Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Improving Cardiovascular Disease Prediction through Comparative Analysis of Machine Learning Models

Nishat Anjum[1], Cynthia Ummay Siddiqua[2], Mahfuz Haider[3], Zannatun Ferdus[4], Md Azad Hossain Raju[5], Touhid Imam[6] and Md Rezwanur Rahman[7]✉

[156]*Department of Computer Science, University of South Dakota, Vermillion, South Dakota, USA*
[2]*Department of Clinical Operations, University of Virginia Physicians Group, USA*
[3]*Department of Clinical Operations Analyst, University of Virginia Physicians Group, USA*
[4]*Department of Science in Information Technology (MSIT), Washington University of Science and Technology, USA*
[7]*Department of Computer Science, University of Colorado Boulder, USA*

Corresponding Author: Md Rezwanur Rahman, **E-mail**: mdra7255@colorado.edu

| ABSTRACT

Cardiovascular diseases, including myocardial infarction, present significant challenges in modern healthcare, necessitating accurate prediction models for early intervention. This study explores the efficacy of machine learning algorithms in predicting myocardial infarction, leveraging a dataset comprising various clinical attributes sourced from patients with heart failure. Six machine learning models, including Logistic Regression, Support Vector Machine, XGBoost, LightGBM, Decision Tree, and Bagging, are evaluated based on key performance metrics such as accuracy, precision, recall, F1 Score, and AUC. The results reveal XGBoost as the top performer, achieving an accuracy of 94.80% and an AUC of 90.0%. LightGBM closely follows with an accuracy of 92.50% and an AUC of 92.00%. Logistic Regression emerges as a reliable option with an accuracy of 85.0%. The study underscores the potential of machine learning in enhancing myocardial infarction prediction, offering valuable insights for clinical decision-making and healthcare intervention strategies.

| KEYWORDS

Machine Learning, Myocardial Infarction, Heart Disease, Coronary Infraction.

| ARTICLE INFORMATION

ACCEPTED: 01 April 2024 **PUBLISHED:** 20 April 2024 **DOI:** 10.32996/jcsts.2024.6.2.7

1. Introduction

The heart, being the body's primary organ responsible for circulating blood to all tissues and organs, holds paramount importance. Any malfunction in its operation can lead to dire consequences, including the shutdown of vital organs like the brain, resulting in rapid fatality. Consequently, cardiovascular diseases have become a focal point in medical research due to their potential lethality and the complex challenges involved in diagnosis and treatment. Factors contributing to the development of cardiovascular diseases include smoking, high cholesterol, family history, obesity, hypertension, and a sedentary lifestyle. Detecting these risk factors often relies on assessing patients' symptoms, which can be influenced by lifestyle choices, workplace stress, and dietary habits, contributing to the increasing prevalence of heart-related disorders worldwide, particularly myocardial infarctions, commonly known as heart attacks.

In response to the critical need for improved prediction accuracy in myocardial infarction, this study delves into the realm of artificial intelligence and machine learning. Leveraging a dataset focused on myocardial disease and established machine learning methodologies, the research aims to illuminate the underlying factors contributing to heart disease development. Through a thorough examination of various machine learning models, including LightGBM, XGBoost, Logistic Regression, Bagging, Support Vector Machine, and Decision Tree, the study seeks to identify the most accurate predictor by conducting a comparative analysis.

The findings of this study not only underscore the potential of machine learning in enhancing myocardial infarction prediction but also offer valuable insights for clinical decision-making and healthcare intervention strategies. The top-performing algorithm, XGBoost, achieves an impressive accuracy of 94.80% and an AUC of 90.0%, closely followed by LightGBM, with an accuracy of 92.50% and an AUC of 92.00%. Additionally, Logistic Regression emerges as a reliable option with an accuracy of 85.0%.

By shedding light on the factors contributing to the development of heart disease and identifying the most accurate predictors, this research contributes to the ongoing efforts to combat cardiovascular diseases effectively. Furthermore, the study opens avenues for future research, including the exploration of advanced technologies like blockchain and deep learning, to further enhance the accuracy and efficiency of prediction models in cardiovascular health management. Ultimately, this study represents a significant step towards advancing cardiovascular health through innovative applications of machine learning and artificial intelligence.

