A Comparison of Methods for Predicting Heart Disease: Neural Network, XGBoost, Gradient Boost, Logistic Regression, and SVM

Mohammad Maydanchi*[1], Mina Basiri[2], Armin Ziaei[3], Fatemeh Haji[4], Mehrbod Ziaei[5], Saman Sargolzaei[6]

[1]*Department of Industrial and Systems Engineering, Auburn University, AL, USA*
[2]*Industrial Engineering Department, Tarbiat Modares University, Tehran, Iran*
[3]*Department of Engineering and Computer Science, The University of Texas at Dallas, TX, USA*
[4]*Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*
[5]*Department of Electrical Engineering, University of Science and Culture, Tehran, Iran*
[6]*Department of Engineering, The University of Tennessee at Martin, TN, USA*
*mzm0181@auburn.edu

***Abstract*—Cardiovascular diseases (CVD) are a major cause of mortality, demanding immediate attention. Early screening for risk factors speeds up CVD identification and treatment, reducing mortality risk efficiently. This study compares the predictive abilities of XGBoost, Gradient Boost, Neural Network, Logistic Regression, and SVM in forecasting CVD symptoms using CDC data from Kaggle. Evaluating against a heart disease risk approach, we assess model performance with AUC and F_1-score for class 1 due to data imbalance. Additionally, we consider training run time as another metric for evaluation. XGBoost, Gradient Boost, and Logistic Regression excel in AUC and runtime, outperforming Neural Network and SVM significantly. The performances are quite similar when comparing the F_1-scores, but Logistic Regression, XGBoost, and Gradient Boost still maintain their superiority. The paper recommends using Logistic Regression and boosting-based decision tree models for future heart disease prediction investigations.**

***Index Terms*—Artificial Intelligence, Machine Learning in Healthcare, Comparison Analytics, Classification Models, Heart Disease, Medical Data.**

I. Introduction

Cardiovascular disease (CVD) refers to conditions that affect the heart and circulatory system. CVD is influenced by genetics, behavior, psychosocial factors, other diseases, delayed diagnosis, and non-compliance with medical and behavioral strategies [1].Globally, CVD stands as one of the major causes of death. The American Heart Association (AHA) recently published a statement on the public health burden of CVD, urging culturally appropriate interventions and prioritized goals to enhance heart health in the country [2].Artificial intelligence(AI) merges computer science and physiology to envision computers emulating human behavior more efficiently than humans [3], [4].Machine learning(ML), a subset of AI, uses algorithms to build models from sample data for predicting or making decisions based on known training data to perform specific tasks [5].Machine learning promises to benefit organizations by providing predictions and aiding in decision-making. Identifying its advantages and disadvantages is crucial for a clear understanding of its applications [6].One of the many techniques to improve the diagnosis of CVD is data mining with the help of machine learning methods [7].These techniques allow extracting hidden knowledge and identifying relationships between features in a dataset and are a promising strategy for CVD classification [8].Ensuring high-quality, cost-effective clinical services is a crucial challenge for health organizations, requiring accurate patient diagnosis and effective treatment to avoid misdiagnoses [9].Early CVD detection reduces costs and mortality. Data mining with a classification algorithm in clinical research efficiently achieves this at a low cost [10], [11].

ML methods help protect against heart disease and movement disorders in medical datasets [12].Discovering such essential data helps researchers gain valuable insight into applying their diagnosis and treatment to a particular patient. Various machine learning methods analyze vast healthcare data, aiding professionals in disease prediction [13].The diagnosis and treatment of CVD are accelerated with the help of early screening, and for this purpose, data analysis methods can be used [14], [15].

Our previous study compared the performance of Decision Tree, AdaBoost, Random Forest, Naïve Bayes, KNN, and Perceptron for Heart Disease Prediction [16]., showing that AdaBoost stood out with the highest AUC at 0.828 and an F_1-score of 0.37. In contrast, the Decision Tree model displayed the lowest AUC, measuring at 0.595, and an F_1-score of 0.25. In this present paper, we have expanded our analysis to include five additional machine-learning methods for predicting heart disease. In the previous paper, we solely focused on the evaluation of F_1-score and AUC, but in the current study, we incorporate these two metrics while introducing the runtime of the models for our comparative analysis. In the upcoming section, we assess the performance of these classification models, namely XGBoost, Gradient Boost, Neural Network, Logistic Regression, and SVM, in predicting CVD symptoms.

