

Machine Learning Stacking Ensemble Model for Predicting Heart Attacks

Muath A. Obaidat

Department of Computer Science
City University of New York
New York, NY 10019
email: muobaidat@ccny.cuny.edu

Alex Alexandrou

Security and Emergency Management
John Jay College of Criminal Justice
New York, NY 10019
email: aalexandrou@jjay.cuny.edu

Samantha Sanacore

Department of Computer Science
City University of New York
New York, NY 10019
email: ssanacore@jjay.cuny.edu

Abstract— To mitigate the extent of one of the world’s leading causes of death, heart attacks, there needs to be an improvement in the technological aspect to predict this disease more accurately. Machine learning methods have come very far in increasing prediction accuracy based on patient data. Ensemble methods have exhibited improvement compared to individual classifier models. For this study, the goal is to develop a Machine Learning model to reach a very high level of accuracy for predicting myocardial infarction, otherwise known as a heart attack. A stacked ensemble model is used in this study and combines a group of three base-level classifiers such as Naive Bayes, Random Forest, and Extreme Gradient Boosting (XGBoost). This model will help identify those who are at risk and prevent heart attacks, therefore, lowering the mortality rate globally. Diversity among strong classifiers used in this model will be a more effective way to achieve the highest accuracy. The metrics used to evaluate the prediction performances are accuracy, Area Under the Curve (AUC), specificity, precision, and sensitivity. This process is carried out using RStudio and the results indicate that the proposed stacked ensemble method had a better performance under every evaluation metric compared to the individual base-level classifiers that were utilized.

Keywords— machine learning; naïve bayes; random forest; extreme gradient boosting; ensemble; heartattack; accuracy.

I. INTRODUCTION

Heart attacks, also known as myocardial infarction, have been one of the leading causes of death worldwide. The number of heart attacks has gone up by millions over the past decade [1]. In The U.S. alone, someone has a heart attack every 40 seconds and 1 in 5 attacks are silent, which results in damage being done to the person even if that person is unaware [2]. These staggering numbers are disturbing and will only get worse if no action is taken. Aside from the conventional dieting and exercise, something else that can be proactively done is to improve how we detect and predict heart attacks.

A heart attack is caused when a blood clot prevents oxygen from entering the heart, causing enough damage to the muscle cells that they start dying. Blood clots are formed when a buildup of plaque, made up of deposits, cholesterol, and other substances, ruptures and causes a blockage in the arteries [3]. It is important to evaluate and diagnose the patients early on and take steps to reduce or eliminate any of the risk factors, whether it be genetic or acquired. This should be done before

there is any irreversible damage, which would eradicate the ability to provide any treatment to the patient. If a patient has a blockage in a heart artery, the most common and effective procedure to reduce the risk of heart attack and improve the supply of blood, oxygen and reduce blockage in a coronary artery to the heart is coronary bypass surgery [4], which is relatively very rarely used nowadays. However, the most used procedure is angioplasty where stents are inserted to enlarge the artery; they are further absorbed by the artery wall [5].

Predicting and diagnosing patients early can possibly save someone’s life. This is where Machine Learning (ML) can have a major impact in the medical field. ML methods can be implemented to determine who is at risk of suffering a heart attack and get treatment.

ML implements data that is used as input and utilizes algorithms that can be trained to predict certain outcomes based on features from a provided dataset. ML continues to constantly evolve and has become beneficial in programming tasks that can predict or classify data. Classifiers separate data into classes and the prediction functions create a trend line, also known as the line of best fit, to fit a shape to get the closest to the data points. ML can fall into three categories: supervised machine learning, unsupervised machine learning, and semi-supervised learning. The classifiers used in this study fall under supervised machine learning [6].

An ensemble model is a valuable ML algorithm and can provide a variety of techniques for classification and regression. This study focuses on classification techniques to improve the prediction accuracy of heart attacks from a dataset. There are several ensemble techniques available such as bagging, boosting, stacking, and blending. Stacking is the technique that is relevant in our proposed model. The stacking ensemble model creates a strong meta-classifier, which is trained on features that are outputs from the combination of weak or base level classifiers [7].

