, p_2)$  and  $q = (q_1, q_2)$ , the distance formula is as follows: [7, 8]

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}. \quad (2)$$

### 3.2 Rotation forest

Classifier produces ensembles of classifiers. In this classifier the dataset features are split into k subsets, and on each subset the principle component analysis is applied and a new set of subsets. Now the data is transformed into new features and along with it, a decision tree is learned. The splits in the decision tree leads to different rotations. Like this, a diverse classifier is obtained. The information about the spread out of the data is completely stored in the new space of extracted features [9]. This is how accurate separate classifiers are built. In rotation forest classifier, the aim is to get diverse and accurate classifiers concurrently. This method is called rotation forest because it's a combination of principle component analysis which is a rotation of coordinate axes and the base classifier model is a decision tree [9].

### 3.3 Simple logistic regression

The simple logistic model is a supervised classification model. It is more like a linear classifier which takes the calculated logits scores and weights to forecast the class label. The score is more like the numerical correspondent to a particular categorical attribute. The weights are like the weightages corresponding to the particular target. Now these scores and weights are multiplied to form logits. The

logits are then passed to a function called normalized exponential function which will return the probabilities for target class. The class with the highest probability will be the predicted class for the given unknown instance. Formula for normalized exponential function: [10].

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^k e^{z_k}} \quad \text{for } j = 1, \dots, k \quad (3)$$

where k-dimensional vector of real values in the range 0–1. The simple logistic classifier is selected as the meta classifier for stacking which is the second phase of stacking process. The meta classifier plays an important role in getting good results in the proposed hybrid model.

### 3.4 Ensemble stacking

Classification is a two phased process. In the first phase base classifiers train the datasets using j-cross-validation,  $c_1, c_2, \dots, c_i$  produces  $y_1, y_2, \dots, y_i$  as output of the classifier and then this output is the input to the meta classifier which is the second phase of stacking. The meta classifier minimises the error in order to optimise the base classifiers. This process is repeated for k-cross validation to get the final stacked classification model [11–18] (Fig. 1).

### 4 Working of proposed hybrid model

In this paper, a unique hybridization of 1-NN model (where the parameter value for  $k = 1$ ) and rotation forest model are taken as the base classification model and meta classifier as simple logistic model in stacking, is proposed. The basic working behind the proposed hybrid model is as follows:

- A training dataset D is send separately to each of the base classification models [1-NN and rotation forest] present in the stacking phase 1.
- The two outputs of the two classification models of stacking phase 1 are combined to form a single new dataset  $D_h$ .

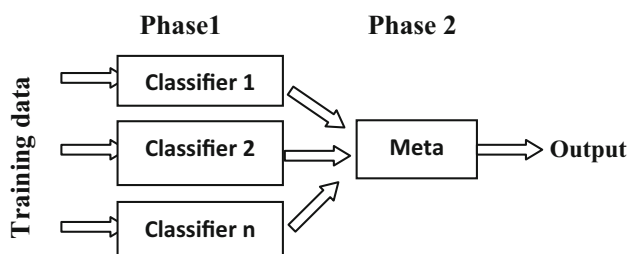


Fig. 1 The stacking process

- This dataset acts as the input training data for the meta classification model [Simple logistic model] of stacking phase 2 which produces the final output  $H$ .  $H$  is the final prediction of the proposed hybrid model.

The proposed hybrid model is compared with other ensemble methods like bagging, boosting, and logitboost, and it has also compared with six other classification models like Naive bayes, Bayesian Network, simple CART (classification and regression tree), J48, decision tree, decision table upon 13 datasets taken from UCI repository [19] (Fig. 2).

### 5 Datasets used

There are thirteen datasets used in this paper which are taken from UCI data repository. The datasets are a mix of all kinds of data, some are categorical in nature, some are numeric and there are dichotomous class problems as well as multi class problems (Table 1).

