

Prediction Of Diabetes And Cholesterol Diseases Based On Ensemble Learning Techniques

R. Karthikeyan, Dr. P. Geetha, Dr. E. Ramaraj

Abstract: In the field of Artificial Intelligence, the ensemble learning methodology is a multiple machine learning model algorithm used to acquire a better performance for a disease prediction. Disease prediction through Ensemble learning can be done by using two different datasets such as Diabetes and Cholesterol. The rise of Non-Communicable Diseases (NCD) like Cardiovascular Disease (CVD) and Diabetes in the world is increasing day by day drastically in a very high manner which leads to deaths in several countries. The main aim of ensemble learning is to process datasets on multiple models and then combine all the predictions to get ensemble learning. The ensemble learning algorithm uses five algorithms Adaboost, Random Forest or Randomization, Bagging or Bootstrap Aggregation, Voting, and Stacking. Datasets inputs are supplied to five Ensemble learning algorithms separately to generate five models and combine all the five models to get a final prediction. For combining the classifier model the weighting methods are best suited for the prediction of diseases. Nowadays NCD can affect anyone due to various parameters like irregular food habits, use of tobacco, alcohol consumption, and high glucose levels, physical inactivity, high blood pressure, and high cholesterol, obesity. Cholesterol is one of the major factors for CVD. When the human body could not produce enough insulin or the body cannot utilize the insulin produces then it leads to chronic conditions or Diabetes. Results suggest that the ensemble learning algorithm performs better than an individual algorithm. Results also suggest that key parameters play a major role in the prediction of diseases in diabetes and cholesterol.

Keywords: Artificial Intelligence, Ensemble Learning, Non-Communicable Diseases, Cardiovascular Disease (CVD), AdaboostM1, Random Forest, Bootstrap Aggregation, Voting, Stacking (key words)

1 INTRODUCTION

ENSEMBLE LEARNING [EL] is used when the expected outcome is not known exactly and it depends on various model outputs are collectively combined to form a supermodel. EL is used in the field of classification and prediction to improve model performance. The need for EL in machine learning is it converts a weak classifier into a strong classifier by combining all base learners to get a new prediction based on ensemble learning combining techniques. In EL a set of hypotheses are combined to form a new prediction whereas ordinary ml algorithms will produce a single hypothesis. The rise of Non-Communicable Disease or cardiovascular disease (CVD) in the world is increasing day by day drastically in a very high manner which leads to death in several countries. A survey from an international study published by Elsevier in the journal of the American college of cardiology states that India's death rate of CVD rose by around 34 percent from 155.7 to 209.1 deaths per one lakh population and during the same period survey in America it is declined significantly due to awareness of CVD. Nowadays CVD can affect by any person due to various parameters like irregular food habits, use of tobacco, alcohol consumption, and high glucose levels, physical inactivity, high blood pressure, and high cholesterol, obesity. Even doctors and nurses are more prone to CVD due to work pressure and irregular food habits. Cholesterol is a waxy, fatty substance. Some cells in the liver produce it and release it into the bloodstream. Cholesterol is a type of body fat or lipid. A serum cholesterol level is a measurement of certain elements in the blood, including the amount of high- and low-density lipoprotein cholesterol (HDL and LDL) in a person's blood.

Low-density lipoprotein (LDL) cholesterol is bad cholesterol, while high-density lipoprotein (HDL) cholesterol is considered as good cholesterol, triglycerides, which are a type of fat bundled with cholesterol. Serum cholesterol levels can indicate their risk for conditions such as heart disease [15]. Cardiovascular disease (CVD) is a big reason for morbidity and mortality in the current living style. Identification of Cardiovascular disease is an important but complex task that needs to be performed very minutely, efficiently and the correct automation would be very desirable. Every human being can not be equally skillful and so as doctors. All doctors cannot be equally skilled in every subspecialty and many places, we don't have skilled and specialist doctors available easily [19]. Ensemble learning is a classification method used to reduce bias and variance in a model for better accuracy [20]. Diabetes and Cholesterol diseases are predicted through ensemble-based techniques.

1.1 Diabetes: Diabetes occurs due to high blood glucose levels and the human body could not produce the required level of insulin or cannot utilize the insulin it produces [2]. It is also called as Diabetes mellitus, which can lead to various complications such as heart disease, stroke, eye, and foot-related problems [1]. The blood sugar levels in diagnosing diabetes in various forms like Normal Levels, Pre-Diabetes, and Diabetes. The various tests are available for diabetes such as Plasma Glucose Test, Fasting, and 2-hour post-prandial Test, Non-Diabetic Check in Before Meals or After Meals Test, Type-1 Diabetes, Type-2 Diabetes, HBA1C Test.