With the use of ML, algorithms are trained to find patterns using a large dataset to make predictions. Some past researchers have used ML models to predict heart attacks using single classifiers and have had some high success rates. This proposed ML model plans to use a combination of three classifiers to achieve a high accuracy percentage, which would then result in a high success rate. This project will use a combination of datasets, such as Cleveland database [8] and Statlog (Heart) [9], which consists of 573 patients, 13 features and 1 target column. The target features determines whether

there is a “presence” or “absence” of heart disease, which is scaled as 0 or 1.

The 13 features in each patient’s data consist of age, sex, chest pain type (4 values) (cp), resting blood pressure (resttbps), serum cholesterol in mg/dl (chol), fasting blood sugar > 120 mg/dl (fbs), resting electrocardiographic results (values 0, 1, 2) (restecg), maximum heart rate achieved (thalach), exercise induced angina (exang), oldpeak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (CA), thal: 1=normal; 2=fixed defect; 3=reversible defect.

This study will use the analysis and evaluation of the features, with some risk factors incorporated in the patient’s data to help identify and accurately predict heart attacks. Out of the data, 70% is the training dataset and the remainder 30% of the data is considered the testing dataset. The training dataset is used to train the three classifier models. The testing dataset is used to test and evaluate the stacked ensemble model using the three classifier models. The three models used for classification are Random Forest, Naïve Bayes, and extreme gradient boosting (XGBoost). These classifiers are included in the ensemble model using the stacking technique. This ensemble method is proposed to achieve the best prediction accuracy of heart attacks and is evaluated by calculating the performances of accuracy, AUC, specificity, precision, and sensitivity.

The contributions made to this study include a unique combination of three base-level classifiers in the stacked ensemble model. The train control uses the 10-fold cross validation method for resampling the data. The meta-classifier uses a generalized linear model method.

The following sections make up the rest of the paper. Section II discusses the related work of this topic. The description of the proposed solution, along with the methodology, equations and implementation are presented in Section III. The results and discussion are reported in Section IV. Section V concludes the paper and provides some future work directions.

II. RELATED WORK AND BACKGROUND

There are many studies involving different machine learning techniques utilized to improve the prediction accuracy on heart attacks or heart disease. There are also some studies discussed in this paper that have used ensemble methods for prediction using various other datasets. The ensuing reviews of these methods and results are portrayed in this section.

Gao et al. [10] used K-nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, and Naïve Bayes to classify heart disease. The models were compared by using a boosting and bagging ensemble method with feature extractions algorithms such as linear discriminant analysis and principal component analysis. The results concluded that the bagging ensemble method with the principal component analysis feature extraction and Decision Tree achieved the best overall performance as compared to other models.

Parthasarathy et al. [11] used data from Cleveland UCI repository to perform heart disease classification by using Random Forest, Decision Tree, Support Vector Machines and Naïve Bayes based on only 9 of the 13 features provided. This study concluded that Random Forest classifier provided the best precision and accuracy when training the model using feature selection. Their model resulted in a prediction accuracy of 79.47%.

Obasi and Shafiq [12] compared Naïve Bayes, Logistic Regression and Random Forest using a dataset of 4838 observations combining Cleveland heart disease dataset, cardiovascular disease dataset and Framingham Heart study dataset. The study determined that Random Forest was the most accurate with 92.44%, followed by Naïve Bayes and Logistic Regression with accuracies of 61.96% and 59.70%, respectively.

Fang et al. [13] discussed how genes are important risk factors for myocardial infarction. The Recursive Feature Elimination (RFE) algorithm was implemented to find the 15 genes with the highest prediction accuracy. They integrated these genes using the GSE61144 dataset to construct a Support Vector Machine to predict the patients who have a high risk for myocardial infarction or heart attack. The outcome of the proposed model resulted in a 92% prediction accuracy.

Alaa et al. [14] developed a machine learning based model using AutoPrognosis to predict cardiovascular disease. The dataset contained more than 400,000 participants from the UK Biobank and included 473 features for each participant. The proposed AutoPrognosis model had a better AUC performance compared to the standard Framingham score and Cox PH models.

Revathi and Kavitha [15] compared Naïve Bayes, Instance-Based learning with parameter k (IBK) and Random Forest using the University of California, Irvine (UCI) heart disease dataset containing 270 observations. The study concluded that Naïve Bayes had the best performance with an 83.70% accuracy followed by Random Forest with 81.48% and IBK with 75.18%.