2. Literature Review

Khan et al. conducted a comprehensive comparison of six distinct machine learning models—LightGBM, XGBoost, Logistic Regression, Bagging, Support Vector Machine, and Decision Tree—to predict myocardial disease, yielding satisfactory outcomes with individual accuracy rates of 79.06%, 72.90%, 83.85%, 84.60%, 72.80%, and 82.01%, respectively. Among these models, LightGBM demonstrated superior performance, suggesting its potential for effective myocardial infarction treatment advancement pending further investigation, particularly in healthcare applications. Kayyum et al.'s study involved dataset compilation and the application of Machine Learning Algorithms for data classification, gathering 345 instances with 26 attributes from myocardial infarction patients. They employed the K-Fold Cross Validation Technique to train Bagging, Logistic Regression, and Random Forest algorithms, achieving accuracy rates of 93.913%, 93.6323%, and 91.0145%, respectively. Additionally, Islam et al. addressed the significance of heart disease accuracy by examining a dataset related to myocardial conditions and aiming to predict myocardial infarction using Machine Learning algorithms like K-Means and Hierarchical Clustering, with a dataset of 345 instances and 26 attributes from various hospitals in Bangladesh, seeking insights from historical myocardial infarction cases to evaluate prediction accuracy.

3. Methodology

3.1 Data Collection

All data were gathered manually from diverse clinics and hospitals situated in Dhaka, Bangladesh. During the data collection process, particular emphasis was placed on capturing the recent health conditions of patients afflicted by heart failure. The team compiled the data from multiple clinics and hospitals, and the list of these sources is detailed in Table I. The dataset encompasses a total of 600 instances, each characterized by 13 distinct attributes, as presented in Table II. Within this dataset, information pertaining to specific patients is incorporated, including a class attribute categorized as either "Distinctive" or "Non-Distinctive." Notably, the dataset consists of 12 distinctive attributes and one non-distinctive attribute. This compilation encapsulates treatment records associated with 600 patients affected by heart failure, with each patient profile encompassing 13 distinct clinical traits. These records were gathered over the duration of the patients' treatment journey. The authors visually depicted the correlation matrix in Figure 2

Table 1: The medical institutions involved in data collection.