1.1.1 Glucose Category Levels: The Glucose level of a human informs the major factor for Diabetic disease prediction. They are divided into three levels as normal. Above normal, high or Diabetic.

1: Normal (Or) Desirable - Refers to Below 42 mmol/mol (6.0%).

2: Above normal (Or) Borderline High (Or) Pre-Diabetic - Refers to 42 to 47 mmol/mol (6.0 to 6.4%).

3: Well Above Normal (Or) High (Or) Diabetic -Refers to 48 mmol/mol (6.5% or over).

- R. Karthikeyan, Ph.D. Research Scholar, Dept. of Computer Science, Alagappa University, Karaikudi- India, karthikeyan.r@srmasc.ac.in.
- Dr. P. Geetha, Associate Professor, Dept. of Computer Science, Dr. Umayal Ramanathan College for Women, Karaikudi- India, geetha.ganesan@gmail.com.
- Dr. E. Ramaraj, Professor & Head, Dept. of Computer Science, Alagappa University, Karaikudi- India, eramaraj@rediffmail.com

1.2 Cholesterol: When a serum cholesterol level results in high may lead to risk factors like an increase in triglyceride levels, obesity, high blood pressure, coronary artery disease, diabetes, stroke, heart attack, cardiovascular disease, organ or tissue damage or family history of heart disease. Due to a sedentary lifestyle and poor diet habits, cholesterol risk factors will arise. If any one or both the parents are prone to high cholesterol, then he/she might be affected by high cholesterol. Total serum cholesterol= HDL level+ LDL level+ 20 % triglyceride level present in a blood sample. The Normal range for a person's serum cholesterol level in the serum cholesterol limit is less than 200, LDL Cholesterol limit is less than 130, HDL Cholesterol limit is 55 for women and 45 for men, Triglycerides limit is less than 150.

1.2.1 Cholesterol Category Levels: The Cholesterol level of humans informs the availability of Cholesterol in blood. They are divided into three levels as normal, above normal, high.

1: Normal (Or) Desirable - Refers to Less than 200mg/DL.

2: Above Normal (Or) Borderline High - Refers to 200-239 mg/dL.

3: Well Above Normal (Or) High - Refers to 240 mg/dL and Above.

1.3 Dataset Description

The dataset for diabetic prediction is from UCI Machine learning repository and for Cholesterol the dataset is from kaggle. The link is provided below as <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>. The various attributes are id, age, gender, height, weight, BP in systolic and diastolic, cholesterol, glucose, smoke, alcohol intake, physical activity, class variable cardio present or not.

2 LITERATURE REVIEW

Authors Divya Zion, *B.K.Tripathy [4] focuses on risk factors and datasets of Non-Communicable Diseases that are 'slow-moving catastrophe'. Predictive analysis-based data mining techniques are used to prevent the risk factors of NCD. Also, the author discusses the Risk Factor of NCD Use of Tobacco, Physical Inactivity, Alcohol Consumption, and Unhealthy Diet. Methodologies used are Naïve Bayes classification, Discriminant Analysis, and decision tree learning methods to predict NCDs. Authors Zhiwen Yu et al. [5],[11] explores the hybrid incremental ensemble learning approach to handle noisy data's by bagging algorithm and final results are obtained through weighted volume. Authors Shaohua Wan, Hua Yang [6] focuses on four ensemble methods to handle classification and regression problems. Also, authors show the various experimental results for thirty datasets using UCI. Authors X. Liu et al.[7] focused on an ensemble transfer learning framework to improve the classification accuracy using weighted resampling method for transfer learning is proposed, which is named TrResampling And authors uses three classic machine learning algorithms, namely, naïve Bayes, decision tree, and SVM, are used as the base learners of TrResampling, where the base learner with the best performance is chosen for transfer learning. The authors Aayushi Verma, Shikha Mehta [8] describes three ensembles "BBS method" which stands for Bagging, Boosting and Stacking with appropriate base classifiers for the classification of the five UCI datasets taken from the field of Bioinformatics. Bagging is performed using Random Tree classifier, Boosting is done using AdaBoost classifier, along with the Decision

Stump as a base classifier, and Stacking is performed using three learning algorithms which are J48, Naive Bayes and Random Forest respectively. The Authors Adil Husain, Muneeb H. Khan [9] focuses on An Ensemble model using a majority voting technique using unweighted prediction on various machine learning models. The authors[10] focus on non-generative methods, the basic classifiers are joined by a combination procedure that depends on its adaptability to input observations and the needs of the output provided by individual learning systems, that is, the type of combination depends on the type of output. This procedure can be improved by assigning a weight to each classifier so that the behavior of the combined classifier in the training set is optimized. Authors Heitor Murilo Gomes et al.,[12] suggest ensemble based methods best suited for data stream classification problem and also he compared ensemble learning with sixty algorithms. The author [13] suggests multiple classifier systems growth in recent years in machine learning for improving accuracy and provides solutions to various machine related problems.