Gupta et al. [16] compared several models using the Cleveland heart disease dataset. The top 3 models with the highest accuracy were Naïve Bayes, AdaBoost and Boosted Tree with results of 86.42%, 86.21% and 85.75%, respectively. These top 3 models are implemented in the ensemble model and the accuracy of this model was the highest of all the models compared, with 87.91%.

Ali et al. [17] applied multiple techniques using the Hungarian and Cleveland datasets to their proposed ensemble approach to predict heart disease. This model is compared with classifiers such as Support Vector Machine, Logistic Regression, Multilayer perceptron, Random Forest, Decision Tree and Naïve Bayes, based on feature fusion, feature selection and weighing techniques. The model used the LogitBoost boosting algorithm and the proposed feature fusion approach to obtain results higher than existing ones with 98.5% accuracy.

Palaniappan and Awang [18] developed a prototype Intelligent Heart Disease Prediction System (IHDP) using Decision Trees, Naïve Bayes, and Neural Network. This system was built using CRISP-DM to build the models and used patient records from the Cleveland dataset. The Naïve Bayes model gives the highest accuracy of predicting heart disease with 95% followed by Decision Tree with 94.93% accuracy and Neural Network with 93.54% accuracy.

Tama et al. [19] used a stacked ensemble model in their research to predict coronary heart disease. This ensemble is a combination of classifiers such as Random Forest, Gradient Boosting Machine and Extreme Gradient Boosting, while applying Particle Swarm Optimization (PSO)-based feature selection and using datasets from Z-Alizadeh Sani, Cleveland, Hungarian and Statlog. This proposed model achieved a prediction accuracy of 98.13%, 93.55% and 91.18% which is the highest among other studies. It is compared with using Z-Alizadeh Sani, Statlog and Hungarian datasets, respectively.

Zhang et al. [20] compared the performances of eight base classifiers and chose the three with the best AUC results to be used for the stacking-based ensemble model to predict the risk of 30-day re-admission in patients with acute myocardial infarction. The proposed stacked model used an under-sampling method of neighborhood cleaning rule and a feature selection method of SelectFrom Model (SFM). The proposed stacked model had an AUC of 0.72, which was better than all the other classifiers the study compared.

Muhammad et al. [21] have researched which models and techniques result in an improved performance for the prediction of heart disease. Ten classifiers were compared and implemented by applying different feature selection algorithms to achieve the best performance possible. The top two classifiers without applying feature selection algorithms were Extra-Tree and Gradient Boost which had 92.09% and 91.34% accuracies, respectively. After experimenting with the feature selection algorithms, Extra-Tree had the highest accuracy, 94.41%, when the Relief algorithm was applied, and Gradient Boost had the highest increase to 93.36% with the Fast Correlation Based Filter algorithm.

To this end, we propose an ensemble model that uses fewer factors and variables to predict heart attacks. Nonetheless, the proposed model has better performance metrics than many previously discussed research studies while some of these studies have better performance in other metrics. The difference, however, is not very significant, which means the proposed scheme is well suited to predict heart attacks with a high degree of accuracy and proactively.

III. METHODOLOGY AND THE PROPOSED SCHEME

The objective of the proposed stacked algorithm is to improve the performance of being able to predict heart attacks. Fig. 1 displays the flow of the proposed scheme. This includes data splitting, training the base-level models, meta-classifier, and the evaluation metrics.

First, the data used for this model is collected from two databases. The Cleveland heart disease dataset has 303

observations and the Statlog (heart) dataset has 270 observations. Both datasets have the same number, 14, types of variables which allows them to be combined for a larger dataset to be used for this study. Once the data is imported, it is split into a training set and a testing set. The training set contains 70% of the data and the testing set contains 30% of the data.

The training set is used to train the models to learn about the features of the data. This is essential for future use to accurately predict the new data from the testing set. The dependent variable, y , is the target variable used for the prediction of heart attacks. The independent variable, x , is the variable or feature from the dataset that is used to determine the prediction results.

