



A nested stacking ensemble model for predicting districts with high and low maternal mortality ratio (MMR) in India

Sourabh Shastri¹ · Paramjit Kour¹ · Sachin Kumar¹ · Kuljeet Singh¹ · Anand Sharma² · Vibhakar Mansotra¹

Received: 15 June 2020 / Accepted: 6 November 2020 / Published online: 22 November 2020
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2020

Abstract The ensemble is an efficacious machine learning framework that combines variety of algorithms for better performance and effective prediction. Over the past few years, numerous researchers proposed wide variety of ensemble methodologies in the field of healthcare industry. In the present research paper, a nested ensemble has been suggested based on Stacking and Voting schemes for prediction and analysis of Maternal Mortality Ratio (MMR) in India. The presented nested ensemble combines Base Learners and Meta Learners by employing different classification algorithms and prediction results were afterwards evaluated by using K-fold cross validation and thus, facilitating the statistical distribution of results. Further, the effectiveness of the ensemble was investigated by comparing its performance with the various single learning

algorithms in terms of accuracy, precision, recall, F-measure and ROC.

Keywords Ensemble · Stacking · Voting · Base learners · Meta learners

1 Introduction

India, a low-middle income country is the second most populous country of the world that has been drawing attention concerning health profile. Globally, the Maternal Mortality Ratio (MMR) declined from 385/100,000 live births in 1990 to 216/100,000 in 2015. In a similar time-frame, estimated MMR of India ranged from 556 to 174 per 100,000 live births in between 1990 and 2015 and by 2015, the country contributed to 15% of global maternal deaths [7]. Maternal mortality highlights a health burden on women during and just after pregnancy when women are at risk of complications, particularly in developing countries like India.

Maternal Mortality Ratio (MMR) is a standard measure for measuring maternal deaths which mainly occur while a woman is pregnant or within 42 days of the termination of pregnancy, during labor or delivery or after childbirth. These deaths mostly happen due to preventable causes [28, 48] which include ante-partum hemorrhage, postpartum hemorrhage, ruptured uterus, high blood pressure or eclampsia, severe bleeding, infection after termination of pregnancy and pulmonary embolism. The prominent factors like early age marriage or pregnancy, poverty, malnutrition, illiteracy, unsafe abortion and less time gap between two deliveries make the condition even more dangerous for mothers living in remote areas [25] and in order to accommodate this, Government of India has

Vibhakar Mansotra: Mentor.

✉ Sourabh Shastri
sourabhshastri@gmail.com

Paramjit Kour
Param.kour.94@gmail.com

Sachin Kumar
skh.sachinkumar@gmail.com

Kuljeet Singh
kuljeetshan94@gmail.com

Anand Sharma
andz24@gmail.com

Vibhakar Mansotra
vibhakar20@yahoo.co.in

¹ Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, Kathua 184104, India

² Department of IT and Research, Guru Kashi University, Punjab, India

implemented bundles of schemes for reducing maternal mortality for the improvement of the health of pregnant women falling in the reproductive age group. A multi-strategy initiative like National Rural Health Mission (NRHM) was launched in 2005 focusing on the reproductive, maternal, newborn or child healthcare strategies in order to strengthen the health system and the same was renamed to National Health Mission (NHM) in 2012 [27]. Despite of all the initiated schemes, application of recent technologies is a requisite for improving and reducing the burden of mortality rate present in India.

Owing to the inclining use of electronic health (e-health) systems by health organizations in India, the flow of medical information has also increased, thereby, necessitating the usage of intelligent automated systems for early detection of problems. This study aims at employing multiple data mining algorithms in medical domain followed by analysis on the basis of outcomes. Afterwards, plan concerning detailed evaluation of problems in pregnancy is sketched out as early stage detection and treatment of causes can reduce the number of deaths of women during childbearing and further, keeping into consideration the problem, various methodologies can be developed with the help of data mining and machine learning algorithms.

Data Mining refers to the identification and extraction of useful information from large collection of raw healthcare data [5] as Data Mining is able to search valuable information in the sector of healthcare, which mainly can be used for predicting various diseases, automated decision system or assistance for the doctors in making decisions [18, 30]. Depending on the type of dataset and how it is implemented, data mining algorithms have different powers in classification, clustering and prediction. Since, single selection algorithm seems to be incapable of ensuring optimal results in terms of prediction and stability, thus the effectiveness of ensemble approach involving the combination of different algorithms [6, 44] was explored. The ensemble approach has been found to be more effective in the growth of healthcare data mining and shows more promising predictive results as compared to a single classification algorithm applied on training and testing dataset. The training dataset is used to train the learners and model building after which the trained learners are combined using stacking ensemble technique and prediction is computed. The result, thus, is evaluated by comparing the predicted results of single learners and ensembles with K-fold cross validation [9, 10]. As a step further, majority voting [41] has been applied as baseline method for wrapping the combined learners and averaging the prediction of learned combiners.

In the second section of the present research paper, work related to proposed ensemble methodologies in the area of medical science was studied and explored whereas the proposed ensemble methodology, architecture and algorithm were scrutinized in the third section. In the section four, the experiments were thoroughly explored followed by the presentation of results and discussion in the fifth section. Conclusion, future scope and benefits of the proposed methodology were examined in the last section of the research paper.

2 Related work

Over the last couple of years, researchers have worked out a lot of ensemble methodologies for analysis and prediction in medical domain. The ensuing paragraphs reflect the review of literature of research work carried out by the researchers.

Abdar et al. [3] proposed two-layer nested ensemble for early detection of breast cancer by employing classifier and Meta Classifiers. They combined the nature of stacking and voting ensemble techniques and variation in classification algorithms was accomplished in Meta Classifier. The Wisconsin Diagnostic Breast Cancer Dataset was used for conducting experiment and evaluation of model was done on the basis of K-fold cross validation, wherein the results indicated that two-layer nested ensemble performed better as compared to single classifiers and SV-NaïveBayes-3-Meta Classifier took less time to build model as it was discerned to be more efficient towards diagnosis of breast cancer. Esener et al. [22] presented a framework for breast cancer diagnosis and prediction employing feature ensemble with multistage classification scheme. They collected a publicly available mammogram dataset during the Image Retrieval in Medical Applications (IRMA) project and three groups of features were concatenated to construct the feature vectors which were local configuration pattern-based, statistical and frequency domain features. After feature extraction, eight well-known classification algorithms were applied in three stages i.e. one-stage study, two-stage study and three-stage study with 11-fold cross-validation. The results indicated that the performances were combined via a majority voting technique to improve the recognition accuracy and multistage classification scheme was found to be more effective than the single-stage classification for prediction. Moreira et al. [40] created an ensemble with nearest-neighbor classifiers using the random subspace algorithm which classified unbalanced pregnancy database. The performance of

proposed ensemble was evaluated by Area under Curve (AUC) and other indicators of confusion matrix using ten-fold cross-validation method. This approach predicted the Apgar score and gestational age during childbirth which could be strongly associated with the neonatal death risk and also predicted fetus-related problems that developed hypertensive disorders in pregnancy. Cong et al. [19] proposed a selective ensemble method using KNN, SVM and Naive Bayes as the Base Classifier with ten-fold cross-validation in order to diagnose breast cancer. Different ultrasound images were combined with mammography images to calculate gray level co-occurrence matrix (GLCM) which can provide a method for generating texture features and extracting morphological features. The selective ensemble method was noted to be efficient in diagnosing the breast cancer along with the classifier-fusion method as compared to the feature fusion method during the study.