1.	Square Hospital
2.	National Heart Foundation Hospital
3.	Crescent Hospital
4.	Appolo Hospital
5.	Farida Clinic

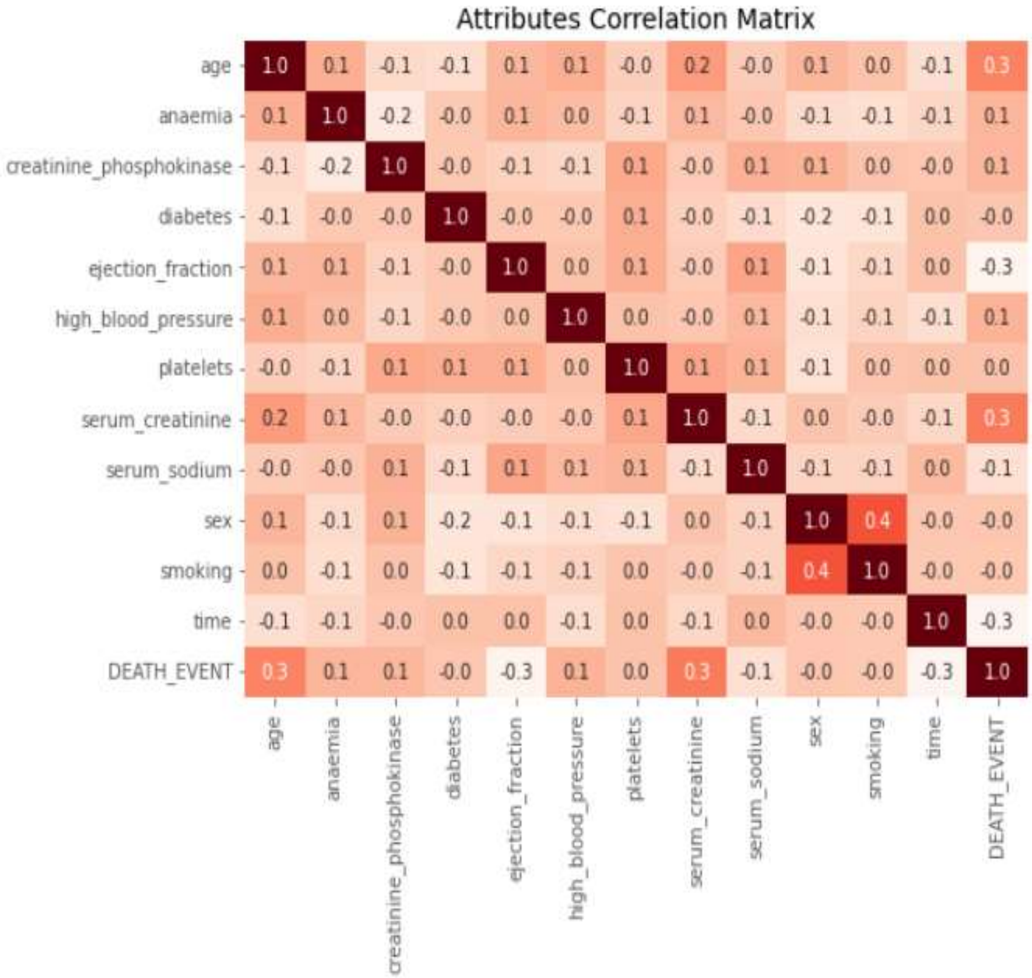

Fig 2: Correlation matrix between dataset attributes

3.2 Data Preprocessing & Filter

We utilized two unsupervised filters during the preprocessing phase using the renowned machine learning tool, Waikato Knowledge Analysis Environment (WEKA 3.8.3). To begin with, we removed the instances with absent items from the dataset, subsequently updating them. In this filtering procedure, we employed the mean, median, and mode to replace missing values in both qualitative and quantitative attributes. As the second step, we employed the Randomly Select filter, which effectively fills in the missing data points without causing substantial speed reductions. Additionally, we utilized the median function, which identifies the middle value within the dataset.