3 EXISTING SYSTEM

Prevention and diagnosis of NCDs approach to predict and diagnose NCDs using a machine learning technique to enhance the quality of treatment. Also, the author suggests risk factors include such as tobacco usage, alcohol usage, lack of physical activity, unhealthy diet, obesity, stress, and environmental factors lead to the major disease burden of NCDs [1]. The main four diseases discussed here are diabetes mellitus: hypertension, cardiovascular diseases: diabetes very dangerous to people. The prediction and diagnosis of non-communicable disease results have shown increases the predictive accuracy for medical datasets by using feature selection and classification. Various machine learning algorithms with diabetes prediction can be achieved through voting based ensemble techniques. Also, the authors apply the majority voting classifier for KNN, Decision Tree, SVM, and Random Forest. The prediction made by all these classifiers on a test case is voted and the prediction with the highest votes is considered as a final prediction [2]. The research using strong data pre-processing techniques like the removal of noisy data, removal of missing data, filling default values if applicable and classification of attributes for prediction and decision making at different levels. They compared individual results of Support Vector Machine, Gradient Boosting, Random Forest, Naive Bayes classifier and logistic regression on the dataset taken in a region to present an accurate model of predicting cardiovascular disease.

4 METHODOLOGY

Ensemble Learning is a machine learning algorithm works on multiple models to get higher performance and the main advantages of EL are minimize the factors like variance, noise, and bias what other learning models lack in? By using these three factors will improve machine learning model accuracies.

4.1 Proposed Architecture

The proposed architecture in figure 1 shows the working of EL with five algorithms such as AdaboostM1, Random Forest, Bootstrap Aggregation, Voting, and

Stacking.

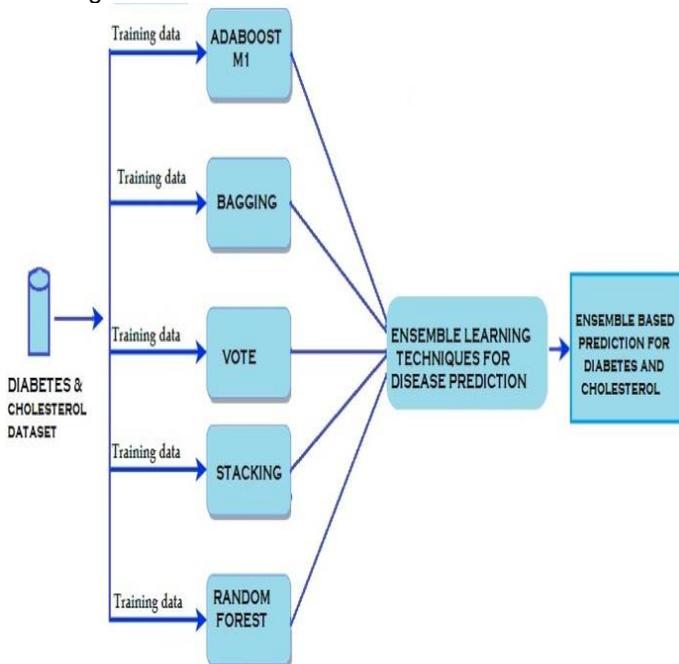


Figure. 1. Ensemble learning methodology for disease prediction in diabetes and cholesterol

Initially diabetes and cholesterol datasets are divided into five subsets and each dataset is trained with a separate model and finally combined using a combinator for better accuracy for diabetes and cholesterol prediction. The advantage of comparing different models in EL will make an informed decision based on the learning model to improve the overall performance. A complex or big problem can be solved by dividing the problem into subproblems and combine the results to get an accurate result like divide and conquer. The power of EL is that it suited for a large or a tiny dataset ie. When a dataset is large it is partitioned into subsets, train with separate classifiers, and then combine using perfect combination rules like voting, averaging, weighted averaging. If the dataset is small then the EL methodology Bootstrap can be used to test with random samples. EL algorithms will ensure better accuracy (low error), higher consistency (avoiding over fitting), and reduced Bias and variance errors.

4.2 Ensemble Learning Methodology

The EL methodology has five algorithms four Meta-learning classifier algorithms like AdaBoostM1, Bagging, stacking and Vote, Random Forest or Randomization. The proposed ensemble approach works as follows.