Researchers Kabir and Ludwig [31] presented a technique called super learning or stacked-ensemble coming up with the optimal weighted average of diverse learning models achieving better performance than the individual base classifier. Bashir et al. [9, 10] conducted a study wherein Naive Bayes, Decision Tree based on Gini Index and Information Gain, Instance-based classifier and Support Vector Machine were applied across heart dataset and achieved the accuracy of 87.37% with ten-fold cross-validation. Rahman et al. [45] utilized ensemble method of three data mining modeling techniques viz. Logistic Regression, Naïve Bayes and Neural Network for Robust Intelligent Heart Disease Prediction System (RIHDPS) and the same could predict accurately just by analyzing the history of heart disease in a patient. The proposed RIHDPS produced an accuracy of 91.26% and logistic regression decision boundary using Principal Component Analysis (PCA). Bashir et al. [11] developed an application ‘IntelliHealth’ based on proposed model called ‘HM-BagMoov’, an ensemble framework with multi-layer classification using enhanced bagging and optimized weighting applied on five different datasets that may be used by hospitals/doctors for diagnosis/advice. Abdar et al. [1, 2] applied novel decision tree based algorithm on liver disease dataset

and observed that C5.0 algorithm via. Boosting technique achieved an accuracy of 93.75% which was found to be better than the boosted CHAID algorithm. Shastri and Mansotra [46] designed a conceptual framework viz. KDD-MHCI based on Knowledge Discovery in Databases (KDD) for discovering knowledge from the databases of maternal health and child immunization (MHCI).

3 Methodology

3.1 Proposed stacking model

This section introduces the new proposed methodology, architecture and algorithm with its fundamental design and the features of each of three parts of proposed method.

Let D_s is the dataset, f_n is the set of feature vectors, t_n is the set of target variables and $L = \{L_1, L_2, L_3, \dots, L_n\}$ is the set of algorithms that will be applied on dataset D_s . In proposed K-level nested stacking, two or more stacking techniques can be combined to arrive at better performance than single learners. The proposed system has multiple levels of stacks (nested stacking) where stacking can be applied K times and is very flexible to use several Base Learners as shown in Fig. 1. Assume, we have stacking learning technique and different combination of algorithms $L_1 = \{L_{1,1}, L_{1,2}, \dots, L_{1,N}\}$, $\{L_{2,1}, L_{2,2}, \dots, L_{2,N}\}$, \dots , $\{L_{M,1}, L_{M,2}, \dots, L_{M,N}\}$ as Base Learners with Meta Learners. The output obtained will be $\{P_{1,1}, P_{1,2}, \dots, P_{1,N}\}$ of Level-1 and if there arises the need of more stacking then we move to next levels of stacking. After implementing the combination of these learners, the majority voting technique is applied for calculation of final output. If optimal result is achieved in 1-Level, there is no further need to apply stacking two or more times.

The proposed K-level nested stacking framework is shown in Fig. 1 and the proposed algorithm for nested stacking with K-fold cross validation is depicted in Algorithm 1.

Algorithm 1: Nested Stacking with K-fold Cross Validation

INPUT: Dataset, $D_S = \{(f_n, t_n); n=1, 2, 3, \dots, N\}$
 where $f_n = \{F_1, F_2, F_3, \dots, F_n\}$ is a set of features vector and
 $t_n = \{t_1, t_2, \dots, t_n\}$ is a set of target variables.
 $L = \{L_1, L_2, L_3, \dots, L_n\}$ is a set of learning algorithms.

OUTPUT: Final_Prediction = P_f

1. **Level-1 Stacking** (D_S, L, K)
 2. Initially $D = \phi$
 3. Apply Feature_Selection_Methods (F_S_M)
 $D = F_S_M(D_S)$ where $D = \{(f_n, t_n); n = 1, 2, 3, \dots, N\}$ s.t. f_n is a set of selected features and $f_n \leq F_n$
 4. **Output:** D is now new dataset that is used for further steps
 5. Randomly split D into K almost equal size subsets: $\{D_1, D_2, D_3, \dots, D_k\}$, where $D_k = \text{Number of records / folds}$
 6. **for** $i \leftarrow 1$ to k **do**
 7. test_set = D_i ; training_set = $D^{(-i)} = D - D_i$
 8. **for** $l \leftarrow 1$ to n **do**
 9. $W_l^{(-i)} = L_l(D^{(-i)})$
 10. **end for**
 11. **end for**
 12. **for** $l \leftarrow 1$ to n **do**
 13. **for** $i \leftarrow 1$ to k **do**
 14. **for each** t_n in D_i **do**
 15. X_{iln} = prediction of model $W_l^{(-i)}$ on t_n in test_set
 16. **end for**
 17. **end for**
 18. **Output:** $D_{SS} = \{f_n, X_{111}, \dots, X_{kn(N/k)}\}$ is new Dataset for Nested Stacking
 19. **end of Level-1 Stacking**
 20. **if** (Nested_Stacking == NO) **do** // Voting
 21. **for** $i \leftarrow 1$ to k
 22. **for** $l \leftarrow 1$ to n
 23. $M_l^{(-i)} = L_l(X_{iln})$
 24. **end for**
 25. **end for**
 26. **for** $l \leftarrow 1$ to n **do**
 27. **for** $l \leftarrow 1$ to k **do**
 28. Y_{iln} = Prediction of Model $M_l^{(-i)}$ on t_n in test set
 29. **end for**
 30. **end for**
 31. $P_f = \text{Majority_Voting}(Y_{iln})$
 32. **return** P_f
 33. **else do** // Nested Stacking
 34. $P_S = \text{Stacking_Prediction} = \phi$
 35. $D_{NS} = \phi$
 36. **Nested_Stacking** (D_{SS})
 37. P_{temp} = Make Prediction on (D_{SS})
 38. $P_S = P_S + P_{temp}$
 39. **if** (P_S) meets Prediction Criteria
 40. **go to** Voting
 41. **else do**
 42. $D_{NS} = \text{Level-1 Stacking}(D_{SS})$
 43. **Nested Stacking** (D_{NS})
 44. **end else**
 45. **end else** // end of Nested Stacking
-