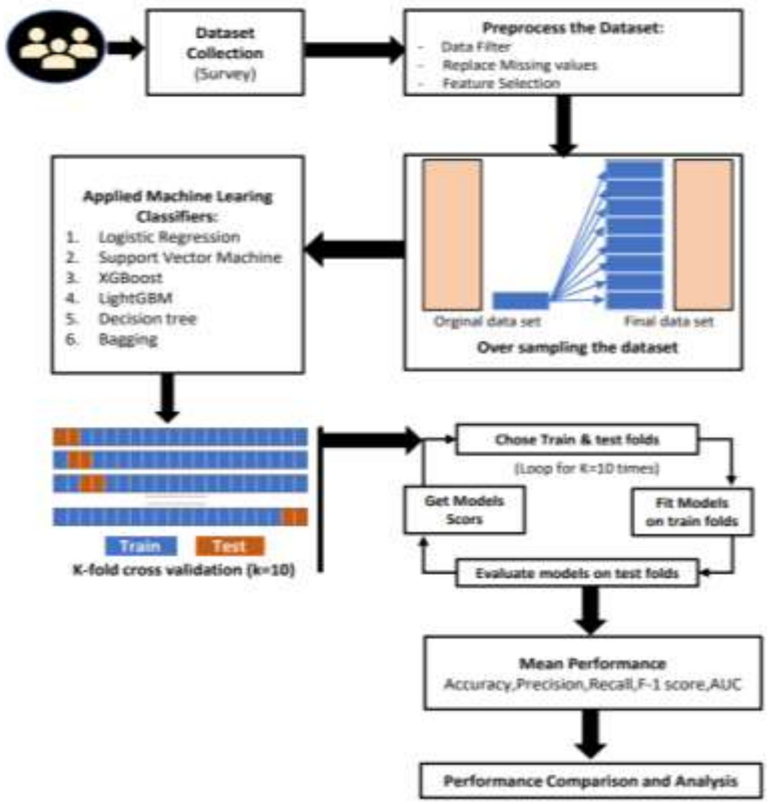

Fig. 1: The overview of our study

3.3 Feature Selection and Validation Technique

Rapid Miner: Rapid Miner finds utility in diverse domains, including academia, training, and research. In our application, this tool was utilized for various tasks, including data preprocessing, visualization of results, model validation, and process optimization. Acknowledged by Gartner as one of the most predictive analytical methods, Rapid Miner swiftly established itself as a leader in the advanced Magic Quadrant Systems Theory. The selection of an appropriate validation technique tailored to specific datasets is crucial. Opting for the most effective approach, the hold-out validation method involves allocating 80% of the dataset for training and 30% for testing. Implementing this method led to favorable outcomes. Metrics such as accuracy, sensitivity, specificity, and F1-Score were all computed using the implied confusion matrix. The presentation of a comprehensive assessment is facilitated through the use of bar graphs, effectively visualizing the performance indicators.

Table 2: Features List (Dataset attribute names)

Attribute names	**Attribute information**
1. **Age**:	The individual's age in years
2. **Sex**	Man or woman (binary)
3. **Diabetes**	Presence of diabetes in the patient (Boolean).
4. **Smoking Habit**	Whether the individual was a smoker or not (Boolean).
5. **Anemia**	Decrease in hemoglobin or red blood cells (Boolean)
6. **Blood Pressure**	If the patient experiences elevated blood pressure (True/False).

7. Creatinine phosphokinase (CPK)	The concentration of CPK enzyme in the blood (mcg/L)
8. Time	Duration of observation (Days)
9. Serum creatinine	Concentration of creatinine in the blood serum (measured in mg/dL)
10. Serum sodium	The concentration of sodium in the bloodstream (measured in mEq/L)
11. Ejection fraction	The proportion of blood expelled from the heart following each contraction expressed as a percentage.
12. Platelets	The existence of blood platelets (platelet count per milliliter)
13. Death event [Target]	At the time of the follow-up, if the patient had deceased (Boolean value).

3.4 Machine Learning Algorithms
After meticulous data preprocessing, rigorous training, and thoughtful categorization, a comprehensive array of machine-learning algorithms was employed in the analysis. These encompassed a diverse set of methodologies such as Logistic Regression, Support Vector Machine, XGBoost, LightGBM, Decision Tree, and Bagging. Following meticulous evaluation, the algorithm demonstrating the most remarkable performance was thoughtfully singled out. In a consolidated presentation, the outcomes of all these algorithms on the dataset are laid bare, offering a comprehensive view of their respective contributions and enabling the identification of the highest-performing algorithm.

3.5 Logistic Regression
In the pursuit of forecasting dependent categorical outcomes, a supervised learning approach known as logistic regression is harnessed. This methodology proves exceptionally valuable when dealing with vast datasets involving regression models. Through this algorithm, the likelihood of specific class probabilities is predicted based on pertinent dependent variables [14]. Mathematically encapsulated within the equation, $y = e^{(b0 + b1x)} / (1 + e^{(b0 + b1x)})$, 'x' represents the input value, 'y' signifies the anticipated outcome, 'b0' denotes the bias or intercept term, and 'b1' stands for the input coefficient.

The efficacy of logistic regression hinges on the Sigmoid function, adept at translating continuous outputs into probabilistic statements between 0 and 1. Elevating precision in this technique involves several pivotal steps: initial library importation, dataset visualization, handling null or missing values, data cleansing by removing extraneous elements, addressing outliers, defining independent and dependent variables, partitioning data into training and testing subsets, leveraging Ensemble and Boosting Algorithms, and engaging in Hyperparameter Tuning.