1. Diabetes and cholesterol dataset is partitioned into five disjoint subsets.
2. Each dataset is trained with every model s such as Bagging, Random Forest, AdaboostM1, Vote, Stacking and results are combined to get higher accuracy.
3. The ensemble-based prediction techniques such as average and weighted average are calculated for every dataset.
4. Best accuracies are considered from average and weighted average for disease prediction in diabetes and cholesterol.

We implemented the ensemble learning disease prediction using WEKA tool and finally, EL based classifiers get higher accuracy in disease prediction than a single classifier model.

4.3 Ensemble Classifiers

4.3.1 AdaboostM1: Adaptive Boosting also called adaboostM1 is the boosting algorithm developed for binary classification and decision tree-related problems. AdaBoostM1 is mainly used to boost the performance of an algorithm. Normally boosting combines the output of various weak classifiers (such as decision tree) to produce a strong classifier. Boosting and bagging algorithms are considered to be homogeneous and they use a single base learner algorithm. To reduce the bias in a model Boosting is mainly used [16]. Boosting is an ensemble technique to build a strong classifier from weak classifiers. Initially, a model is trained with training data and the outputs are passed as input to the second model. It is also called Adaptive Boosting which is used to boost the nominal classifier; it is a type of a sequential EL algorithm to convert a weak classifier to a strong classifier by using boosting techniques. It reduces the bias errors and the base learners are generated sequentially. The boosting model works well with the training dataset. The adaboostm1 generates a weak learner at each stage and updates the weights of training data; at last, it combines all weak learners to form a strong learner [6].

4.3.2 Bagging: It is Also called (Bootstrap AGGREGatING) bootstrap aggregating, a simple ensemble learning algorithm in which the training set is supplied to various models and predicted results are aggregated to get a new prediction. Bagging is also called unstable learning algorithms, which is a small change in the dataset that leads to a major change in results. It works well when every model makes the prediction, based on voting the best classifier is chosen. Here the base learners are generated in parallel in nature and collections of multiple models decrease variance. Random forest is one of the best examples of bagging and mainly used for reducing the variance of a classifier. Bagging algorithm works on different training datasets and the predictions are combined by either voting or average method[6]. To reduce the variance in a model Bagging is used and the best example for bagging is the Random Forest algorithm.

4.3.3 Stacking: This is ensemble model used for combining several classifiers to improve the prediction of the model by adding a multiple base classifier at level 0 and a single Meta classifier in level 1. Stacking is considered to be heterogeneous since it uses various kinds of a base learner to improve the performance of the algorithm. Stacking is achieved by combining two or more previous models.

4.3.4 Vote: This is used for combining several classifiers to improve the prediction of the model by maximum voting. Normally voting ensemble model has low accuracy while comparing with other Meta classifiers, but when base classifier j48 is compared with dataset its results in high accuracy.

4.3.5 Random Forest: It is used for constructing a forest of random trees. Random forest are a combination of tree predictors. It is unexcelled in accuracy among current algorithms. it runs efficiently on large databases. it can handle thousands of input variables without variable deletion[6].

4.4 Evaluating the performance of Disease Prediction by Ensemble Learning Techniques-Weighted Average

The weighted average formula is used to calculate the average value of a particular set of numbers with different levels of relevance. The relevance of each number is called its weight. The weights should be represented as a percentage of the total relevancy. Therefore, all weights should be equal to 100% or 1[17]. The weights are given as {0.20%, 0.30%, 0.20%, 0.20%, and 0.10%} substituted for each algorithm. The weighted average can be calculated with percentages, every category value must be multiplied by its percentage [18]. While making calculation % must be converted to decimal before multiplying. Then all of these values must be added together [18]. From the above table multiply the accuracy by the weight % given above. For example, the dataset DB_T4 has a weighted average is 91.05%. It can be calculated using the formula below.

$$\text{Weighted Average} = w_1X_1 + w_2X_2 + \dots + w_nX_n$$

Where w = relative weight (%), X = value, for example DB_T4 (153) dataset is expanded below for weighted average

Bagging [92.81*(.20) =18.56] + Random Forest [100*(.30) =30] + AdaboostM1 [86.27*(.20) =17.25] + Vote [88.24*(.20) =17.64] + Stacking [75.82*(.10) =7.582].

