IJIBC 24-3-31

Enhancing Heart Disease Prediction Accuracy through Soft Voting Ensemble Techniques

Byung-Joo Kim

Professor, Dept. EE, Youngsan University, Korea
bjkim@ysu.ac.kr

Abstract

We investigate the efficacy of ensemble learning methods, specifically the soft voting technique, for enhancing heart disease prediction accuracy. Our study uniquely combines Logistic Regression, SVM with RBF Kernel, and Random Forest models in a soft voting ensemble to improve predictive performance. We demonstrate that this approach outperforms individual models in diagnosing heart disease. Our research contributes to the field by applying a well-curated dataset with normalization and optimization techniques, conducting a comprehensive comparative analysis of different machine learning models, and showcasing the superior performance of the soft voting ensemble in medical diagnosis. This multifaceted approach allows us to provide a thorough evaluation of the soft voting ensemble's effectiveness in the context of heart disease prediction. We evaluate our models based on accuracy, precision, recall, F1 score, and Area Under the ROC Curve (AUC). Our results indicate that the soft voting ensemble technique achieves higher accuracy and robustness in heart disease prediction compared to individual classifiers. This study advances the application of machine learning in medical diagnostics, offering a novel approach to improve heart disease prediction. Our findings have significant implications for early detection and management of heart disease, potentially contributing to better patient outcomes and more efficient healthcare resource allocation.

Keywords: Heart Disease Diagnosis, Ensemble Learning, Soft Voting

1. INTRODUCTION

1.1 RESEARCH BACKGROUND

Heart disease remains one of the most significant health challenges worldwide, leading as a primary cause of morbidity and mortality across diverse populations [1]. The prevalence of heart disease highlights an urgent need for effective strategies for early detection, prevention, and management. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the number one cause of death globally, with an estimated 17.9 million lives lost each year. Among these, coronary heart disease and stroke account for the majority of CVD deaths, underscoring the critical importance of addressing heart health at both individual and

public health levels. The impact of heart disease extends beyond the substantial health burden, encompassing economic implications for healthcare systems and society at large. The costs associated with medical treatment, hospital care, and loss of productivity due to heart disease are significant, making it a major economic challenge for countries around the world [2]. Advancements in medical science have led to better diagnostic tools and treatments for heart disease. However, the ability to predict heart disease early and accurately remains a crucial unmet need. Early detection of heart disease can significantly improve patient outcomes by enabling timely intervention and preventing the progression of the disease. In this context, machine learning (ML) emerges as a powerful tool to enhance the prediction and diagnosis of heart disease [3-6]. ML algorithms can analyze vast amounts of health data, identify patterns, and learn from them to predict health outcomes with remarkable accuracy. The application of ML in heart disease prediction involves using clinical and non-clinical data to train models that can forecast the likelihood of heart disease in individuals. This predictive capability is particularly valuable in heart disease, where early indicators may be subtle or non-specific. Traditional risk assessment models, based on a limited set of variables, often fail to capture the complex interplay of factors contributing to heart disease. Machine learning models, by contrast, can integrate diverse data types—including demographic information, clinical metrics, lifestyle factors, and even genetic markers—to offer a more nuanced risk assessment. The development and validation of ML-based heart disease prediction systems have the potential to revolutionize cardiovascular healthcare. By providing healthcare professionals with more accurate and personalized risk assessments, these systems can inform targeted prevention strategies, optimize patient management, and ultimately reduce the burden of heart disease. This paper aims to explore the application of machine learning in developing a predictive model for heart disease, utilizing a dataset of clinical and demographic variables from patients at the Cleveland Clinic Foundation. Through this study, we seek to demonstrate the efficacy of machine learning algorithms in enhancing heart disease prediction and contributing to the broader efforts to combat this global health challenge.