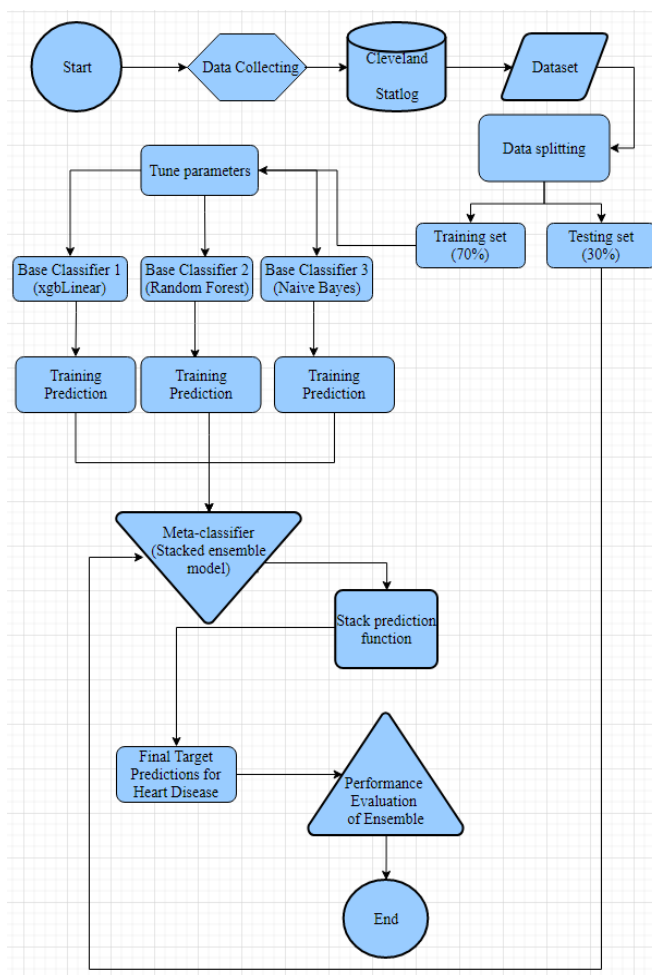


Figure 1. Scheme of proposed model to predict heart attacks.

The next step is to tune the parameters with ‘trainControl’ before training the base-level classifiers: Random Forest, XGBoost and Naïve Bayes. The tuning parameters would cross-validate (cv) the results 10-fold and save the predictions for later use with the class probabilities set to

‘True’. The ‘trainControl’ is utilized in the three base-level classifiers.

The three base-level classifiers use the training set and each are resampled. This set of data is used to train the models to learn the features to predict the target variable, the factor determining the possibility of suffering a heart attack. After each model is optimized with the parameters and trained, the prediction accuracy results are shown. The results of each model are averaged and merged into a single array by implementing the ‘caretList’ function. This function includes the target variable, the training set and the ‘trainControl’ parameters. This will be the input data for the meta-classifier/stacked model.

The Naive Bayes classifier is based on Bayes Theorem [22]:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

For this model, A and B are replaced with y, which is the output prediction variable, and x, which is the input variable. This equation solves for the probability of y given the input features, x, from the training set. The equation is re-written to adapt to the independent variables:

$$P(X|y) = P(x_1|y) * P(x_2|y) * \dots * P(x_n|y) \quad (2)$$

The probability of X, P(X), is a constant so it can be removed and replaced with a proportionality: $P(y|X) \propto P(X|y) * P(y)$. The last step of the Naïve Bayes classifier is to choose the class y with the maximum probability. This is accomplished by using “argmax” operation which finds the argument that results in the maximum y value:

$$y = \operatorname{argmax}_y [P(y) * \prod_{i=1}^n P(x_i|y)] \quad (3)$$

The Random Forest Classifier is made up of multiple decision trees for each patient observation. The algorithm starts with using random observations from the training set. The next step in the algorithm is to create a decision tree for every one of these observations to produce results containing the prediction for heart attacks. For each variable, a prediction is made, and voting is performed to determine the final prediction result for that observation [23].

The stacked ensemble model, ‘stack.model’, is implemented by using the ‘caretStack’ function with a generalized linear model (glm). The model uses the merged prediction results from the previous step, as well as the ‘Accuracy’ metric. The same tuning parameters to train the base-level models are utilized for this model under ‘stackControl’. The next step is to apply the ‘stack.model’ to the stack prediction function. This function produces the probability and uses ‘as.data.frame’, which returns a frame containing columns comparing the estimated prediction results using the input data from ‘stack.model’ and the actual results from the testing set.