By combining all the weights we get 91.05% as weighted average of DB_T4 in diabetes prediction. The weighted average ratio is given in figure 2 below

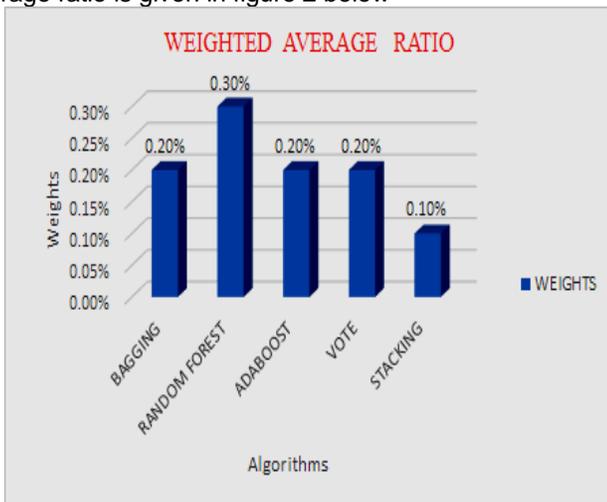


Figure 2 Weighted Average Ratios for Ensemble Learning Algorithms

4.5 Ensemble Learning Methodology for diabetes and cholesterol

4.5.1 Diabetes Disease Prediction Through Ensemble Learning: Diabetes prediction through ensemble learning methodology is given in Table 1 below with five ensemble learners such as AdaBoostM1, Bagging, Stacking, Vote, and Random Forest. Initially, the 768 instances are divided into five subsets and each dataset is tested with separate classifiers using the WEKA tool. First dataset DB_T1 with 153 instances are tested with the five algorithms to get different accuracies. Bagging, Random Forest, AdaboostM1 algorithms score reasonable accuracies whereas the Vote and Stacking algorithms have less accuracy 64.7059 %

compared to the other three algorithms. To increase the accuracy for Stacking the Random Forest in classifier and J48 in the Meta classifier is added to get the accuracy over 100 % and for the Vote algorithm it uses J48 as a base classifier to get the accuracy over 77.12 in Weka.Gui.GenericObjectEditor. Next, the remaining four datasets are compared to get various accuracies are given in the table below. Each time we use various base classifiers and Meta classifiers in the stacking algorithm and for a vote, we use J48 based classifier to get improved accuracy. For examining the ensemble learning techniques we use two methods such as average, weighted average. By using averaging rule all five model accuracies are 87.58%, 82.35%, 83.66%, 88.63%, 83.21% and by using a weighted average the accuracies are 87.58%, 85.29%, 87.45%, 91.05%, 86.79% as shown in table 1 below.

Dataset / Algorithm	DB_T1 (153)	DB_T2 (153)	DB_T3 (153)	DB_T4 (153)	DB_T5 (156)
Bagging	83.01%	83.66%	88.89%	92.81%	85.90%
Random Forest	100%	100%	100%	100%	100%
Adaboostm1	77.78%	73.86%	81.05%	86.27%	79.49%
Vote	77.12%	83.66%	86.27%	88.24%	86.54%
Stacking	100%	70.59%	62.09%	75.82%	64.10%
Average	87.58%	82.35%	83.66%	88.63%	83.21%
Weighted Average	87.58%	85.29%	87.45%	91.05%	86.79%

TABLE I. DIABETES PREDICTION THROUGH ENSEMBLE LEARNING TECHNIQUES

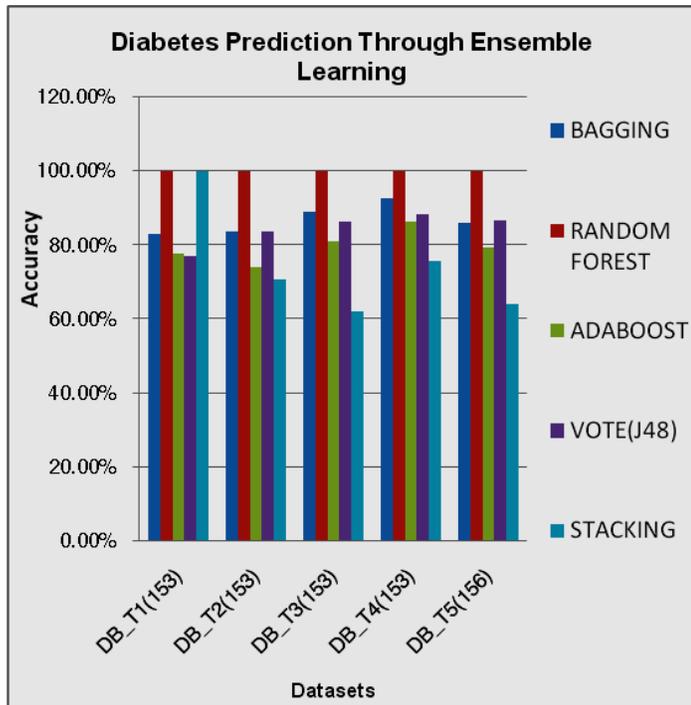


Figure 3 Diabetes Predictions through Ensemble Learning
 From the above figure, 3 show the graph with accuracies with five different datasets in diabetes prediction using ensemble learning methodology. The above table1 shows the average and weighted average values calculated for five different datasets. Compared to average, the weighted average has higher accuracies in all dataset DB_T1 to DB_T5 in the prediction of diabetes is shown in figure 4.