II. Materials and Methods

This study used the publicly available 'Behavioral Risk Factor Surveillance System (BRFSS)' dataset obtained from Kaggle [17]. The BRFSS dataset has been collected by the Centers for Disease Control and Prevention (CDC) as a major component of their Population Health Surveillance Branch under the Division of Population Health at the National Center for Chronic Disease Prevention and Health Promotion [18]. The BRFSS conducts annual telephone surveys nationwide, including all 50 states, the District of Columbia, and three US territories. It is the world's largest continuously conducted health survey system. The CDC states that the 2020 dataset comprises comprehensive health-related survey data from 400,000 US adults, encompassing risk behaviors, chronic health conditions, and preventive services.

The target variable in our heart disease prediction study is a binary variable indicating the presence (class 1) or absence (class 0) of heart disease risk. The dataset includes 27,373 samples of class 1 and 292,422 samples of class 0, indicating an imbalance with only 8.5% of samples in class 1. This imbalance requires careful consideration as it may lead to biased model predictions, affecting the accurate identification of individuals at risk of heart disease. This dataset contains a variety of categorical and continuous features that can be used to predict heart disease. There are 13 categorical variables and four continuous variables in addition to the target variable [16]. In order to extract and use patterns in the data and evaluate the models, it is needed to do the preprocessing steps. To convert categorical data into numeric format, we employ classical encoding for binary variables and one-hot encoding for those with more than two unique values. The dataset used in this analysis has 42 dimensions, resulting from encoding the original 18 dimensions.

We divide the data into a training dataset (80%) and a test dataset (20%). For scaling, this paper uses the standard scaling (Z-score) method. The formula for Z-score is given by:

$$Z = \frac{X - \mu}{\sigma}$$

Where:

$$Z = \text{Standardized value (Z-score)}$$
$$X = \text{Raw score or data point}$$
$$\mu = \text{Mean (average) of the data set}$$
$$\sigma = \text{Standard deviation of the data set}$$

In this study, we use the Synthetic Minority Oversampling Technique (SMOTE) to deal with the imbalance present in the dataset. [19]. The initial sample count was 319,795, and applying SMOTE increased it to 467,928. In this paper, we used k-fold cross-validation to reduce overfitting in order to obtain better models for predicting new data. Supervised learning models in this study include Support Vector Machines (SVMs), Neural Network, Logistic Regression, Gradient Boost, and XGBoost, introduced in the following sentences. SVMs constitute a powerful class of supervised learning algorithms extensively employed within the realm of machine learning [20]. SVMs excel in various applications, owing to their robustness and capability to handle complex data distributions. Neural Network is a machine learning model designed to enable computers to process data in a manner that emulates the human brain's cognitive processes [21]. These networks consist of interconnected nodes, or neurons, which collectively collaborate to tackle complex problem-solving tasks. Logistic regression is a fundamental statistical method employed in predictive modeling and data analysis [22]. It uses mathematical principles, particularly the logistic function, to predict binary target variables [23]. The last two models are in the family of decision trees. The Gradient Boost method is a machine learning technique in which weak prediction models are combined to create a prediction model [24]. Typically, weak models can be described as simple decision trees that make few assumptions about the data in order to fit the model. As a result of the use of the best possible next model in combination with the previous model, the overall prediction error is minimized as much as possible. XGBoost, short for Extreme Gradient Boost, is a prominent machine learning library [25]. Like Gradient Boost, XGBoost employs the strategy of utilizing weak decision tree models to assemble a potent predictive model. We have selected these five traditional classification models for their excellence in predicting tasks with numerous parameters.

Table I shows the main parameters of these models.

TABLE I: Parameters of the classification models used

Model	**Parameter**
SVM	kernel = rbf
Neural Network	optimizer = adam
XGBoost	gamma = 0.12
Logistic Regression	penalty = l2
Gradient Boost	learning_rate = 0.1

We used Python 3.10.10 for the computations. All codes are available at https://github.com/Machine-Learning-Projects1/2020-BRFSS-Codebook-CDC/blob/main/CDC_ML_2.ipynb. In evaluating the performance of various machine learning models, we considered F_1-score, AUC of ROC and runtime. All computations were executed on a system equipped with a 12th Gen Intel(R) Core(TM) i7-12700H processor running at 2.30 GHz, accompanied by 16.0 GB of RAM. In the following section, we present the results of this comparison.