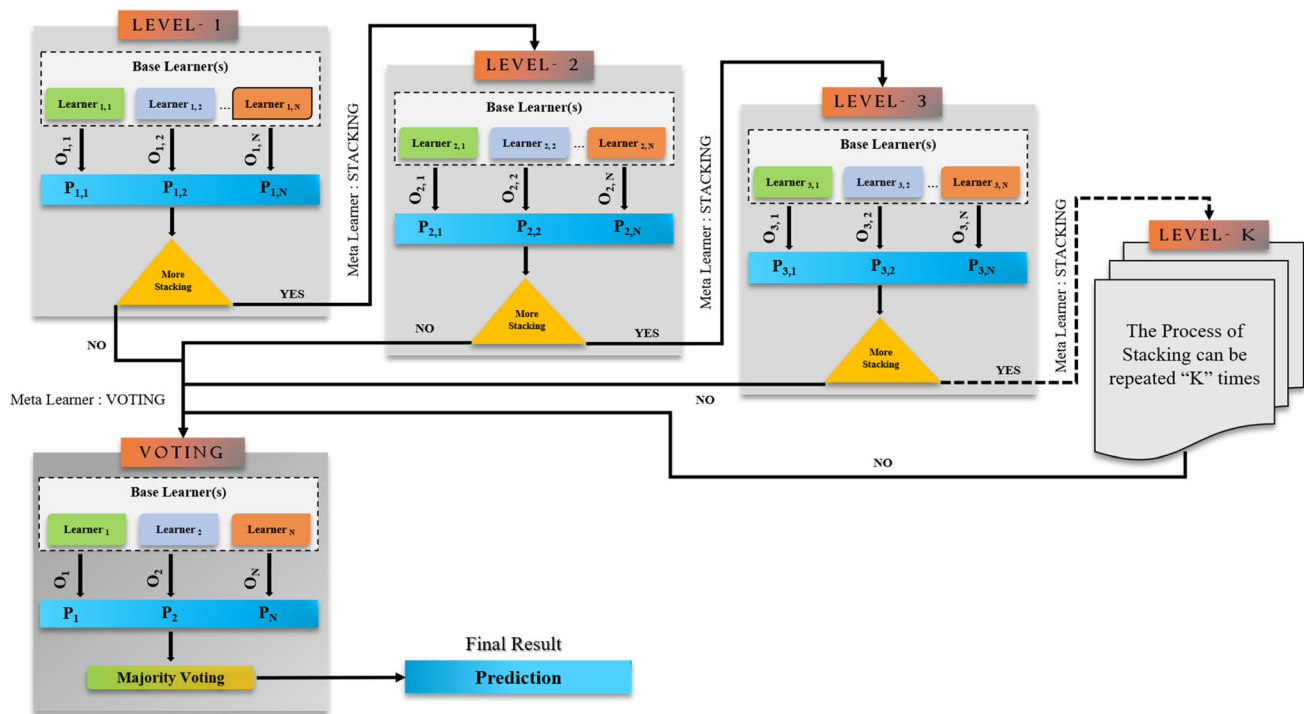


Fig. 1 Proposed K-level nested stacking framework

The present paper deals with the heterogeneous ensemble [42] method based on Stacking and Voting techniques. The nested stacking with K-fold cross-validation [17, 35] includes two learners namely Base Learner and Meta Learner and the performance of the model is verified using K-fold cross-validation [52]. The ensemble in this study is called SV-(n-Base, n-Meta) where SV is Stacking and Voting, n is as many numbers of learners used. The different combinations of these Base and Meta Learners are indicated in Table 1.

The individual classification algorithm was applied on extracted subset of features of training data and their performance was assessed based on K-fold cross-validation. As per nature of stacking, several level learners are primarily trained followed by the prediction. Test data classification is later on accomplished, firstly by producing the output of the Base Learner and then passing these outputs to the Meta Learner to give rise to the final prediction [15, 36]. The present work is only up to 1-level of stacking and the architecture for the same is depicted in Fig. 2.

The proposed heterogeneous ensemble is implemented by using eight different classification algorithms viz. Random Forest, Logistic, CART, JRip, PART (Base Learners) and Hoeffding tree, REPTree, J48 (Meta Learners). The description of techniques used is presented in Table 2.

4 Experiments

4.1 Dataset

The dataset used in this study was taken from Health Management Information System (HMIS) portal of Ministry of Health and Family Welfare (MoHFW), Government of India and it comprised of all 674 districts of India for the years 2014–18 and contained 33 parameters. Out of 674 districts, 386 districts were reflecting high MMR and the rest 288 were found to be low MMR districts. Out of 33 input parameters, the important 16 parameters were selected for modeling by using Forward Feature Selection technique of Wrapper Method which is depicted in Table 3. Additionally, a flag variable i.e. MMR with 2 values viz. High MMR and Low MMR was used as the class label for the present work.

4.2 Evaluation measures

To evaluate the performance, several performance measures were used:

- (a) Accuracy: The ratio of the number of correct predictions to the total number of predictions [20].
- (b) Precision: The ratio of number of true positives to the number of positives [29].

- (c) Recall: The ratio of correctly classified positives out of the total positives in that particular class [4, 8].
- (d) F-measure: The weighted average of precision and recall [39].
- (e) ROC: Receiver Operating Characteristic (ROC) curve is a probability curve showing the performance of a classification model by plotting true positive rate against true negative rate at varied threshold values [26, 33].

Furthermore, nested stacking ensemble approach was applied to verify substantial improvement, if any, in the prediction method.

4.3 Working environment

The experiments were done on WEKA environment. Waikato Environment for Knowledge Analysis (WEKA) is a free and open source software used for the study, implementation, construction or development of machine learning schemes [38] since it is freely available to the public and is widely used for research in the data mining and machine learning fields, as it represents a conglomeration of diverse machine learning methods for data visualization, classification, clustering, regression etc.

5 Results and discussion

5.1 Results without ensemble techniques

In this section, individual learners assessment was performed on Maternal Health dataset by varying the values of K-fold cross validation. Table 4 indicates the comparison of accuracy, precision, recall, F-measure and ROC results of individual learner techniques using CART, Random

Forest and JRip. It can be seen from Table 4 that Random Forest reflected better performance than other learners.

5.2 Results with ensemble techniques

The performances of SV-(n-Base, n-Meta) with variation of 2 or 3 Base and Meta Learners are reported in this section. The aim of using nested stacking ensemble was to comprehend the best suitability of methods to the available data and corroborate their effect on classification accuracy. The proposed ensemble model was applied on each test set with varied fold of cross validation for result calculation followed by analysis to verify the superiority of the same. In the present piece of research, the value of batch-size was 200. Table 5 reported the performance of JRip algorithm as Base with various combinations, out of which SV-(3-Base, 3-Meta) achieved better accuracy of 90.06% when K = 15. Table 6 depicted the performance of Random Forest algorithm as Base Learner with four combinations. Out of four combinations, three combinations viz. SV-(2-Base, 2-Meta), SV-(2-Base, 3-Meta) and SV-(3-Base, 2-Meta) were performing better at K = 15. The combinations SV-(2-Base, 2-Meta) and SV-(2-Base, 3-Meta) had the same accuracy of 90.80% whereas SV-(3-Base, 2-Meta) gave an accuracy of 91.10%. The difference of accuracy between 91.10% and 90.80% was discerned to be very minimal. So, to select the best model among these three combinations, other measures were also used and shown in Sects. 5.3 and 5.4. It is further recognizable from Table 7 that SV-(3-Base, 3-Meta) outperformed better with 90.65% accuracy using CART algorithm as Base at K = 15.