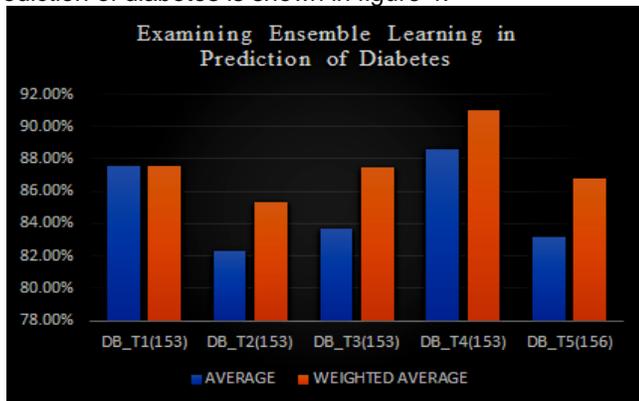


Figure 4 Examining Ensemble Learning for Diabetes

The single or individual classifier performance through WEKA is shown in figure 5 below as stacking has 65.10%, bagging has 75.39%, the random forest has 75%, adaboostm1 has 74.34%, and the vote has 65.10% accuracies. While comparing ensemble learning-based diabetes prediction with individual classifier performance, the EL has higher advantages than a single classifier model.

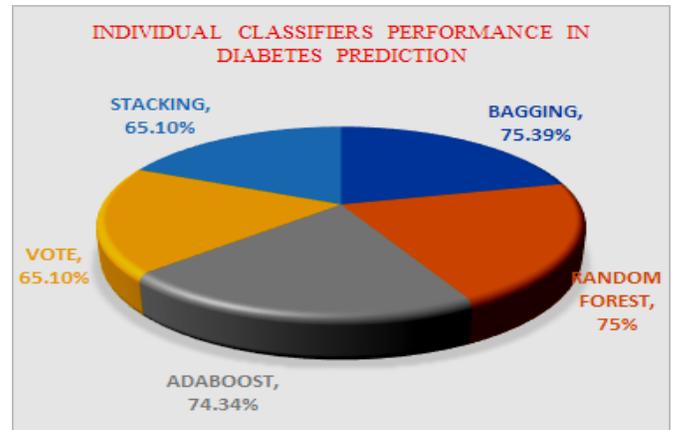


Figure 5 Individual Classifier Performance of Diabetes

4.5.2 Cholesterol Disease Prediction Through Ensemble Learning: Diabetes prediction through ensemble learning methodology is given in Table 2 below with five ensemble learners such as AdaBoostM1, Bagging, Stacking, Vote, and Random Forest. Initially, the 1000 instances are divided into five subsets and each dataset is tested with separate classifiers using the WEKA tool. First dataset CH_T1 with 200 instances are tested with the five algorithms to get different accuracies. Bagging, Random Forest, AdaboostM1 algorithms score reasonable accuracies whereas the Vote and Stacking algorithms have less accuracy 50.00 % compared to the other three algorithms. To increase the accuracy for Stacking the Random Forest in classifier and J48 in the Meta classifier is added to get the accuracy over 98.50% and for Vote algorithm it uses J48 as a base classifier to get the accuracy over 87.50% in Weka.Gui.GenericObjectEditor. Next, the remaining four datasets are compared to get various accuracies are given in the table below. Each time we use various base classifiers and Meta classifiers in the stacking algorithm and for a vote, we use J48 based classifier to get improved accuracy. For examining the ensemble learning techniques we use two methods such as average, weighted average. By using averaging rule all five model accuracies are 87.58%, 82.35%, 83.66%, 88.63%, 83.21% and by using a weighted average the accuracies are 87.58%, 85.29%, 87.45%, 91.05%, 86.79% as shown in table 2 below.