III. Result and Disussion

This section aims to analyze the performance of five classification models on the prediction of heart disease. These models include Neural Network, SVM, Logistic Regression, XGBoost, and Gradient Boost. Table II shows the performance of all five models with respect to four evaluation metrics. After training, the XGBoost model demonstrated a peak accuracy of 89%, whereas the Neural Network model achieved a maximum accuracy of 75%. As indicated in the previous section, each

instance's target variable is classified as either class 0 or class 1, with a dataset showing moderate imbalance—over tenfold more class 0 samples than class 1. Hence, accuracy isn't a reliable measure for class differentiation. Prioritizing predicting heart disease in at-risk individuals is more significant than forecasting its absence. Emphasizing the importance of class 1 over class 0, our study focuses on evaluating models for predicting class 1. Within this dataset, the precision of class 1 indicates the proportion of patients who belong to class 1 and who were correctly predicted by the model to belong to that class. Similarly, the recall value for class 1 represents the proportion of samples belonging to class 1 that are correctly classified. Based on the findings presented in Table II, the XGBoost model exhibits the highest precision of 30%, whereas SVM and Neural Network have the lowest precision of 22%. Furthermore, Logistic Regression achieved the highest recall of 78%. On the other end of the spectrum, the XGBoost model attained the lowest recall value, standing at 49%. As a result of the combination of these two metrics, it can be seen that XGBoost has the highest F_1-score of 38%, whereas Neural Network has the lowest F_1-score of 33%. In this study, we use the Receiver Operating Characteristic (ROC) curve as an additional metric to evaluate and compare models. The ROC curve's x-axis shows the false positive rate, and the y-axis represents the true positive rate. The true positive rate is synonymous with recall, while the false positive rate is the fraction of samples wrongly classified as class 1 to all samples not belonging to class 1. Calculating these rates for all classifiers, we plot ROC curves. The ROC curves of five models are shown in Figure 1. When comparing models, the one with the highest Area Under the Curve (AUC) is considered the best. Logistic Regression outperforms with the highest AUC of 0.843, while SVM has the lowest AUC at 0.723, making it the least-performing model in this group.

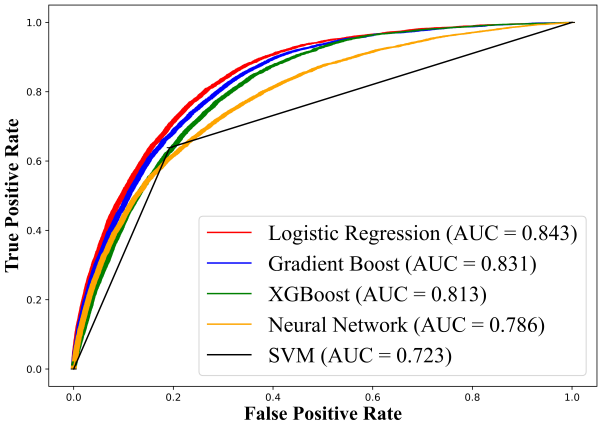

Fig. 1: The ROC curves for the classification

Additionally, this paper includes a comparison of models based on their training runtime. Table III presents the outcomes of this comparison.

The research uncovers that Logistic Regression demonstrates the quickest runtime, completing in a mere 38.3 seconds, whereas the SVM model has the slowest runtime, taking around 134 minutes.

IV. Conclusions

This study conducts a thorough performance analysis of five classification models, namely Support Vector Machine (SVM), Neural Network (NN), XGBoost, Logistic Regression, and Gradient Boost, for predicting heart diseases. This investigation marks a significant step in understanding the utility of machine learning in diagnosing heart disease, emphasizing the efficacy of different models under various conditions. For evaluation, key metrics used for model comparison include F_1-score and Area Under the Curve (AUC) of the ROC curve. These metrics were chosen over accuracy due to the imbalanced dataset. The analysis focuses intentionally on class 1, related to individuals with heart disease. Another considered metric is the training runtime of the models.