5.3 Overall comparison

The overall comparisons of accuracy, ROC and F-measure were made with single learner and proposed 1-level

Table 1 Combinations of selected base and meta learners

Models		Base learners	Meta learners
JRip	SV-(2-Base, 2-Meta)	JRip, Logistic	Hoeffding, REPTree
	SV-(2-Base, 3-Meta)	JRip, Logistic	Hoeffding, REPTree, J48
	SV-(3-Base, 2-Meta)	JRip, Logistic, PART	Hoeffding, REPTree
	SV-(3-Base, 3-Meta)	JRip, Logistic, PART	Hoeffding, REPTree, J48
Random Forest	SV-(2-Base, 2-Meta)	Random Forest, Logistic	Hoeffding, REPTree
	SV-(2-Base, 3-Meta)	Random Forest, Logistic	Hoeffding, REPTree, J48
	SV-(3-Base, 2-Meta)	Random Forest, Logistic, PART	Hoeffding, REPTree
	SV-(3-Base, 3-Meta)	Random Forest, Logistic, PART	Hoeffding, REPTree, J48
CART	SV-(2-Base, 2-Meta)	CART, Logistic	Hoeffding, REPTree
	SV-(2-Base, 3-Meta)	CART, Logistic	Hoeffding, REPTree, J48
	SV-(3-Base, 2-Meta)	CART, Logistic, PART	Hoeffding, REPTree
	SV-(3-Base, 3-Meta)	CART, Logistic, PART	Hoeffding, REPTree, J48

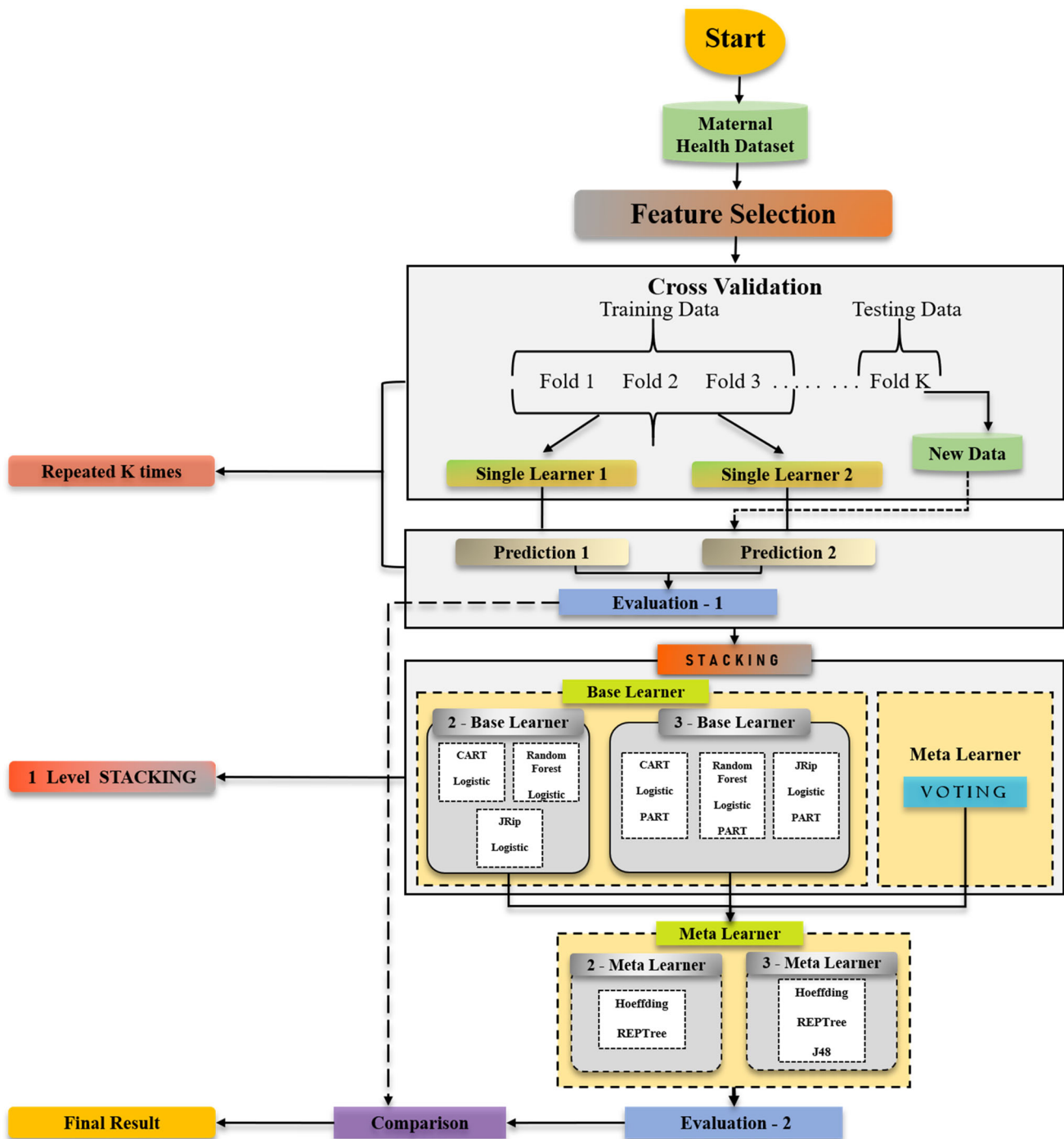


Fig. 2 1 level stacking architecture

heterogeneous ensemble with variation of K-fold cross-validation which is reflected in histograms in Figs. 3, 4 and 5. As per figures, the proposed 1-level nested stacking ensemble performed much better than traditional classification techniques. The best results achieved by RF: [SV-(3-Base, 2-Meta)] indicated an accuracy rate of 91.10% at

K = 15 cross-validation which was found to be greater than other created models. Figure 4 demonstrated a remarkable prediction ROC of 95.10 by RF: [SV-(2-Base, 3-Meta)] model with K = 10 cross-validation. Figure 5 explained that the value of F-measure is foremost by RF: [SV-(3-Base, 2-Meta)] i.e. 91.10. To conclude, the combinations of

Table 2 Learners description table

Learners	Description
Random Forests	It is an ensemble technique [16, 37] and widely used supervised learning algorithm which consists of large number of individual decisions trees. Each individual tree is grown by randomly selecting a group of features to determine the split and the most voted class of an instance is selected as the final result [24]
PART	Projective Adaptive Resonance Theory (PART) is a separate & conquer rule learner that use planned set of rules for building partial C4.5 decision tree in each iteration and after comparing new data to each rule, a class of first matching rule get assigned to item and makes the “best” leaf into a rule [23, 43]
JRip	Repeated Incremental Pruning to Produce Error Reduction (RIPPER) implements a propositional rule learner [51] which grows through four stages namely initialization stage, building stage including growing and pruning phase, optimization stage and deletion stage
CART	Classification and Regression Trees (CART) use Gini Index as selection measure to find best attribute for splitting, thereby, resulting in the formation of Binary Decision Tree. The splitting step in CART is greedy [50, 53] but it works well with continuous values and uses a series of dichotomous splits
Logistic	It is a class in Weka Data Mining tool used for multinomial logistic regression model with ridge estimator
REPTree	Reduced Error Pruning Tree REPTree [13] creates numerous trees in different iterations by using logic of regression tree, followed by the selection of the best tree on basis of lowest mean-square-error measure. It is a fast decision tree based on C4.5 algorithm which builds a regression/decision tree using information gain/variance [47] and prunes it using reduced-error pruning. Depending on the type of response variable; if response variable is continuous, regression tree is developed and in case of discrete response variable, classification tree is created
Hoeffding Tree	The Hoeffding Tree is also known as VFDT (Very Fast Decision Tree) that deals with streaming data [14, 21, 34]. It grows incremental based on Hoeffding bound and anytime decision tree induction algorithm which is capable of learning from large data streams assuming that no changes take place in data distribution over the time. It is based on idea that small evidence can be often sufficient to choose an optimal splitting attribute and for decision based on the distribution –independent Hoeffding bound [49]
J48	This algorithm generates decision tree using C4.5 algorithm [1, 2] which employs top-down and greedy search through all possible branches to construct a decision tree where nodes are either decision node or leaf node. The target value of the new instance can be predicted by selecting attribute having highest information gain [32] as leaf node indicates the class value. It is also the successor of ID3 algorithm and its classification process is applied to all the tuples in the database and modelled using the binary tree [12]