The stacked ensemble model is constructed by using the dataset as the input. The dataset is $D = \{x_i, y_i\}_{i=1}^m$, where x_i

represents the feature vectors and y_i represents their classifications. The first phase of the model consists of a set of three base-level classifiers, h_i . The XGBoost classifier is h_1 , Random Forest classifier is h_2 and Naïve Bayes classifier is h_3 . Each base-level classifier, h_i , is trained by applying level 0. Level 0 is known as the data from the training set inputs which the level 0 classifiers make predictions from [24]. After the classifiers are trained, the next step is to construct a new dataset that contains $\{x_i^{new}, y_i\}$ [25], where:

$$X_i^{new} = \{h_1(x_i), h_2(x_i), \dots, h_T(x_i)\} \quad (4)$$

This step uses the predictions from the previous step as the new input for the meta-classifier. The final step is to learn the meta-classifier by applying a generalized linear model (glm) and training it. This new classifier, h^{new} , is based on the recently constructed dataset. The final output is displayed with H being the stacked ensemble model:

$$H(x) = h^{new}(h_1(x), h_2(x), \dots, h_T(x)) \quad (5)$$

The last step is to evaluate the metrics for the stacked ensemble model of the predicted results and the actual results using a confusion matrix. The metrics used to evaluate the prediction performance are accuracy, AUC, specificity, precision, and sensitivity. A confusion matrix shows the prediction results which are categorized into four sections. The True Positive (TP) outcome is defined as predicted true and true in reality. Another outcome is True Negative (TN), which means it is predicted false and false in reality. The False Positive (FP) outcome is defined as predicted true and false in reality, whereas the False Negative (FN) means that it is predicted false and true in reality [26]. Accuracy is defined as the number of correct predictions divided by the total number of predictions made. The closer the accuracy is to 100 percent, the stronger the performance of the model; the equation follows [26]:

$$\text{Acc} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (6)$$

For AUC, the closer the score is to 1, the better the model would be at differentiating between positive and negative predicted and actual values, and vice versa, with an underperforming model. Precision is defined as the number of correct positive results divided by the number of positive results predicted by the model. The precision equation is [27]:

$$\text{precision} = \frac{TP}{(TP+FP)} \quad (7)$$

The specificity metric gives us the true negative rate and is defined as the proportion of actual negative results that are correctly identified, also known as true negatives, to everything classified as negative. The equation for specificity follows [28], where TNR is true negative rate.

$$\text{TNR} = \frac{TN}{(TN+FP)} \quad (8)$$

Sensitivity gives us the true positive rate, which is the proportion of actual positive predictions that are correctly identified to everything classified as positive [29], where TPR is true positive rate.

$$TPR = \frac{TP}{(FN+TP)} \tag{9}$$

IV. RESULTS AND DISCUSSION

This section discusses the results of the proposed model’s performance along with the performance of the three individual classifier models. The overall performance of the ensemble model outperforms Random Forest, XGBoost and Naïve Bayes. The stacked ensemble model had the highest performance in accuracy, AUC, specificity, precision, and sensitivity with 80.11%, 85.53%, 84.21%, 85.89% and 76.84%, respectively.

In Fig. 2, the ensemble model achieved the highest accuracy performance at 80.12%. The second-best model was the Random Forest classifier model with an accuracy of 78.95%. The third best model was XGBoost with an accuracy of 74.85%. The worst accuracy performance of the group was Naïve Bayes with 73.1%.

Displayed in Fig. 3 is the AUC performance of each model. The ensemble model had the best results with 85.53% AUC followed by Random Forest with 78.97% and XGBoost with 74.15%. Fig. 5 shows the specificity results of the ensemble model, Random Forest, XGBoost and Naïve Bayes, which are 84.21%, 82.05%, 78.67% and 77.78%, respectively.

The specificity performance metric is presented in Fig. 4. As can be seen from Fig. 4, the ensemble model has the highest specificity, which reaches around 84.21%. It is clear that the proposed model achieves a very significant specificity as compared to other benchmark models in the study.