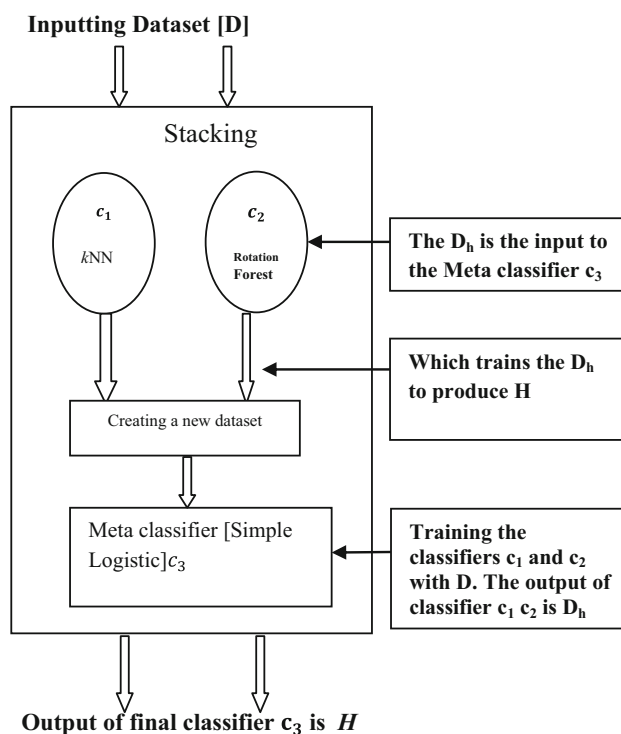


Fig. 2 Proposed hybrid model

**Table 1** Datasets taken from UCI repository

Datasets	No. attributes	Classes	Instances
Diabetes	9	2	768
Horse-colic	23	2	368
Anneal	39	10	898
Balance scale	5	3	625
Cylinder bands	40	2	550
Ecoli	8	8	336
Ionosphere	35	2	351
Mfeat-factor	217	10	2000
Postoperative	90	3	90
Sonar	61	2	208
Spambase	58	2	4601
Thoracic surgery	17	2	470
Hypothyroid	30	4	3772

## 6 Results and discussions

### 6.1 Computation of performance of classification models

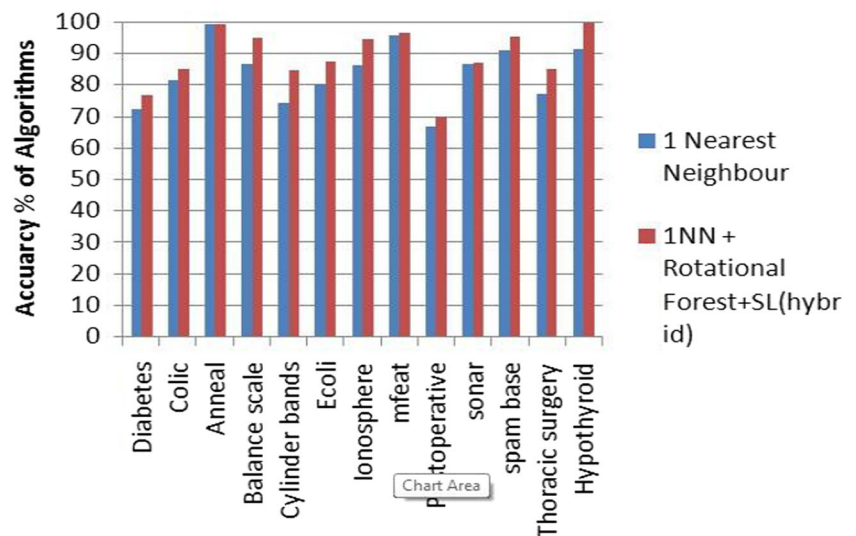
There are certain measures to evaluate the performance of the classification model. Here in this paper binary classification problem as well as multi class problems are taken under consideration. The classification algorithms selected in this whole study can tackle multiclass problems as well.

Quality of model is treated as good if the maximum number of items is correctly classified. True positive and True negative values are evaluated, which tells when the classifier is getting things right, while False positive and False negative tells when the classifier is getting things wrong. There are many other measures to estimate a classifier like TP rate, FP rate, F-measure, precision and

**Table 2** Shows the comparison between proposed hybrid model and standard k-nearest neighbor model in terms of accuracy values

Data sets	1 nearest neighbour	1NN + rotation forest + SL(hybrid)
Diabetes	72.3776	76.8229
Colic	81.25	85.0543
Anneal	99.1091	99.2205
Balance scale	86.56	95.04
Cylinder bands	74.4444	84.6296
Ecoli	80.3571	87.5
Ionosphere	86.3248	94.302
Mfeat	95.85	96.6
Postoperative	66.6667	70
Sonar	86.5385	87.0192
Spam base	90.7846	95.3271
Thoracic surgery	77.234	84.8936
Hypothyroid	91.5164	99.5758