Table 3 Selected feature set

S. No	Features	Code	Data Type
1	Maternal Mortality Ratio	MMR	Flag
2	Total number of pregnant women registered for ANC	T_PW_ANC	Integer
3	As regards item 2, number registered within first trimester (within 12 weeks)	T_PW_1TRI	Integer
4	Number of pregnant women who received 3 or more ANC checkups during pregnancy	PW_3ANC	Integer
5	Total number of pregnant women who were provided with 100 IFA tablets	PW_100IFA	Integer
6	Number of new cases of hypertension (BP > 140/90) detected in pregnant women at the institution	N_PW_HYP	Integer
7	Number of pregnant women with anaemia i.e. Hb level < 11 (tested cases)	N_PW_HB11	Integer
8	Number of pregnant women having severe anaemia (Hb < 7) treated at institution	N_PW_HB7	Integer
9	Number of deliveries conducted at home and attended by trained SBA (i.e. doctor or nurse or ANM)	N_HD_SBA	Integer
10	Number of deliveries conducted at home and attended by non-trained SBA (i.e. trained TBA or relatives etc.)	N_HD_NSBA	Integer
11	Deliveries conducted at public institutions (including C-sections)	D_PI_C	Integer
12	As regards item 11, number discharged under 48 h of delivery	N_DIS_48H	Integer
13	Total number of caesarean (C-section) deliveries performed at public facilities i.e. PHC, CHC, SDH, DH and other State owned public institutions	T_CS_PF	Integer
14	Number of still births	N_SB	Integer
15	Number of abortions (spontaneous or induced)	N_ABO	Integer
16	Women getting a postpartum check-up within 48 h after delivery	W_PP_48H	Integer
17	Women getting a postpartum checkup between 48 h and 14 days after delivery	W_PP_14D	Integer

Table 4 Results without stacking

Measures	CART				Random forest				JRip			
	K = 3	K = 5	K = 10	K = 15	K = 3	K = 5	K = 10	K = 15	K = 3	K = 5	K = 10	K = 15
Accuracy	79.52	78.93	79.97	80.56	86.65	86.79	86.50	87.98	80.41	80.41	80.71	81.01
Precision	79.50	78.90	79.90	80.50	86.60	86.80	86.50	88.00	80.80	80.40	80.80	81.00
Recall	79.50	78.90	79.90	80.60	86.60	86.80	86.50	88.00	80.40	80.40	80.80	81.00
F-Measure	79.50	78.90	80.00	80.50	86.60	86.80	86.50	87.90	80.50	80.40	80.70	80.90
ROC	84.00	82.50	83.30	81.80	93.80	94.70	94.40	94.90	84.20	81.90	82.70	83.10

Table 5 1-Level stacking results using JRip as base learner

Measures (JRip)	SV-(2-Base, 2-Meta)			SV-(2-Base, 3-Meta)			SV-(3-Base, 2-Meta)			SV-(3-Base, 3-Meta)							
	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 15				
Accuracy	88.87	88.43	88.28	89.76	88.13	88.28	87.98	88.72	88.43	87.39	88.13	88.43	90.06	87.98	88.28	88.28	89.76
Precision	88.90	88.40	88.30	89.80	88.10	88.30	88.00	88.80	88.50	87.40	88.20	88.50	90.10	88.00	88.30	88.30	89.80
Recall	88.90	88.40	88.30	89.80	88.10	88.30	88.00	88.70	88.40	87.40	88.10	88.40	90.10	88.00	88.30	88.30	89.80
F-Measure	88.80	88.40	88.30	89.70	88.10	88.30	88.00	88.70	88.40	87.40	88.10	88.40	90.10	87.90	88.30	88.30	89.70
ROC	92.90	92.40	92.70	92.90	92.90	92.70	93.40	93.20	92.70	91.50	92.80	92.70	93.80	91.80	93.10	93.60	94.10

Table 6 1-Level stacking results using random forest as base learner

Measures (RF)	SV-(2-Base, 2-Meta)			SV-(2-Base, 3-Meta)			SV-(3-Base, 2-Meta)			SV-(3-Base, 3-Meta)							
	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 15				
Accuracy	88.87	90.20	89.76	90.80	88.28	89.32	88.72	90.80	89.91	88.72	89.91	90.06	91.10	88.72	89.91	89.32	90.65
Precision	88.90	90.20	89.70	90.80	88.30	89.30	88.70	90.80	89.90	88.70	89.90	90.10	91.10	88.70	89.90	89.30	90.70
Recall	88.90	90.20	89.80	90.80	88.30	89.30	88.70	90.80	89.90	88.70	89.90	90.10	91.10	88.70	89.90	89.30	90.70
F-Measure	88.80	90.20	89.80	90.80	88.20	89.30	88.70	90.80	89.90	88.70	89.90	90.10	91.10	88.70	89.90	89.30	90.60
ROC	93.10	94.50	94.70	94.20	94.20	94.80	95.10	94.30	94.20	93.30	94.20	94.70	93.50	93.70	94.90	94.70	94.30

Table 7 1-Level stacking results using CART as base learner

Measures (CART)	SV-(2-Base, 2-Meta)			SV-(2-Base, 3-Meta)			SV-(3-Base, 2-Meta)			SV-(3-Base, 3-Meta)			
	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 3	K = 5	K = 10	K = 15
Accuracy	87.83	89.02	89.32	87.24	88.43	89.32	86.94	88.13	89.61	87.24	89.02	89.32	90.65
Precision	87.80	89.00	89.30	87.20	88.40	89.30	87.00	88.20	89.60	87.20	89.00	89.30	90.60
Recall	87.80	89.00	89.30	87.20	88.40	89.30	86.90	88.10	89.60	87.20	89.00	89.30	90.70
F-Measure	87.80	89.00	89.30	87.20	88.40	89.30	87.00	88.10	89.60	87.20	89.00	89.30	90.60
ROC	93.10	92.90	93.40	93.50	93.90	93.80	91.60	93.00	93.20	92.70	93.80	93.50	94.00

Random Forest exhibited outstanding accuracy, ROC and F-measure with variation of Base and Meta Learners. Further, to choose the finest ensemble model among all these combinations of Random Forest, the time comparison was also calculated which is discussed in Sect. 5.4.

5.4 Computational time comparison

The overall time taken comparisons to train the model are shown in Table 8 and RF: [SV-(3-Base, 2-Meta)] was observed to perform better among others. Therefore, RF: [SV-(3-Base, 2-Meta)] was considered as the best performance model among all the models used in this research work in terms of accuracy, F-measure and training time.