The precision metric is presented in Fig. 5. It shows that the best result is at 85.89%, which comes from the ensemble model. The next best result is at 83.53% from Random Forest and the worst performing models based on precision are both XGBoost and Naïve Bayes with 81.18%.

The sensitivity results conclude the evaluation of the performing models in Fig. 6, with the ensemble model having the best result with 78.84%; the next highest sensitivity result is Random Forest with 76.34%. The last two-classifier models, XGBoost and Naïve Bayes, have results of 71.88% and 69.7%, respectively.

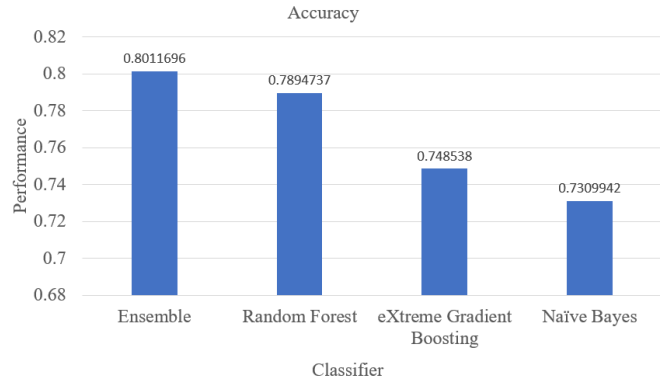


Figure 2. Accuracy

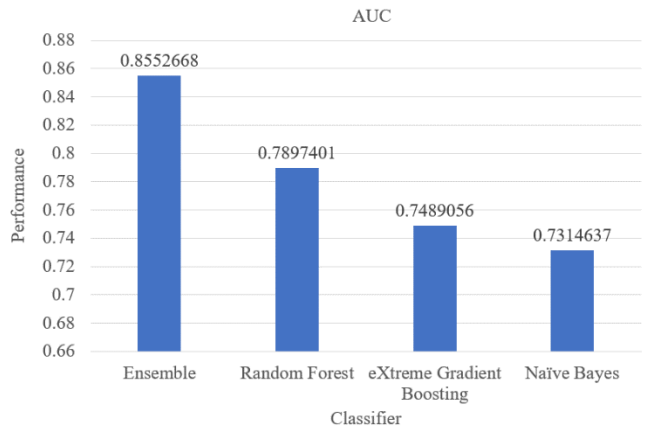


Figure 3. Area under the curve (AUC)