**Fig. 3** Comparison between the proposed hybrid model and the standard kNN



**Table 3** Shows the comparison of proposed hybrid model and other classification models

Datasets	Hybrid	NBayes	BayesNet.	ABOost.	Baggi.	Logitb.	D. table	D. tree	j48	CART
Diabetes	<b>76.8229</b>	76.3021	74.349	74.349	75.7813	74.0885	71.224	71.875	73.8281	75.1302
Colic	85.0543	77.9891	81.25	81.25	86.413	81.5217	82.8804	81.5217	85.3261	<b>86.1413</b>
Anneal	<b>99.2205</b>	86.3029	96.2138	83.6303	98.5523	98.5523	94.3207	77.1715	98.441	98.3296
Balance scale	<b>95.04</b>	90.4	72.32	72.32	83.36	88	73.12	55.04	76.64	79.04
Cylinder bands	<b>84.6296</b>	72.2222	73.5185	72.5926	60.3704	72.037	64.6296	69.2593	57.7778	59.8148
Ecoli	<b>87.5</b>	85.4167	81.25	64.5833	81.5476	85.7143	76.7857	64.5833	84.2262	83.9286
Ionosphere	<b>94.302</b>	82.6211	89.4587	90.8832	91.1681	91.1681	89.4587	82.6211	91.453	89.7436
Mfeat factor	<b>96.6</b>	92.65	93.15	19.75	93.8	94.3	67.95	19.75	88.95	88.05
Postoperative	<b>70</b>	66.6667	64.4444	<b>70</b>	<b>70</b>	66.6667	68.8889	<b>70</b>	<b>70</b>	71.1111
Sonar	<b>87.0192</b>	67.7885	80.2885	71.6346	76.9231	79.3269	69.2308	73.0769	71.1538	71.1538
Spam base	<b>95.3271</b>	79.2871	89.8066	90.0674	94.0013	92.0017	90.3065	78.0483	92.9798	92.4364
Thoracic surgery	84.8936	78.5106	81.9149	84.0426	83.8298	83.617	84.4681	84.0426	84.4681	<b>85.1064</b>
Hypothyroid	<b>99.5758</b>	95.281	98.5949	93.2131	99.5228	<b>99.5758</b>	99.3372	95.3871	<b>99.5758</b>	99.5493

Bold values represent the highest accuracy values

recall. True positives ( $T_p$ ): when a classification model correctly predicts the label as positive. True negatives ( $T_n$ ): when a classification model correctly predicts the label as negative. False positives ( $F_p$ ): when a classification model predicts the label incorrectly as positive. False negatives ( $F_n$ ): when a classification model predicts the label incorrectly as negative [20].

The performance of the classification is obtained by calculating accuracy, precision, and recall. Accuracy measures the rate of total correct  $T_p$  and  $T_n$  predictions to all predictions. Precision quantifies the correctness rate of the class predictions as positive by the classification model and recall measures the rate of positives are correctly predicted as positive. They both are opposite so comparison between them are difficult. So, a single measure from precision and recall could be computed, called F-measure [20].

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{4}$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{5}$$

$$\text{Recall} = \frac{T_p}{T_p + T_n} \tag{6}$$

$$\text{Fmeasure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{7}$$

F-measure is the weighted harmonic mean of Precision and Recall.

### 6.2 Experimental results

The results of this experiment are shown in two different tables. The Table 2 shows that the proposed hybrid model has outperformed the traditional nearest neighbor model significantly in terms of accuracy, in most of the cases.

Following is the graphical representation of comparison between the proposed hybrid model and the standard kNN. For both models the values for  $k = 1$  (Fig. 3).

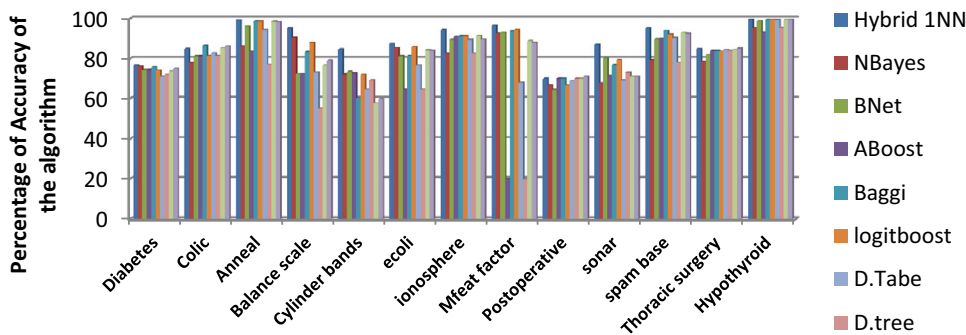
With respect to each datasets in the graph, it is observed that the proposed hybrid method is significantly better than the standard 1NN model.