6 Conclusion

In all walks of life, people gradually focus on collecting and utilizing data. The experts from various fields now-a-days are investigating the dataset by applying data mining algorithms for well-being of the society. In this study, the aim was to ameliorate the current status of the deaths of women by undertaking prediction procedures via Base and Meta Learners. The well-known algorithms of data mining are considerably important in medical field. Thus, through this paper, a novel attempt was made to investigate maternal dataset by way of more effective nested stacking technique. There is limited research on India's Maternal Mortality Ratio (MMR) when stacking techniques were applied, therefore, researchers endeavored to find various combinations of Base and Meta Learners with the intent to examine which learners are preminent for making predictions in maternal deaths. The reliability of the system was evaluated by computing the accuracy and ROC area of algorithms without stacking and with stacking ensemble. In case of Base Learners, Random Forest attained accuracy of 87.98% among other learners. However, it was observed that in case of proposed nested stacking, accuracy of 91.10% was achieved from combination RF: [SV-(3-Base, 2-Meta)] and among three Base Learners, Random Forest achieved noteworthy prediction accuracy. Furthermore, this accuracy was obtained by varying K-fold cross validation and by working on best feature subset obtained after feature selection process. Therefore, Random Forest showed its potential in terms of efficiency and effectiveness based on accuracy, F-measure and training time. Thus, from this study, it has been concluded that Random Forest when using as Base Learner with the combination of two other Base Learners viz. Logistic and PART and two Meta Learners viz. Hoeffding and REPTree was best suited model for classifying the Maternal Health data into High MMR and Low MMR. In future, the same model shall also