2. LITERATURE REVIEW

Machine learning-based research has significantly advanced heart disease prediction, offering more accurate, efficient approaches. [7] explores the use of the Random Forest classifier to analyze factors contributing to Coronary Artery Disease (CAD), finding that demographic characteristics significantly impact CAD risk. [8] utilizes the Particle Swarm Optimization (PSO) algorithm to generate rules for heart disease diagnosis, comparing results with the C4.5 algorithm to showcase PSO's effectiveness. [9] introduces a heart disease prediction system developed utilizing neural network technology, which incorporates thirteen medical attributes to predict heart disease. Comparative analysis conducted within the scope of this research delineates the superior performance of the proposed system relative to existing state-of-the-art methodologies. Existing approaches are low accuracy and high computation time. In order to tackle these problems new methods are needed to detect correctly. The improvement in prediction accuracy is a big challenge and research gap. Current methods suffer from low accuracy and extended computational times. To address these issues, novel approaches are required for accurate detection. Enhancing prediction accuracy represents a significant challenge and an area needing further research.

3. THEORETICAL BACKGROUND OF SOFT VOTING

In this chapter, we present the theoretical background for employing the soft voting technique in this study. Ensemble methods have consistently demonstrated their ability to outperform single classifier systems in various predictive modeling tasks. Among these, the soft voting ensemble approach stands out for its

effectiveness in combining probabilistic estimates from multiple classifiers to make a final prediction [10]. Theoretical background of soft voting is as follows. First, ensemble methods reduce the overall error by averaging out the biases and variances of individual classifiers [11]. Soft voting, specifically, mitigates errors by weighing the probabilistic predictions of each classifier, thereby minimizing the impact of erroneous or overconfident predictions from any single model. Second, the strength of an ensemble lies in the diversity of its constituent classifiers. Soft voting benefits from the varied decision boundaries and predictive behaviors of different models, which collectively contribute to a more comprehensive exploration of the feature space[11]. Third, soft voting takes advantage of the probabilistic outputs of classifiers, incorporating the degree of certainty associated with each prediction. This allows the ensemble to make nuanced decisions based on the collective confidence of its components, leading to more accurate and calibrated predictions [12]. Fourth, the ensemble's reliance on the consensus among diverse models makes it less susceptible to overfitting, a common pitfall of complex single classifiers [13]. Fifth, the incorporation of multiple classifiers in a soft voting ensemble inherently leads to better generalization abilities [14]. By aggregating the predictive strengths of diverse classifiers, the ensemble is less likely to overfit on the training data, enhancing its capability to perform well on unseen data. This generalization is crucial for real-world applications where the data may vary significantly from the samples used during the model's training phase. Sixth, soft voting ensemble methods offer significant flexibility in the selection of constituent models [15]. Practitioners can combine various types of classifiers, regardless of their underlying algorithms, to harness their unique strengths. This versatility allows for the creation of a more robust and adaptable predictive system, capable of tackling the nuances and complexities of specific datasets. Seventh, soft voting not only considers the output probabilities of each classifier but also allows for weighting the contributions of different classifiers based on their performance [16]. By assigning higher weights to more accurate classifiers, the ensemble can further refine its predictions, leveraging the most reliable estimates to drive the final decision-making process. Finally in scenarios where datasets are imbalanced, soft voting ensembles can demonstrate remarkable robustness [17]. By integrating the probabilistic outputs from various classifiers, soft voting can effectively navigate the challenges posed by disproportionate class distributions, mitigating the risk of bias towards the majority class. In conclusion, the soft voting ensemble method represents a powerful and flexible approach to predictive modeling, addressing many of the limitations associated with single classifiers. Its theoretical foundation underpins its ability to deliver improved accuracy, robustness, and adaptability, making it an attractive choice for complex predictive tasks across various domains.

4. DATA

The dataset encompasses various variables related to heart disease, including age, sex, type of chest pain (cp), resting blood pressure (trestbps), cholesterol levels (chol), fasting blood sugar levels (fbs), electrocardiogram results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), the slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), results from the thallium stress test (thal), and the target variable indicating the presence of heart disease. Detailed description of data set in[18]. The dataset contains 303 observations with no missing values. Basic statistical summaries revealed, for example, that the average age is approximately 54.4 years, with a range from 29 to 77 years. The average cholesterol level was observed to be around 246.7 mg/dl. Males were more prevalent than females, and the presence of heart disease was nearly evenly split among the subjects.