The Table 3 shows the comparison between the proposed Hybrid model and 9 other standard classification models, out of these 9 models 3 are ensemble based models, bagging, Ad boost and logitboost. For this comparison, the same above mentioned 13 datasets has been taken. In 9 out of 13 datasets cases, the proposed hybrid model has shown highest accuracy, and in 2 cases proposed hybrid model has same highest values as other two models in this study and in only 2 cases the proposed system has the lower accuracy values. The models used here are Naive Bayes, Bayesian networks, decision table, decision tree, J48, simple CART.

The experiment was aimed at comparing performance of the proposed hybrid model using stacking against the standard kNN model and it was also compared with six other standard models as well as three ensemble models. Here T-paired test has done to show the comparison of the proposed hybrid method using stacking with other standard method in terms of accuracy, Fmeasure and ranking. The table also shows the accuracy of the models and the standard deviations values. Here confidence level is taken .05.



**Fig. 4** Performance of the proposed hybrid model compared to other classification models



Dataset	meta.Stacking	(1) bayes.Naive	(2) bayes.Bayes	(3) meta.AdaBoos	(4) meta.Baggin	(6) rules.Decisi	(7) trees.Decisi	(8) trees.J48	(9) trees.Simple
pima_diabetes	76.83(4.47)	76.31(5.52)	74.36(4.71)	74.35( 4.73) *	75.78(2.65)	71.23( 4.05) *	71.87(5.08) *	73.83(5.66) *	75.13( 4.09)
horse-colic	85.06(6.39)	78.00(4.10) *	81.25(4.84)	81.26( 5.86) *	86.38(5.68)	82.87( 6.03)	81.52(4.59)	85.30(4.75)	86.11( 4.98)
anneal	99.22(0.76)	86.30(2.51) *	96.21(2.24) *	83.63( 0.48) *	98.55(1.18)	94.32( 3.03) *	77.16(4.02) *	98.44(1.08) *	98.33( 1.08) *
balance-scale	95.04(1.59)	90.39(1.72) *	72.30(5.76) *	72.31( 4.93) *	83.36(4.07) *	73.13( 5.16) *	55.06(4.91) *	76.65(3.77) *	79.05( 4.53) *
cylinder-bands	84.63(5.86)	72.22(4.45) *	73.52(8.60) *	72.59( 6.10) *	60.37(4.11) *	64.63( 4.14) *	69.26(5.53) *	57.78(0.78) *	59.81( 4.10) *
ecoli	87.49(4.83)	85.40(5.39)	81.23(4.57) *	64.58( 1.68) *	81.53(7.86) *	76.77( 4.68) *	64.58(1.68) *	84.23(7.51)	83.90( 6.79) *
ionosphere	94.33(3.71)	82.62(5.47) *	89.46(4.47) *	90.90( 4.16) *	91.18(4.30) *	89.47( 5.20) *	82.62(4.76) *	91.46(3.27) *	89.76( 4.22) *
mfeat	96.60(1.33)	92.65(2.19) *	93.15(2.01) *	19.75( 0.35) *	93.80(2.16) *	67.95( 2.89) *	19.75(0.35) *	88.95(2.52) *	88.05( 1.50) *
postoperative-patient-dat	70.00(5.37)	66.67(5.24)	64.44(4.68) *	70.00( 5.37)	70.00(5.37)	68.89( 8.76)	70.00(5.37)	70.00(5.37)	71.11( 5.74)
sonar	87.02(5.52)	67.88(9.29) *	80.29(6.12) *	71.67(10.87) *	76.90(9.27) *	69.21(12.21) *	73.05(8.28) *	71.17(7.10) *	71.21(10.19) *
spambase	95.33(0.94)	79.29(1.55) *	89.81(0.92) *	90.07( 1.86) *	94.00(1.01) *	90.31( 1.22) *	78.05(1.50) *	92.98(1.17) *	92.44( 1.34) *
Thoracic_Surgery_Data	84.89(0.67)	78.51(9.27)	81.91(4.17)	84.04( 1.12) *	83.83(1.49)	84.47( 1.03)	84.04(1.12) *	84.47(1.03)	85.11( 0.00)
hypothyroid	99.58(0.22)	95.28(0.83) *	98.60(0.50) *	93.21( 1.01) *	99.52(0.27)	99.34( 0.29) *	95.39(0.53) *	99.58(0.29)	99.55( 0.25)