Dataset/ Algorithm	CH_T1 (200)	CH_T2 (201)	CH_T3 (201)	CH_T4 (201)	CH_T5 (201)
Bagging	85.00 %	87.06 %	84.08 %	85.57 %	85.57 %
Random Forest	100%	100%	100%	100%	100%
Adaboostm 1	79.50 %	73.63 %	68.16 %	78.61 %	77.61 %
Vote	87.50 %	88.56 %	77.50 %	83.08 %	88.06 %
Stacking	98.50 %	85.57 %	51.24 %	50.75 %	51.24 %
Averaging	90.10 %	86.97 %	76.20 %	79.60 %	80.50 %

Weighted Average	90.25 %	88.41 %	81.07 %	84.53 %	85.37 %
------------------	---------	---------	---------	---------	---------

TABLE II. CHOLESTEROL PREDICTION THROUGH ENSEMBLE LEARNING TECHNIQUES

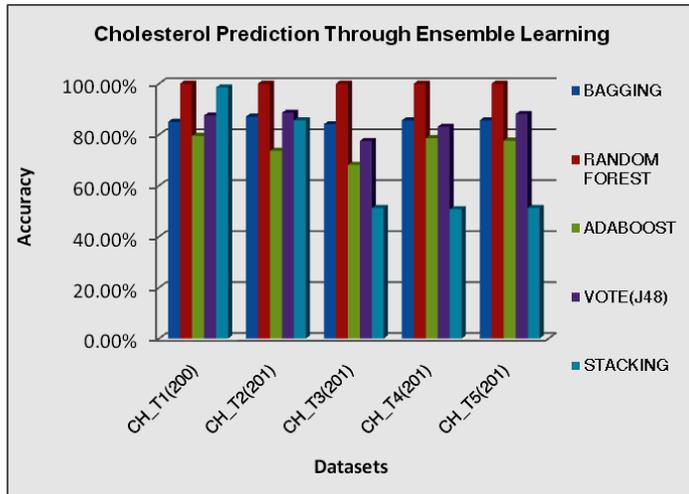


Figure 6 Cholesterol Predictions through Ensemble Learning

From the above figure, 6 show the graph accuracies with five different datasets in cholesterol prediction using ensemble learning methodology. The above table2 shows the average and weighted average values calculated for five different datasets. Compared to average, the weighted average has higher accuracies in all dataset CH_T1 to CH_T5 in the prediction of cholesterol is shown in figure 7.

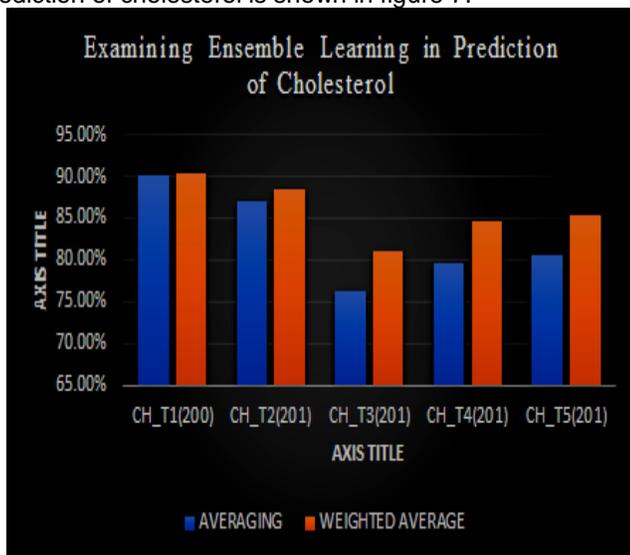


Figure 7 Examining Ensemble Learning for Cholesterol

The single or individual classifier performance through WEKA is shown in figure 7 below as stacking has 50.70%, bagging has 73.70%, the random forest has 74.00%, adaboostm1 has 74.40%, and the vote has 50.70% accuracies. While comparing ensemble learning-based cholesterol prediction with individual classifier performance, the EL has higher advantages than a single classifier model.

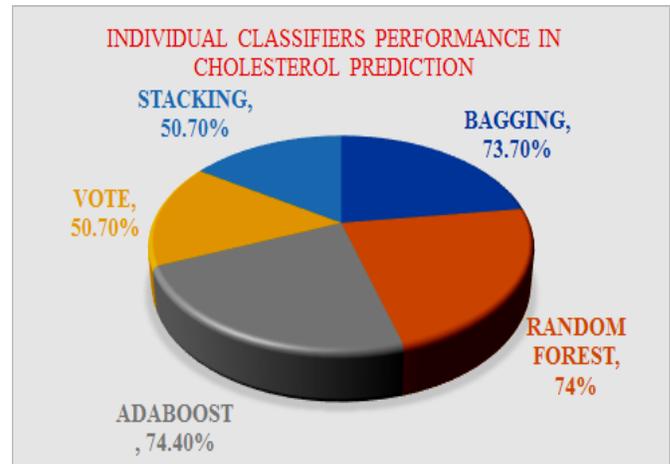


Figure 8 Individual Classifier Performance of Cholesterol

5 RESULTS AND DISCUSSIONS

From the above table [1] shows the diabetic prediction and table [2] shows the cholesterol prediction, both ensemble learning-based disease prediction has higher accuracies than the individual classifier performance in diabetes and cholesterol shown in figure [5],[8] shows diabetic and cholesterol prediction through WEKA J48 classifier that ensemble learning method has higher accuracy over all other methodologies in diabetes and cholesterol disease prediction.