In F_1-score, the model ranking is: 1- XGBoost, 2- Gradient Boost, 3- Logistic Regression, 4- SVM, and 5- Neural Network. In AUC, it is: 1- Logistic Regression, 2- Gradient Boost, 3- XGBoost, 4- Neural Network, 5- SVM. Finally, for training runtime, the ranking is: 1- Logistic Regression, 2- XGBoost, 3- Gradient Boost, 4- Neural Network, and 5- SVM. The top three performers across all five metrics are Logistic Regression, Gradient Boost, and XGBoost, suggesting a focus on these models for future investigations with similar datasets. Less attention is advised for Neural Network and SVM. Logistic Regression's runtime is under a minute, while SVM takes more than 134 minutes. This suggests that if runtime is a critical metric for healthcare specialists in the field of heart disease, they should prioritize Logistic Regression and disregard SVM. This issue becomes even more pronounced when dealing with larger datasets. Neural Network results defied expectations, yet exploring other deep learning methods shows potential for future work to improve outcomes.

In our previous study [16], AdaBoost outperformed with the highest F_1-score of 37% and AUC of 0.825 among six models. Naïve Bayes had the second-highest F_1-score of 32% and AUC of 0.813. Comparing these to the current paper, among eleven models, XGBoost, Gradient Boost, AdaBoost, Logistic Regression, and Naïve Bayes exhibit the best performances. It is noteworthy that their performances are quite close to each other. Furthermore, it is evident that among the top five models, three of them belong to the Decision Tree family and utilize boosting methods. This suggests that efforts to enhance the boosting method can lead to improving the prediction of heart disease.

This research highlights machine learning's potential for heart disease diagnosis, aiding emerging medical professionals. Top models, with validity and interpretability, can enhance clinicians' diagnostic speed and accuracy. The models' transparency supports adoption in healthcare, laying the foundation for advanced diagnostic systems and personalized healthcare initiatives. These findings advocate a strategic

TABLE II: Models performance metrics

Model	Accuracy	Class 0 (No)			Class 1 (Yes)		
		Precision	Recall	F_1-score	Precision	Recall	F_1-score
XGBoost	86%	95%	89%	92%	30%	49%	38%
Gradient Boost	84%	95%	87%	91%	29%	53%	37%
Logistic Regression	75%	97%	75%	84%	23%	78%	35%
SVM	76%	97%	76%	85%	22%	71%	34%
Neural Network	75%	96%	76%	85%	22%	71%	33%

TABLE III: Run time of the models

Model	Runtime
Logistic Regression (LR)	38.3s
XGBoost	3m 14.3s
Gradient Boost (GB)	29m 29.8s
Neural Network (NN)	101m 29.3s
SVM	134m 31s

reassessment of machine learning in clinical diagnostics.

REFERENCES

[1] Y. Fatemi, M. Nikfar, A. Oladazimi, J. Zheng, H. Hoy, and H. Ali, "Machine Learning Approaches for Cardiovascular Death Prediction Among Nash Liver Transplant Recipients," Rochester, NY, Oct. 2023. [Online]. Available: https://papers.ssrn.com/abstract=4605366

[2] C. J. Rodriguez, M. Allison, M. L. Daviglus, C. R. Isasi, C. Keller, E. C. Leira, L. Palaniappan, I. L. Piña, S. M. Ramirez, B. Rodriguez, and M. Sims, "Status of Cardiovascular Disease and Stroke in Hispanics/Latinos in the United States," *Circulation*, vol. 130, no. 7, pp. 593–625, Aug. 2014, publisher: American Heart Association.

[3] M. D. Flather, M. C. Shibata, A. J. Coats, D. J. Van Veldhuisen, A. Parkhomenko, J. Borbola, A. Cohen-Solal, D. Dumitrascu, R. Ferrari, P. Lechat, J. Soler-Soler, L. Tavazzi, L. Spinarova, J. Toman, M. Böhm, S. D. Anker, S. G. Thompson, and P. A. Poole-Wilson, "Randomized trial to determine the effect of nebivolol on mortality and cardiovascular hospital admission in elderly patients with heart failure (SENIORS)," *European Heart Journal*, vol. 26, no. 3, pp. 215–225, Feb. 2005.

[4] C. Cai, B. Gou, M. Khishe, M. Mohammadi, S. Rashidi, R. Moradpour, and S. Mirjalili, "Improved deep convolutional neural networks using chimp optimization algorithm for Covid19 diagnosis from the X-ray images," *Expert Systems with Applications*, vol. 213, p. 119206, Mar. 2023.