5. EXPERIMENT

The study conducted a comprehensive analysis using a normalized heart disease dataset, focusing on three primary models: Logistic Regression, SVM with RBF Kernel, and Random Forest. Each model was optimized using GridSearchCV or RandomizedSearchCV to find the best hyperparameters. The soft voting ensemble was then constructed with these optimized models, utilizing their probabilistic predictions to make a final decision on the presence of heart disease. In the context of heart disease, evaluating the performance of predictive models is critical for determining their effectiveness in accurately diagnosing heart disease. The key evaluation metrics are accuracy, recall, precision, F1 score and ROC curve. Accuracy measures the overall correctness of a model's predictions across all classes. It is the ratio of correct predictions to the total number of cases evaluated [19]. Recall, also known as sensitivity, measures the completeness of positive predictions. It quantifies the proportion of actual positive cases that were correctly identified, indicating the model's ability to find all positive instances [20]. Precision indicates the accuracy of positive predictions. It measures the proportion of predicted positive cases that are actually positive, reflecting the model's ability to avoid labeling negative instances as positive [21]. The F1 score is a statistical measure used to evaluate the performance of binary classification models. It is the harmonic mean of precision and recall, providing a single metric that balances both the accuracy of positive predictions (precision) and the completeness of capturing positive instances (recall). This score is particularly useful when dealing with imbalanced datasets [22]. The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of a binary classifier system as its discrimination threshold is varied. The curve plots the True Positive Rate against the False Positive Rate at various threshold settings. The ROC curve starts at the point (0,0) and ends at (1,1). A curve closer to the top-left corner indicates better classifier performance, representing a higher true positive rate and a lower false positive rate [23-24]. The Area Under the ROC Curve (AUC) is used as a single scalar value to summarize the overall performance of the classifier. A higher AUC value (closer to 1) indicates better model performance, while a value of 0.5 suggests performance no better than random chance.

5.1 SVM

In this study, we employed a SVM with a Radial Basis Function (RBF) kernel to predict the presence of heart disease. The objective was to optimize the SVM parameters, specifically the cost parameter (C) and the gamma parameter (γ), to enhance model performance. Utilizing a dataset normalized prior to analysis, we divided the data into training and testing sets, ensuring a representative split of the underlying patterns within the data. We allocated 80% of the dataset for training and the remaining 20% for testing. To find the optimal C and gamma values for the SVM with an RBF kernel, we conducted a grid search across a predefined range of values for both parameters. The grid search involved fitting the model using five-fold cross-validation on the training set and evaluating accuracy to determine the best combination of parameters. The grid search identified optimal parameters of C=10.0 and γ=0.01. The model's performance on the test set, evaluated by various metrics and the ROC curve, is as follows. Table 1 shows that the model produced an ROC curve, achieving an AUC of 0.92. This indicates a high level of model performance, suggesting that the SVM model with optimized parameters effectively distinguishes between the presence and absence of heart disease.

Table 1. Performance of SVM by evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
SVM	91.80%	93.55%	90.63%	92.06%	95.37%

5.2 RANDOM FOREST

Using RandomizedSearchCV for the optimization of hyperparameters in the Random Forest model yielded the following best parameters: n_estimators=200, min_samples_split=4, min_samples_leaf=4, and max_depth=10. The model's performance on the test set, evaluated by various metrics and the ROC curve, is shown in Table 2. The ROC curve, with an AUC of 0.94, indicates a high true positive rate (TPR) across different thresholds, showcasing the model's effective discrimination between the presence and absence of heart disease.

Table 2. Performance of Random Forest by evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
Random Forest	90.16%	96.43%	84.38%	90.00%	94.29%

5.3 LOGISTIC REGRESSION

The Logistic Regression model's performance on the test set, evaluated through various metrics and the ROC curve, is presented in Table 3. The ROC curve, with an AUC of 0.94, indicates a high TPR across different thresholds, showcasing the model's effective discrimination between the presence and absence of heart disease.

Table 3. Performance of Logistic Regression by evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
Logistic Regression	88.52%	87.88%	90.63%	90.23%	94.61%

5.4 SOFT VOTING

Using a Soft Voting Classifier with probability estimation enabled for SVM, along with Logistic Regression and Random Forest models, yielded the following performance metrics on the test set. Table 4 shows the ROC curve, which has an AUC of approximately 0.95, demonstrating the ensemble model's excellent capability to discriminate between the presence and absence of heart disease. The combination of these three models through soft voting effectively leverages their individual strengths, resulting in a robust classifier with high predictive performance.