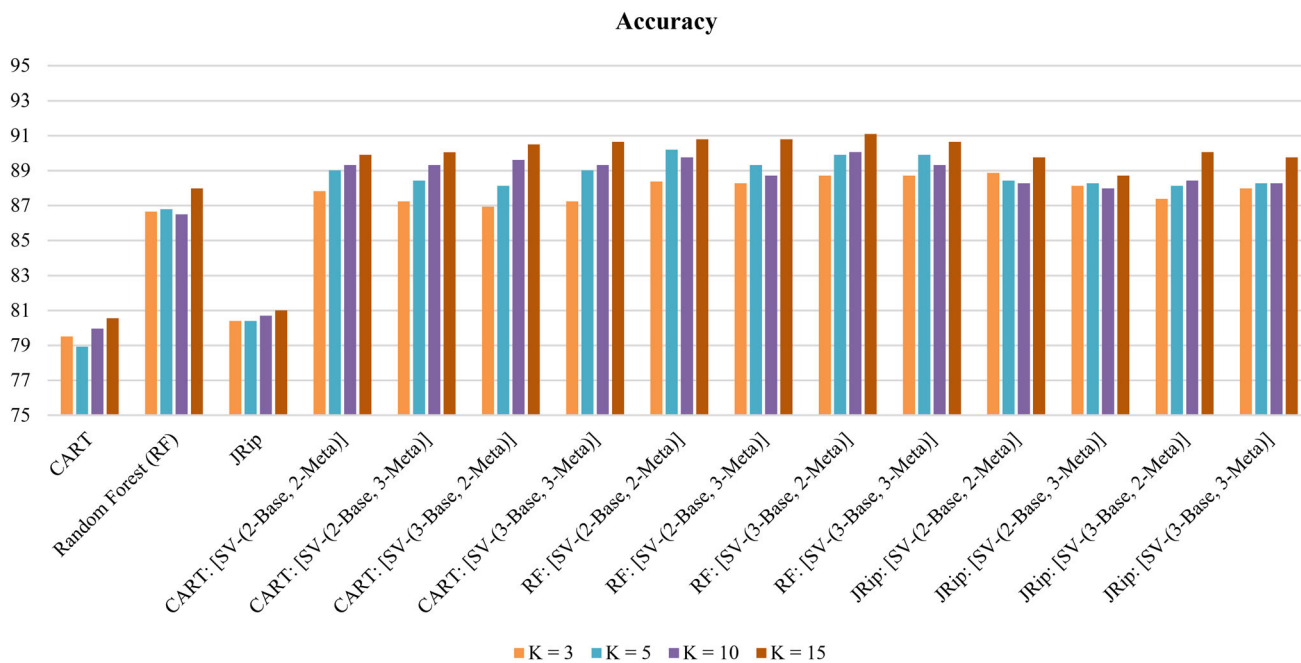


Fig. 3 Overall accuracy comparison

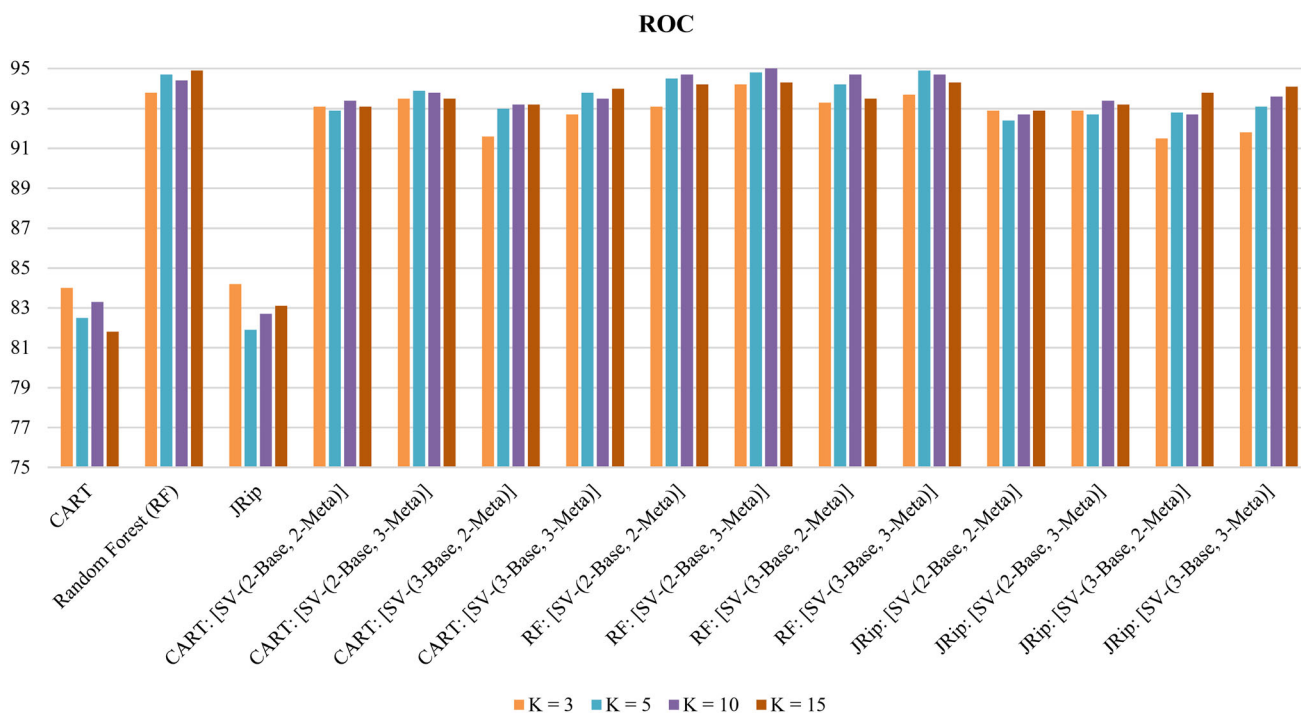


Fig. 4 Overall ROC comparison

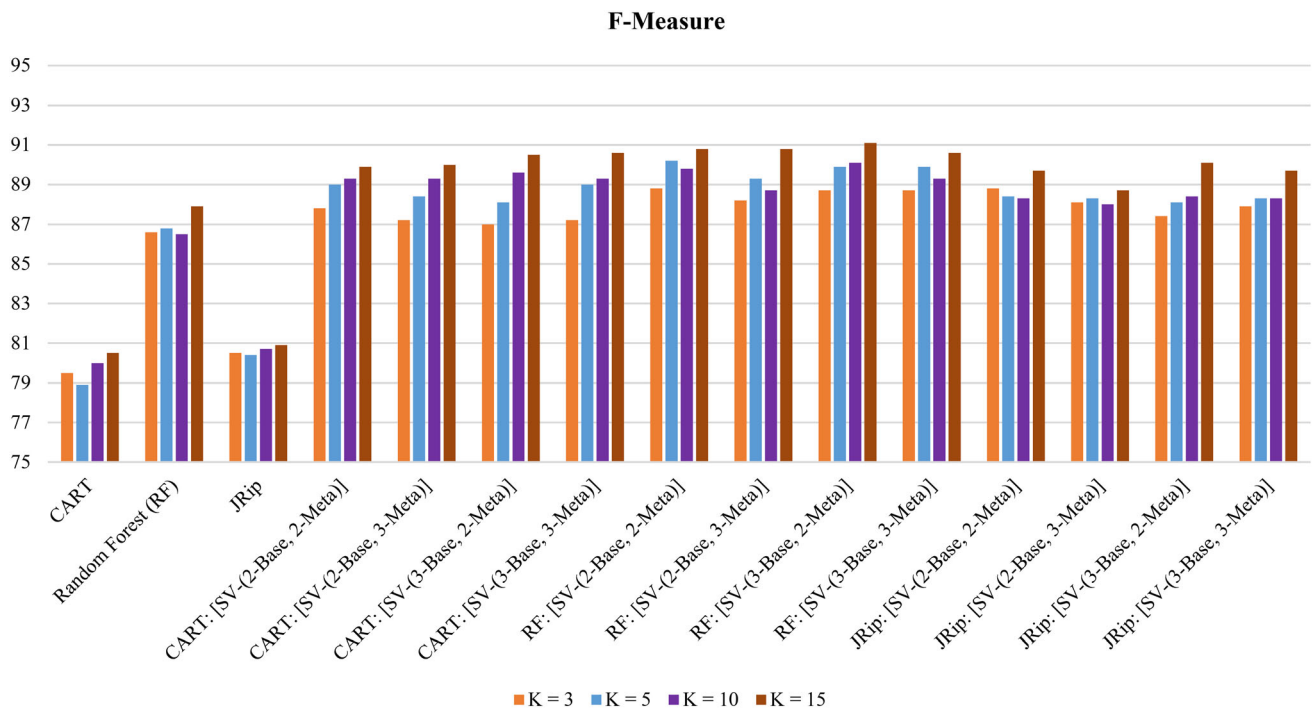


Fig. 5 Overall F-measure comparison

Table 8 Overall time comparison

Methods		Time (in s)			
		K = 3	K = 5	K = 10	K = 15
JRip	Simple	0.11	0.11	0.11	0.11
	SV-(2-Base, 2-Meta)	1.46	1.55	1.58	3.33
	SV-(2-Base, 3-Meta)	1.64	1.74	1.50	2.20
	SV-(3-Base, 2-Meta)	1.85	1.98	2.04	2.93
	SV-(3-Base, 3-Meta)	1.91	2.09	2.00	1.88
Random Forest	Simple	0.24	0.25	0.22	0.22
	SV-(2-Base, 2-Meta)	2.46	2.55	3.32	2.54
	SV-(2-Base, 3-Meta)	2.34	2.33	2.46	2.62
	SV-(3-Base, 2-Meta)	2.86	3.62	2.83	2.48
	SV-(3-Base, 3-Meta)	3.08	3.25	3.51	3.52
CART	Simple	0.14	0.14	0.13	0.14
	SV-(2-Base, 2-Meta)	1.52	1.68	1.71	1.74
	SV-(2-Base, 3-Meta)	1.72	1.71	2.47	2.69
	SV-(3-Base, 2-Meta)	1.87	2.77	2.91	2.87
	SV-(3-Base, 3-Meta)	1.75	2.93	2.15	2.73

be implemented on other datasets related to healthcare. Moreover, the current work shall also be extended to n-level stacking.

Acknowledgements The authors are thankful to Health Management Information System (HMIS), Ministry of Health & Family Welfare (MoHFW), Government of India (GoI), India for providing free online access of the Healthcare data.

Funding No financial support was received for this work.

Compliance with ethical standards

Conflict of interest There is no conflict of interest in this work.