[5] P. H. Campbell, S. Milbourne, L. M. Dugan, and M. J. Wilcox, "A Review of Evidence on Practices for Teaching Young Children to Use Assistive Technology Devices," *Topics in Early Childhood Special Education*, vol. 26, no. 1, pp. 3–13, 2006, publisher: PRO-ED, Inc ERIC Number: EJ762928.

[6] K. C. A. Khanzode and D. R. D. Sarode, "ADVANTAGES AND DISADVANTAGES OF ARTIFICIAL INTELLIGENCE AND MACHINE-LEARNING: A LITERATURE REVIEW," Aug. 2022, publisher: Open Science Framework.

[7] H. Jamali, P. C. Shill, D. Feil-Seifer, F. C. Harris, and S. M. Dascalu, "A Schedule of Duties in the Cloud Space Using a Modified Salp Swarm Algorithm," in *Internet of Things. Advances in Information and Communication Technology*, ser. IFIP Advances in Information and Communication Technology, D. Puthal, S. Mohanty, and B.-Y. Choi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 62–75.

[8] M. Dehghani, Z. Montazeri, E. Trojovská, and P. Trojovský, "Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems," *Knowledge-Based Systems*, vol. 259, p. 110011, Jan. 2023.

[9] M. Nayebi Kerdabadi, A. Hadizadeh Moghaddam, B. Liu, M. Liu, and Z. Yao, "Contrastive Learning of Temporal Distinctiveness for Survival Analysis in Electronic Health Records," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, ser. CIKM '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 1897–1906.

[10] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, p. 104672, Sep. 2021.

[11] H. Jamali and A. Al-Hamood, "A New Method for the Analysis of Misaligned Journal Bearing - ProQuest," *Tribology in Industry*, vol. 40, 2018.

[12] M. E. Hamdi, R. Dezhkam, A. Hajizade, and A. Shamloo, "Prediction of Aqueous Solubility of Drug Molecules by Embedding Spatial Conformers Using Graph Neural Networks," in *2022 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME)*, Dec. 2022, pp. 46–50.

[13] S. Moshtaghi Largani and S. Lee, "Efficient Sampling for Big Provenance," in *Companion Proceedings of the ACM Web Conference 2023*, ser. WWW '23 Companion. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1508–1511.

[14] H. Hoorfar, S. M. Largani, R. Rahimi, and A. Bagheri, "Minimizing Turns in Watchman Robot Navigation: Strategies and Solutions," Aug. 2023, arXiv:2308.10090 [cs].

[15] A. Hadizadeh Moghaddam and S. Momtazi, "A semantic modular framework for events topic modeling in social media," *Multimedia Tools and Applications*, Jun. 2023.

[16] M. Maydanchi, A. Ziaei, M. Basiri, A. N. Azad, S. Pouya, M. Ziaei, F. Haji, and S. Sargolzaei, "Comparative Study of Decision Tree, AdaBoost, Random Forest, Naïve Bayes, KNN, and Perceptron for Heart Disease Prediction," in *SoutheastCon 2023*, Apr. 2023, pp. 204–208, iSSN: 1558-058X.

[17] Kaggle, "Indicators of Heart Disease (2022 UPDATE)," 2022. [Online]. Available: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease

[18] CDC, "CDC Works 24/7," 2020. [Online]. Available: https://www.cdc.gov/index.htm

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.

[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer, 2000.

[21] P. Picton, "What is a Neural Network?" in *Introduction to Neural Networks*, P. Picton, Ed. London: Macmillan Education UK, 1994, pp. 1–12.

[22] T. Rymarczyk, E. Kozłowski, G. Kłosowski, and K. Niderla, "Logistic Regression for Machine Learning in Process Tomography," *Sensors*, vol. 19, no. 15, p. 3400, Jan. 2019, number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[23] M. P. LaValley, "Logistic Regression," *Circulation*, vol. 117, no. 18, pp. 2395–2399, May 2008, publisher: American Heart Association.

[24] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, 2013.

[25] H. Li, Y. Cao, S. Li, J. Zhao, and Y. Sun, "XGBoost Model and Its Application to Personal Credit Evaluation," *IEEE Intelligent Systems*, vol. 35, no. 3, pp. 52–61, May 2020.