**Fig. 5** Classification accuracy and standard deviation of proposed hybrid models and other standard models. \*proposed model is significantly better than the other models

**Table 4** Ranking of the classification models

Models	Ranking (win-loss)	Wins	Losses
Hybrid INN model	<b>75</b>	<b>75</b>	<b>0</b>
Bagging	29	42	13
Simple CART	21	41	20
J48	11	32	21
Bayes net	- 3	30	33
Naive Bayes	- 11	25	36
Decision table	- 27	18	45
Ad boost	- 37	13	50
Decision tree	- 58	6	64

Bold values represent a rank of all models

Below is the graphical presentation of the above mentioned table of comparison between the proposed hybrid model and other classification models in terms of accuracy. The proposed hybrid model has scored the highest in all cases (Fig. 4).

In Fig. 5 The Fmeasure of proposed hybrid model has scored high in maximum cases. The last row shows the win/tie/loss of the entire models in terms of Fmeasure. It is observed that none of the models has obtained any wins against the proposed hybrid method and ranking a model takes place by the number of times a given model beat the other models in j-cross validation. Table 4 shows the ranking of the classification models according to the difference between the number of times each model has been significantly better and worse than another models.

Dataset	meta.Stacking	(1) bayes.Naiv	(2) bayes.Baye	(3) meta.AdaBo	(4) meta.Baggin	(6) rules.Deci	(7) trees.Deci	(8) trees.J48	(9) trees.Simp
pima_diabetes	0.83(0.03)	0.82(0.04)	0.81(0.03)	0.81(0.03) *	0.82(0.02)	0.78(0.03) *	0.79(0.04) *	0.80(0.04) *	0.82(0.03)
horse-colic	0.88(0.05)	0.82(0.04) *	0.85(0.04)	0.85(0.05) *	0.90(0.04)	0.87(0.04)	0.85(0.05) *	0.89(0.04)	0.89(0.04)
anneal	0.70(0.48)	0.63(0.46)	0.67(0.47)	0.00(0.00) *	0.50(0.53)	0.18(0.23) *	0.00(0.00) *	0.50(0.53)	0.40(0.52)
balance-scale	0.97(0.01)	0.94(0.02) *	0.75(0.06) *	0.75(0.05) *	0.88(0.04) *	0.76(0.05) *	0.54(0.11) *	0.80(0.04) *	0.82(0.05) *
cylinder-bands	0.82(0.07)	0.66(0.05) *	0.69(0.10) *	0.62(0.11) *	0.29(0.12) *	0.51(0.14) *	0.42(0.15) *	0.00(0.00) *	0.25(0.16) *
ecoli	0.96(0.04)	0.95(0.03)	0.94(0.04) *	0.79(0.03) *	0.94(0.04) *	0.91(0.04) *	0.79(0.03) *	0.95(0.03)	0.94(0.04) *
ionosphere	0.92(0.05)	0.78(0.06) *	0.85(0.08) *	0.86(0.07) *	0.87(0.06) *	0.84(0.08) *	0.68(0.10) *	0.87(0.05) *	0.85(0.06) *
mfeat	0.98(0.02)	0.94(0.05) *	0.96(0.04)	0.16(0.17) *	0.96(0.03) *	0.62(0.12) *	0.16(0.17) *	0.94(0.05) *	0.95(0.05) *
postoperative-patient-dat	0.82(0.04)	0.80(0.04)	0.79(0.04)	0.82(0.04)	0.82(0.04)	0.81(0.06)	0.82(0.04)	0.82(0.04)	0.83(0.04)
sonar	0.85(0.07)	0.70(0.09) *	0.77(0.07) *	0.69(0.11) *	0.73(0.10) *	0.66(0.13) *	0.69(0.09) *	0.70(0.08) *	0.69(0.11) *
spambase	0.96(0.01)	0.80(0.02) *	0.92(0.01) *	0.92(0.01) *	0.95(0.01) *	0.92(0.01) *	0.83(0.02) *	0.94(0.01) *	0.94(0.01) *
Thoracic_Surgery_Data	0.00(0.00)	0.15(0.13) v	0.07(0.11)	0.02(0.06)	0.00(0.00)	0.04(0.09)	0.02(0.06)	0.02(0.06)	0.00(0.00)
hypothyroid	1.00(0.00)	0.98(0.00) *	0.99(0.00) *	0.97(0.00) *	1.00(0.00)	1.00(0.00) *	0.99(0.00) *	1.00(0.00)	1.00(0.00)
	(w/ /*)	(1/4/8)	(0/6/7)	(0/2/11)	(0/6/7)	(0/3/10)	(0/2/11)	(0/6/7)	(0/6/7)

**Fig. 6** Shows the Fmeasure of all the models against the proposed hybrid model

Proposed hybrid model has the maximum wins in ranking (Fig. 6).