Table 4. Performance of Soft Voting by evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
Soft Voting	91.80%	93.55%	90.63%	92.06%	95.37%

6. EXPERIMENT RESULT

Table 5 shows a comprehensive comparison of the Logistic Regression, SVM with RBF Kernel, Random Forest, and Soft Voting Classifier models based on accuracy, precision, recall, F1 score, and AUC. The Soft Voting Classifier shows excellent performance across all metrics, closely matching or slightly surpassing the performance of the SVM with RBF Kernel model. This indicates the effectiveness of ensemble methods in leveraging the strengths of individual models to achieve higher predictive performance.

Table 5. Comparative performance results of different classifiers by evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1 score	AUC
Soft voting	91.80%	93.55%	90.63%	92.06%	95.37%
SVM	91.80%	93.55%	90.63%	92.06%	95.26%
Random forest	90.16%	96.43%	84.38%	90.00%	94.29%
Logistic regression	88.52%	87.88%	90.63%	89.23%	94.61%

Figure 1 shows the ROC curves for the various models we have discussed. This visualization clearly demonstrates the predictive performance of each model in terms of their ability to discriminate between the presence and absence of heart disease. The Soft Voting Classifier and SVM with RBF Kernel exhibit the highest AUC, indicating their superior performance among the models evaluated.
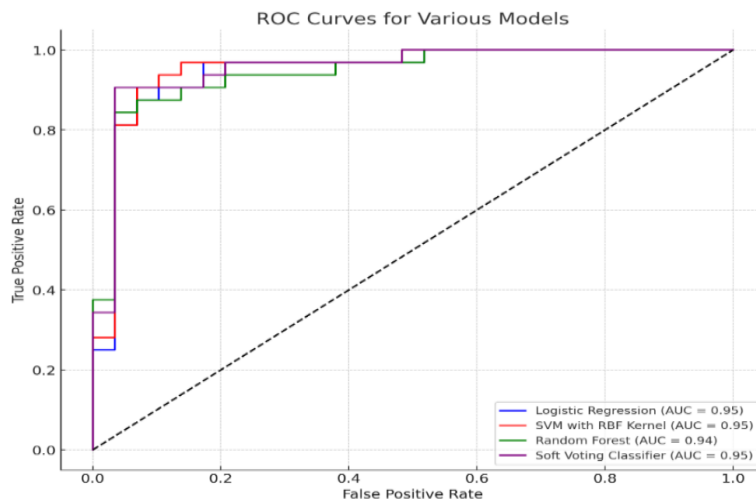

Figure 1. ROC curves for various models

7. CONCLUSION

We demonstrate the potential of soft voting ensemble methods in improving heart disease prediction through our comprehensive study. We uniquely integrate the probabilistic predictions of Logistic Regression, SVM with RBF Kernel, and Random Forest models into a soft voting ensemble method, significantly enhancing prediction accuracy. Our research showcases the viability of this approach as a robust tool in medical diagnostic processes, particularly for heart disease prediction. We have provided evidence that the soft voting ensemble method outperforms individual models, offering a more reliable and accurate prediction framework. Our findings contribute to the growing body of knowledge in predictive analytics for healthcare, specifically in the domain of cardiovascular diseases. For future work, we suggest exploring the integration of additional models into the ensemble and applying these ensemble techniques to other diseases. This approach could potentially broaden the scope of predictive analytics in healthcare, leading to improved diagnostic processes across various medical fields. By extending our methodology to other areas, we aim to contribute to the advancement of data-driven decision-making in healthcare diagnostics.

References

[1] D. R. Labarthe, Epidemiology and Prevention of Cardiovascular Diseases: A Global Challenge, Jones & Bartlett Publishers, 2010.

[2] A. P. Kengne, A. E. Moran, K. Sliwa, and F. Mbanya, Cardiovascular diseases and diabetes as economic and developmental challenges in Africa, Prog. Cardiovasc. Dis., Vol. 56, No. 3, pp. 302-313, 2013.