6 CONCLUSION

An ensemble is a supervised learning algorithm that can be trained to make a better prediction [14]. Diabetes and cholesterol datasets are subdivided into five disjoint subsets and apply divide and conquer to get better accuracy with several ensembles learning classifiers. The ensemble learning disease prediction of diabetic and cholesterol is carried out through five ensemble learning methodologies and the accuracies have been significantly increased compared with several individual classifier models. The advantages of ensemble learning are one bagging, used to reduce the variance in a model [16]. Two Boosting, used to reduce the bias in a model. Three Stacking, used to increase the prediction accuracy of a model [16]. Four random forests, which gives the best result for classification and decision tree-related problems. Five vote, which is used for the majority voting classifier. From the above tables[1],[2]shows the first, diabetes dataset DB_T4(with 153 instances) has higher average and weighted average such as 88.63% and 91.05%than other averages and weighted averages second, cholesterol dataset CH_T1(with 200 instances) has higher average and weighted average such as 90.10% and 90.25% than other averages and weighted averages. So form the above results always ensemble classifiers are better than a single classifier. The ensemble-based examining methods such as average and weighted average have more advantages in increasing the accuracies in the prediction of diseases diabetes and cholesterol.

REFERENCES

[1]. M. Ambika and Dr. K. Latha, Non-Communicable Diseases: An Approach for Prediction using Machine Learning Technique, International Journal of Applied

- Engineering Research, ISSN 0973-4562 Vol. 10 No.55 (2015)
- [2]. Prema N S, Varshith V, Yogeshwar J, Prediction of Diabetes using Ensemble Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S4, April 2019
- [3]. Dinesh Kumar G, Santhosh Kumar D, Arumugaraj K, Mareeswari V Prediction of Cardiovascular Disease Using Machine Learning Algorithms, Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India
- [4]. Divya Zion, *B.K. Tripathy, Handling Non-Communicable Disease Using Predictive Analysis of Data Mining Techniques, 2017 Helix ISSN 2319 – 5592[4]
- [5] Zhiwen Yu et al., Hybrid Incremental Ensemble Learning for Noisy Real-World Data Classification, IEEE Transactions on Cybernetics
- [6]. Shaohua Wan, Hua Yang*1, 2, Comparison among Methods of Ensemble Learning, 2013 International Symposium on Biometrics and Security Technologies
- [7] X. Liu et al.: Ensemble Transfer Learning Algorithm, IEEE Access
- [8] Aayushi Verma, Shikha Mehta, A Comparative Study of Ensemble Learning Methods for Classification in Bioinformatics, 2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence
- [9] https://link.springer.com/chapter/10.1007/978-981-13-1810-8_10
- [10] Esteban Alfaro, Matías Gámez and Noelia García, Ensemble Classification Methods with Applications in R
- [11] <https://www.ncbi.nlm.nih.gov/pubmed/29990215>
- [12] Heitor Murilo Gomes et al., A Survey on Ensemble Learning for Data Stream Classification, ACM Computing Surveys, Vol. 50, No. 2, Article 23, Publication date: March 2017.
- [13]. Cha Zhang • Yunqian Ma, Ensemble Machine Learning Methods and Applications, SPRINGER
- [14] en.wikipedia.org/wiki/Ensemble_learning
- [15]. <https://www.medicalnewstoday.com/articles/321519.php>.
- [16]. <https://medium.com/@saugata.paul1010/ensemble-learning-bagging-boosting-stacking-and-cascading-classifiers-in-machine-learning-9c66cb271674>
- [17] https://financeformulas.net/Weighted_Average.html#calcHeader
- [18] <https://study.com/academy/lesson/calculating-weighted-average-method-formula-example.html>
- [19] Purushottam ; Kanak Saxena ; Richa Sharma, Efficient heart disease prediction system using decision tree, International Conference on Computing, Communication & Automation
- [20]. N. L. Fitriyaniet al.: Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension, IEEE ACCESS