References

- Abdar M et al (2017a) Educational data mining based on multi-objective weighted voting ensemble classifier. *Int Conf Comput Sci Comput Intell*. <https://doi.org/10.1109/CSCI.2017.192>
- Abdar M et al (2017b) Performance analysis of classification algorithm on early detection of liver disease. *Expert Syst Appl* 67:239–251. <https://doi.org/10.1016/j.eswa.2016.08.065>
- Abdar M et al (2018) A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recogn Lett*. <https://doi.org/10.1016/j.patrec.2018.11.004>
- Adekitan AI et al (2019) Data mining approach for predicting the daily Internet data traffic of a smart university. *J Big Data*. <https://doi.org/10.1186/s40537-019-0176-5>
- Amin MS et al (2019) Identification of significant features and data mining techniques in predicting heart disease. *Telematics Inform* 36:82–93. <https://doi.org/10.1016/j.tele.2018.11.007>
- Ammar A, Ahmed Q, Maheswari D (2019) An enhanced ensemble classifier for telecom churn prediction using cost based uplift modelling. *Int J inf Tecnol* 11(2):381–391. <https://doi.org/10.1007/s41870-018-0248-3>
- Aruldas K, Kant A, Mohanan PS (2017) Care-seeking behaviors for maternal and newborn illnesses among self-help group households in Uttar Pradesh, India. *J Health Popul Nutr*. <https://doi.org/10.1186/s41043-017-0121-1>
- Bairagi V (2018) EEG signal analysis for early diagnosis of Alzheimer disease using spectral and wavelet based features. *Int J Inf Tecnol* 10(3):403–412. <https://doi.org/10.1007/s41870-018-0165-5>
- Bashir S, Qamar U, Khan FH (2015a) A multicriteria weighted vote-based classifier ensemble for heart disease prediction. *Comput Intell*. <https://doi.org/10.1111/coin.12070>
- Bashir S, Qamar U, Khan FH (2015b) BagMOOV: a novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. *Austr Coll Phys Scientists Engineers Med*. <https://doi.org/10.1007/s13246-015-0337-6>
- Bashir S, Qamar U, Khan FH (2016) IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework. *J Biomed Inform* 59:185–200. <https://doi.org/10.1016/j.jbi.2015.12.001>
- Bansal D et al (2018) Comparative analysis of various machine learning algorithms for detecting dementia. In: *International conference on computational intelligence and data science (ICCIDS 2018)*, pp 1497–1502. <https://doi.org/10.1016/j.procs.2018.05.102>
- Boodhun N, Jayabalan M (2018) Risk prediction in life insurance industry using supervised learning algorithms. *Complex Intell Syst*. <https://doi.org/10.1007/s40747-018-0072-1>
- Boidol J et al (2015) Probabilistic Hoeffding trees. *Ind Conf Data Min*. https://doi.org/10.1007/978-3-319-20910-4_8
- Bonet I et al (2011) Ensemble of classifiers based on hard instances. *Mexican Conf Pattern Recogn*. https://doi.org/10.1007/978-3-642-21587-2_8
- Bowes D et al (2017) Software defect prediction: do different classifiers find the same defects? *Softw Qual J*. <https://doi.org/10.1007/s11219-016-9353-3>
- Cawley GC, Talbot NLC (2008) Efficient approximate leave-one-out cross-validation for kernel logistic regression. *Mach Learn* 71:243–264. <https://doi.org/10.1007/s10994-008-5055-9>
- Chakraborti S et al (2018) A machine learning based method to detect epilepsy. *Int J Inf Tecnol* 10(3):257–263. <https://doi.org/10.1007/s41870-018-0088-1>
- Cong J et al (2017) A selective ensemble classification method combining mammography images with ultrasound images for breast cancer diagnosis. *Comput Math Methods Med*. <https://doi.org/10.1155/2017/4896386>
- Das R, Sengur A (2010) Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Syst Appl* 37:5110–5115. <https://doi.org/10.1016/j.eswa.2009.12.085>
- Domingos P, Hulten G (2000) Mining high-speed data streams. *Int Conf Knowl Discov Data Min*. <https://doi.org/10.1145/347090.347107>
- Esener II, Ergin S, Yuksel T (2017) A new feature ensemble with a multistage classification scheme for breast cancer diagnosis. *J Healthcare Eng*. <https://doi.org/10.1155/2017/3895164>
- Frank E et al (1998) Generating accurate rule sets without global optimization. In: *Fifth International conference on machine learning*, pp 144–151
- Gomes HM et al (2017) Adaptive random forests for evolving data stream classification. *Mach Learn* 106:1469–1495. <https://doi.org/10.1007/s10994-017-5642-8>
- GBD (2015) Eastern Mediterranean Region Maternal Mortality Collaborators, Maternal mortality and morbidity burden in the Eastern Mediterranean Region: findings from the Global Burden of Disease 2015 study. *Int J Public Health*. <https://doi.org/10.1007/s00038-017-1004-3>
- Ghosh M, Sanyal G (2018) An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. *J Big Data*. <https://doi.org/10.1186/s40537-018-0152-5>
- Gupta M et al (2017) Utilization of intergovernmental funds to implement maternal and child health plans of a multi-strategy community intervention in Haryana, North India: A Retrospective Assessment. *PharmacoEconomics*. <https://doi.org/10.1007/s41669-017-0026-3>
- Hamal M et al (2018) How do accountability problems lead to maternal health inequities? A review of qualitative literature from Indian public sector. *Public Health Rev*. <https://doi.org/10.1186/s40985-018-0081-z>
- Hasan M et al (2019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Thing*. <https://doi.org/10.1016/j.iot.2019.100059>
- Jothi N et al (2015) Data mining in healthcare—a review, the third information systems international conference. *Procedia Comput Sci* 72:306–313. <https://doi.org/10.1016/j.procs.2015.12.145>
- Kabir MF, Ludwig SA (2019) Enhancing the performance of classification using super learning. *Data-Enabled Discov Appl*. <https://doi.org/10.1007/s41688-019-0030-0>
- Kiranmai SA, Laxmi AJ (2018) Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy. *Protect Control Modern Power Syst*. <https://doi.org/10.1186/s41601-018-0103-3>
- Kour P et al (2019) Classification of maternal healthcare data using naïve bayes. *Int J Comput Sci Eng* 7(3):388–394
- Kourtellis N et al (2016) VHT: vertical Hoeffding tree. *IEEE Int Conf Big Data (Big Data)*. <https://doi.org/10.1109/BigData.2016.7840687>
- Kumari D, Kilam S, Nath P, Swetapadma A (2018) Prediction of alcohol abused individuals using artificial neural network. *Int J Inf Tecnol* 10(2):233–237. <https://doi.org/10.1007/s41870-018-0094-3>

36. Large J, Lines J, Bagnall A (2019) A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates. *Data Min Knowl Disc.* <https://doi.org/10.1007/s10618-019-00638-y>
37. Mantas CJ et al (2018) A comparison of random forest-based algorithms: random credal random forest versus oblique random forest. *Soft Comput.* <https://doi.org/10.1007/s00500-018-3628-5>
38. Ma H et al (2018) Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on non-alcoholic fatty liver disease in China. *Hindawi BioMed Res Int.* <https://doi.org/10.1155/2018/4304376>
39. Mehdiyev N et al (2016) Evaluating forecasting methods by considering different accuracy measures. *Procedia Comput Sci* 95:264–271. <https://doi.org/10.1016/j.procs.2016.09.332>
40. Moreira MWL et al (2018) Predicting neonatal condition at birth through ensemble learning methods in pregnancy care. In: *Proceedings of the 18th Brazilian symposium on computing applied to healthcare*, Natal, 2018. <https://doi.org/10.5753/sbcas.2018.3671>
41. Nagi S, Bhattacharyya DK (2013) Classification of microarray cancer data using ensemble approach, network modeling analysis in healthcare informatics. *Bioinformatics.* <https://doi.org/10.1007/s13721-013-0034-x>
42. Nguyen TT et al (2018) Combining heterogeneous classifiers via granular prototypes. *Appl Soft Comput* 73:795–815. <https://doi.org/10.1016/j.asoc.2018.09.021>
43. Parsania VS et al (2014) Applying Naïve Bayes, BayesNet, PART, JRip and OneR algorithms on hypothyroid database for comparative analysis. *Int J Darshan Inst Eng Res Emerg Technol* 3:1
44. Pes B (2019) Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-019-04082-3>
45. Rahman MdJU et al (2019) Ensemble of multiple models for robust intelligent heart disease prediction system. *Int Conf Electr Eng Inf Commun Technol.* <https://doi.org/10.1109/CEEICT.2018.8628152>
46. Shastri S, Mansotra V (2019) KDD-based decision making: a conceptual framework model for maternal health and child immunization databases. *Adv Comput Commun Comput Sci Adv Intell Syst Comput.* https://doi.org/10.1007/978-981-13-6861-5_21
47. Snousy MBA et al (2011) Suite of decision tree-based classification algorithms on cancer gene expression data. *Egypt Informatics* 12:73–82. <https://doi.org/10.1016/j.eij.2011.04.003>
48. Sourabh VM (2019) Data mining probabilistic classifiers for extracting knowledge from maternal health datasets. *Int J Innovat Technol Explor Eng* 9(2):2769–2776. <https://doi.org/10.35940/ijitee.B6633.129219>
49. Srimani PK, Patil MM (2015) Performance analysis of Hoeffding trees in data streams by using massive online analysis framework. *International Journal of Data Mining. Model Manag* 7:4. <https://doi.org/10.1504/IJDM.2015.073865>
50. Venkatasubramaniam A et al (2017) Decision trees in epidemiological research. *Emerg Themes Epidemiol.* <https://doi.org/10.1186/s12982-017-0064-4>
51. Weitschek E, Fiscon G, Felici G (2014) Supervised DNA Barcodes species classification: analysis, comparisons and results. *Bio Data Min.* <https://doi.org/10.1186/1756-0381-7-4>
52. Wu H et al (2018) Type 2 diabetes mellitus prediction model based on data mining. *Informatics Med Unlocked* 10:100–107. <https://doi.org/10.1016/j.imu.2017.12.006>
53. Zimmerman RK, Jackson ML, Gaglani M (2016) Classification and regression tree (CART) analysis to predict influenza in primary care patients. *BMC Infect Dis.* <https://doi.org/10.1186/s12879-016-1839-x>