[3] Y. Khan, C. S. Pythian, and I. Jaswinder, Machine learning techniques for heart disease datasets: A survey, in Proc. 2019 11th Int. Conf. Mach. Learn. Comput., pp. 27-35, 2019.

[4] K. K. L. Wong, G. Fortino, and D. Abbott, Deep learning-based cardiovascular image diagnosis: a promising challenge, Future Gener. Comput. Syst., Vol. 110, pp. 802-811, 2020.

[5] N. Kagiyama, R. Shrestha, S. Farjo, and P. Sengupta, Artificial intelligence: practical primer for clinical research in cardiovascular disease, J. Am. Heart Assoc., Vol. 8, No. 17, e012788, 2019.

[6] S. S. Yadav, M. Jadhav, S. Nagrale, and N. Patil, Application of machine learning for the detection of heart disease, in Proc. 2020 2nd Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA), pp. 165-172, 2020.

[7] A. S. Abdullah and R. Rajalaxmi, A data mining model for predicting the coronary heart disease using random forest classifier, in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, 2012.

[8] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, Using PSO algorithm for producing best rules in diagnosis of heart disease, in Proc. 2017 Int. Conf. Comput. Appl. (ICCA), 2017.

[9] N. Al-Milli, Backpropagation neural network for prediction of heart disease, J. Theor. Appl. Inf. Technol., Vol. 56, No. 1, pp. 131-135, 2013.

[10] M. A. Khan, M. Umair, M. A. S. Choudhry, and M. K. Chattha, An optimized ensemble prediction model using AutoML based on soft voting classifier for network intrusion detection, J. Netw. Comput. Appl., Vol. 212, 103560, 2023.

[11] S. Karlos, G. Kostopoulos, and S. Kotsiantis, A soft-voting ensemble based co-training scheme using static selection for binary classification problems, Algorithms, Vol. 13, No. 1, 26, 2020.

[12] S. Kumari, D. Kumar, and M. Mittal, An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier, Int. J. Cogn. Comput. Eng., Vol. 2, pp. 40-46, 2021.

[13] M. Shahhosseini, G. Hu, and H. Pham, Optimizing ensemble weights and hyperparameters of machine learning models for regression problems, Mach. Learn. Appl., Vol. 7, 100251, 2022.

[14] S. Kumari, D. Kumar, and M. Mittal, An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier, Int. J. Cogn. Comput. Eng., Vol. 2, pp. 40-46, 2021.

[15] A. Özçift, Medical sentiment analysis based on soft voting ensemble algorithm, Yönetim Bilişim Sist. Derg., Vol. 6, No. 1, pp. 42-50, 2020.

[16] A. Onan, S. Korukoğlu, and H. Bulut, A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification, Expert Syst. Appl., Vol. 62, pp. 1-16, 2016.

[17] D. Yang, B. Xiao, M. Cao, and H. Shen, A new hybrid credit scoring ensemble model with feature enhancement and soft voting weight optimization, Expert Syst. Appl., Vol. 238, 122101, 2024.

[18] Heart Disease Cleveland; https://www.kaggle.com/datasets/ritwikb3/heart-disease-cleveland.

[19] A. Gunawardana and G. Shani, A survey of accuracy evaluation metrics of recommendation tasks, J. Mach. Learn. Res., Vol. 10, No. 12, 2009.

[20] B. Juba and H. S. Le, Precision-recall versus accuracy and the role of large data sets, in Proc. AAAI Conf. Artif. Intell., Vol. 33, No. 01, pp. 4039-4048, July 2019.

[21] E. J. Michaud, Z. Liu, and M. Tegmark, Precision Machine Learning, Entropy, Vol. 25, No. 1, pp. 175, 2023.

[22] D. Chicco and G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics, Vol. 21, pp. 1-13, 2020.

[23] J. Fan, S. Upadhye, and A. Worster, Understanding receiver operating characteristic (ROC) curves, Can. J. Emerg. Med., Vol. 8, No. 1, pp. 19-20, 2006.

[24] H. I. Hahn, Comparative Analysis of CNN Techniques designed for Rotated Object Classification, Int. J. Internet Broadcast. Commun., Vol. 24, No. 1, 2024.