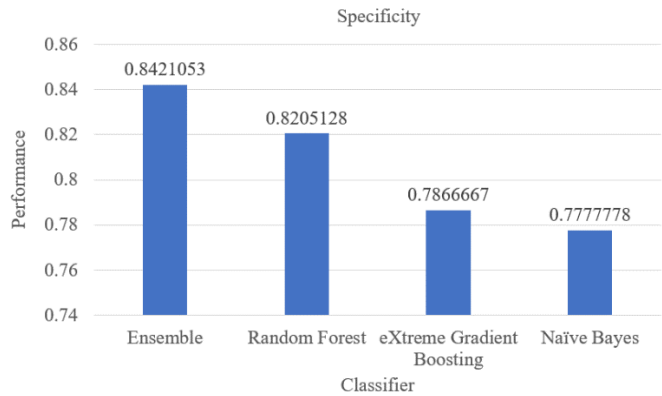


Figure 4. Specificity

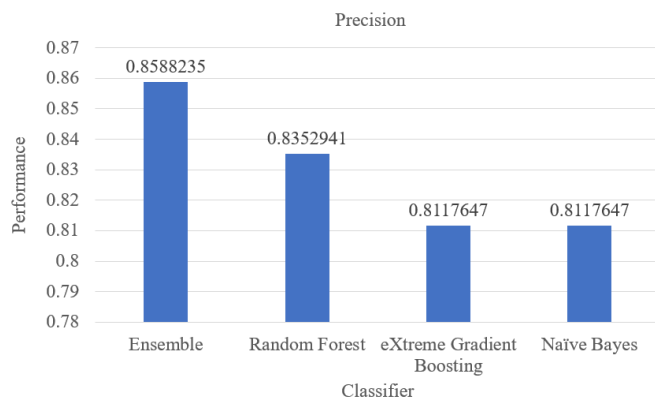


Figure 5. Precision

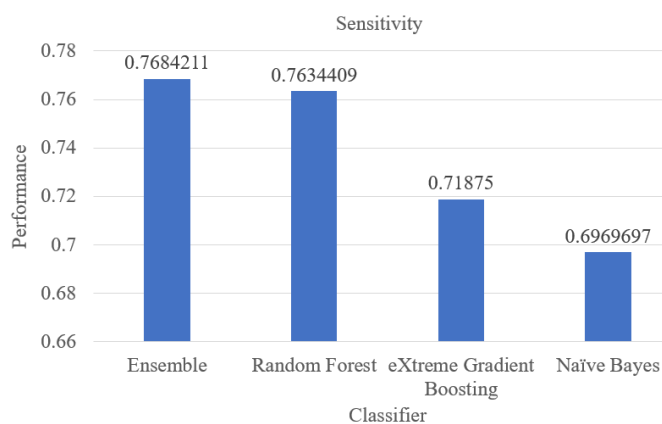


Figure 6. Sensitivity

V. CONCLUSION AND FUTURE WORK

In this work, we proposed a stacked ensemble model to improve the level of accuracy for the prediction of heart attacks. The stacked ensemble model achieved the highest results when evaluating the accuracy, AUC, specificity, precision, and sensitivity metrics compared to other schemes used in the study. The method had the highest accuracy performance, which was 1.16% higher than the next best performance from Random Forest. The performances of the three classifiers such as, Random Forest, Naïve Bayes and XGBoost, underperformed the ensemble model. Although these classifiers had lower performances individually, when they were combined, the performance improved significantly.

The proposed model can help reduce the number of deaths caused by heart attacks worldwide. Patients would be able to get early diagnoses and receive treatment by medical professionals in a timely manner. If the risk factors associated with the patient are determined to be genetic, treatments can help reduce the risk. An acquired risk factor due to lifestyle can be resolved by making changes to the patient's daily routine, which can decrease the probability of a heart attack.

This study can be enhanced by incorporating a larger dataset or by applying feature selection algorithms and

adjusting the hyper parameters of the base-level classifier models. More classification algorithms can be compared to determine which outcome produces the best performance model that more accurately predicts heart attacks. In addition, investigating when these ML models fail would be of interest focus of research.

REFERENCES

- [1] World Health Organization, "The top 10 causes of death," World Health Organization: WHO, Dec. 09, 2020. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. [retrieved: Jan, 2022]
- [2] Centers for Disease Control and Prevention, "Heart disease facts & statistics," Centers for Disease Control and Prevention, Sep. 08, 2020. <https://www.cdc.gov/heartdisease/facts.htm> [retrieved: Jan, 2022]
- [3] "Heart Attack," *Hopkins Medicine*, 2019. <https://www.hopkinsmedicine.org/health/conditions-and-diseases/heart-attack> [retrieved: Jan, 2022]
- [4] "Cardiac Procedures and Surgeries," www.heart.org, 2017. <https://www.heart.org/en/health-topics/heart-attack/treatment-of-a-heart-attack/cardiac-procedures-and-surgeries> [retrieved: Jan, 2022]
- [5] J. R. Stevens, A. Zamani, and J. I. A. Osborne, "Critical evaluation of stents in coronary angioplasty: a systematic review," *BioMed Eng OnLine* 20, 46 (2021). <https://doi.org/10.1186/s12938-021-00883-7>
- [6] IBM Cloud Education, "What is Machine Learning?," www.ibm.com, Jul. 15, 2020. <https://www.ibm.com/cloud/learn/machine-learning> [retrieved: Jan, 2022]
- [7] M. Alhamid, "Ensemble Models," *Medium*, Mar. 15, 2021. <https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c> [retrieved: Jan, 2022]
- [8] N. Bhat, "Health care: Dataset on Heart attack possibility," [kaggle.com](https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility). <https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility> [retrieved: Jan, 2022]
- [9] "UCI Machine Learning Repository: Statlog (Heart) DataSet," Uci.edu, 2019. <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29> [retrieved: Jan, 2022]
- [10] X.-Y. Gao, A. A. Ali, H. S. Hassan, and E. M. Anwar, "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method," *Complexity*, pp. 1–10, Feb. 2021, doi: 10.1155/2021/6663455.
- [11] G. Parthasarathy, L. S. Kesaragopp, M. M. Vishal, S. Manigandan and K. Kulkarni, "Analysis of machine learning algorithm for prediction of heart disease", *International Journal of Advanced Research in Computer Science*, vol. 11, no. 3, pp. 42–46, Jun. 2020, doi: 10.26483/ijarcs.v11i3.6532
- [12] T. Obasi and M. O. Shafiq, "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2393–2402, doi: 10.1109/BigData47090.2019.9005488.
- [13] H. Fang, D. Hu, Q. Li, and S. Tu, "Risk gene identification and support vector machine learning to construct an early diagnosis model of myocardial infarction," *Molecular Medicine Reports*, vol. 22, no. 3, pp. 1775–1782, Jun. 2020, doi: 10.3892/mmr.2020.11247.
- [14] A. Alaaid, T. Bolton, E. Angelantonio, J. Ruddid, and M. V. Der Schaar, "Cardiovascular disease risk prediction using