## 7 Conclusion and future scope

In this paper, a novel technique is implemented using ensemble method in which it is observed that the proposed hybrid model has given very good results as compared to other ensemble models and classic models tested upon 13 datasets from UCI repository. The method not only proved to be a good method for binary class problems but also good enough for multiclass problems as well. This hybridization of three classifiers 1NN, rotation forest as base classifiers and simple logistic regression as meta classifier built in stacking ensemble method has improved the accuracy rate of the standard 1NN significantly. In terms of future application this technique can be used with other classification algorithms as well.

## References

- Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques. Morgan Kaufmann, Waltham
- Yu Z, Chen H, Liu J, You J, Leung H, Han G (2015) Hybrid k-nearest neighbor classifier. *IEEE Trans Cybern* 46(6):1263–1275
- Vinoth R, Jayachandran A, Balaji M, Srinivasan R (2014) A hybrid text classification approach using KNN and SVM. *Int J Adv Found Res Comput (IJAFRC)* 1(3):20–26
- Gill NS, Mittal P (2016) Computational hybrid model, with two level classification using SVM and neural network for predicting diabetes disease. *J Theor Appl Info Technol* 87(1):1–10
- Aci M, İnan C, Avci M (2010) A hybrid classification method of k nearest neighbor, Bayesian methods and genetic algorithm. *Expert Syst Appl* 37(7):5061–5067
- Miloud-Aouidate A, Baba-Ali AR (2012) A hybrid KNN-ant colony optimization algorithm for prototype selection. *Laboratory of Robotics, Parallelism and Embedded University of Sciences and Technology Houari Boumediene, USTHB, T. Huang et al. (Eds.): ICONIP 2012, © Springer-Verlag Berlin Heidelberg*
- Sahu SK, Kumar P, Singh AP (2018) Modified K-NN algorithm for classification problems with improved accuracy. *Int J Inf Technol* 10(1):65–70
- Zhang S, Li X, Zong M, Zhu X, Cheng D (2017) Learning k for kNN classification. *ACM Trans Intell Syst Technol* 8(3):43
- Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 28(10):19–30
- Sumner M, Frank E, Hall M (2005) Speeding up logistic model tree induction. In: 9th European conference on principles and practice of knowledge discovery in databases
- Wolpert DH (1992) Stacked generalization. *Neural networks. IEEE Trans Pattern Anal Mach Intell* 28(10):241–259
- Hsu KW (2017) A theoretical analysis of why hybrid ensembles work. *Hindawi Comput Intell Neurosci*. <https://doi.org/10.1155/2017/1930702>
- Salunkhe UR, Mali SN (2016) classifier ensemble design for imbalanced data classification: a hybrid approach. In: International conference on computational modelling and security (CMS)
- Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF (2017) SVM and SVM ensembles in breast cancer. *PLOS ONE* 12(1):e0161501
- Huda S, Yerarwood J, Jelinek HF, Hasssan MM, Fortin G, Buckland M (2016) A hybrid feature selection with ensemble classification for imbalanced access. *IEEE Access* 4:9145–9154
- Chen Y, Zhao X, Lin Z (2014) Optimizing subspace SVM ensemble for hyperspectral imagery classification. *IEEE J Sel Topics Appl Earth Observations Remote Sens* 7(4):1295–1305
- Campos R, Canuto S, Salles T, de Sá CC, Gonçalves MA (2017) Stacking bagged and boosted forests for effective automated classification. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval
- El Bialy R, Salama MA, Karam O (2016) An ensemble model for heart disease data sets: a generalized model. In: Proceedings of the 10th International Conference on Informatics and Systems, INFOS, May 09–11. ACM
- Tang J, Alelyani S, Liu H (2015) Data classification: algorithms and applications, data mining and knowledge discovery series. CRC Press, Boca Raton
- Agaoglu M (2016) Predicting instructor performance using data mining techniques in higher education. *IEEE Access* 4:2379–2387