- automated machine learning: A prospective study of 423,604 UK Biobank participants,” May 2019. [retrieved: Jan, 2022]
- [15] K. K. Revathi, and K. K. Kavitha “Comparison of classification techniques on heart disease dataset,” *International Journal of Advanced Research in Computer Science*, vol. 8, no. 9, pp. 276–280, Sep. 2017, doi: 10.26483/ijarcs.v8i9.4870.
- [16] N. Gupta, N. Ahuja, S. Malhotra, A. Bala, and G. Kaur, “Intelligent heart disease prediction in cloud environment through ensembling,” *Expert Systems*, vol. 34, no. 3, p. e12207, Apr. 2017, doi: 10.1111/exsy.12207.
- [17] Ali Farman et al., “A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Ensemble Deep Learning and Feature Fusion.” *Information Fusion*, vol. 63, 2020, pp. 208–222., doi:10.1016/j.inffus.2020.06.008.
- [18] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," 2008 IEEE/ACS International Conference on Computer Systems and Applications, 2008, pp. 108-115, doi: 10.1109/AICCSA.2008.4493524.
- [19] A. Tama, B. I. Sun, and S. Lee, “Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble.” *BioMed research international* 2020 (2020): 9816142–10.
- [20] Z. Zhen et al., “A Stacking-Based Model for Predicting 30-Day All-Cause Hospital Readmissions of Patients with Acute Myocardial Infarction.” *BMC medical informatics and decision making* 20.1 (2020): 335–335.
- [21] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, “Early and accurate detection and diagnosis of heart disease using intelligent computational model,” *Scientific Reports*, vol. 10, no. 1, Nov. 2020, doi: 10.1038/s41598-020-76635-9.
- [22] S. Terence, “A Mathematical Explanation of Naive Bayes in 5 Minutes.” *Medium, Towards Data Science*, 6 June 2020, towardsdatascience.com/a-mathematical-explanation-of-naive-bayes-in-5-minutes-44adebcdb5f8.
- [23] Classification Algorithms Random Forest, www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm.
- [24] Johnny's Machine Learning Blog, “How to Develop a stacking Ensemble for Deep Learning Neural Networks in Python with Keras.” 18 Jan. 2020, johdev.com/ensemble/2020/01/18/Ensemble_Keras.html.
- [25] “Stacked Ensemble Learning,” [web.njit.edu. https://web.njit.edu/~avp38/projects/multi_projects/ensemble.html](https://web.njit.edu/~avp38/projects/multi_projects/ensemble.html) [retrieved: Jan, 2022]
- [26] N. S. Chauhan, “Model Evaluation Metrics in Machine Learning,” *KDnuggets*, May 2020. <https://www.kdnuggets.com/2020/05/model-evaluation-metrics-machine-learning.html> [retrieved: Jan, 2022]
- [27] A. Emma, “Machine Learning Evaluation Metrics.” *Medium, ML Cheat Sheet*, 3 Mar. 2020, medium.com/ml-cheat-sheet/machine-learning-evaluation-metrics-b89b8832e275.
- [28] S. Tavish, “Evaluation Metrics machine Learning” *Analytics Vidhya*, 5 Aug. 2020, <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>