Furthermore, the study's primary focus involved the assessment of diverse predictive modeling classifications, which entailed the amalgamation of key metrics, including accuracy, precision, recall, the F1-Score, and the area under the curve (AUC). The culmination of this comprehensive analysis is evident in Table III, which effectively summarizes the research findings. The comparative investigation extended across a range of machine learning classifiers, encompassing logistic regression, support vector machines, XGBoost, LightGBM, decision trees, and bagging. The evaluation process meticulously employed metrics such as accuracy, precision, recall, F1-Score, and AUC, with the outcomes of these evaluations thoughtfully compiled and visually presented in Figure 3.

4. Result and Discussion
The investigation uncovered that upholding a standard platelet count enhances the likelihood of survival. Nonetheless, the connection between these factors exhibits minimal strength. Furthermore, sustaining a usual sodium level diminishes the fatality risk after heart failure. Conversely, elevated blood pressure amplifies the likelihood of demise following cardiac failure. The researchers observed that possessing a higher ejection fraction seems to decrease the peril of fatality after cardiac failure, although

due to the limited sample size, it remains impractical to draw any inferences from extreme values. The association between the variables showcases limited strength, except for the link between gender and smoking.

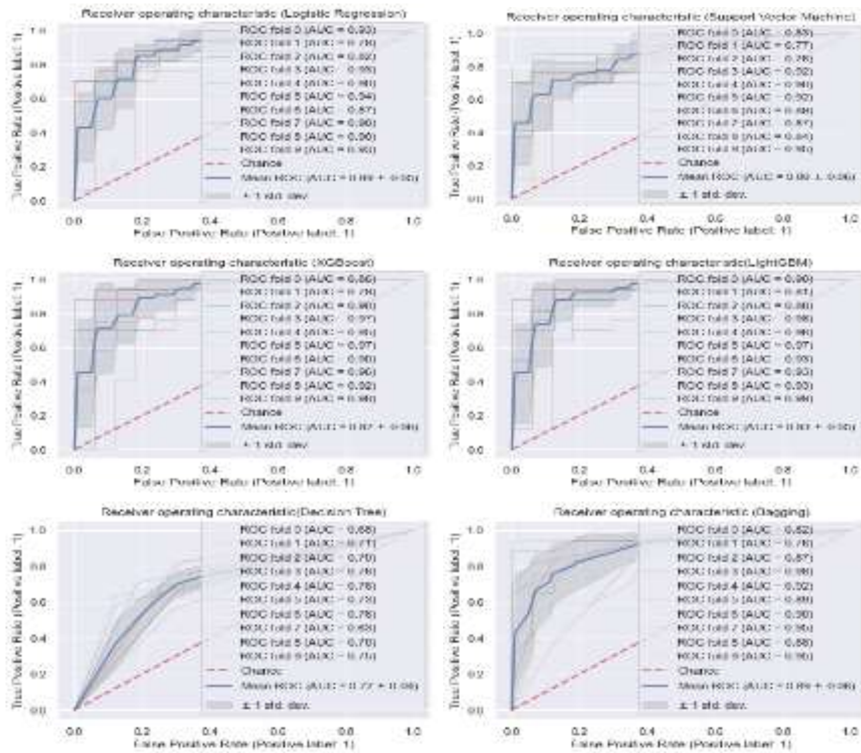

Fig 3**:** Area under curve output of a different kind of machine learning Algorithm

The interpretation of an Area under the Curve (AUC) value holds significance in gauging the performance of predictive models. An AUC of 0.5 signifies a scenario where no differential treatment is evident, indicating an inability to predict a patient's likelihood of cardiovascular disease based on the test outcome. Moving up the scale, an AUC between 0.7 and 0.8 indicates a commendable level of achievement, while a range of 0.8 to 0.9 reflects outstanding results. Remarkably, an AUC surpassing 0.9 reflects an exceptionally high level of performance from the system [13]. To visually capture the performance of specific machine learning models, a 10-fold cross-validation methodology is employed. This comprehensive visual representation encapsulates the comparative prowess of the various machine learning models, providing a clear perspective on their effectiveness and aiding in informed decision-making. The estimation of a patient's vulnerability to cardiovascular disease is reliant on test outcomes. An Area under the Curve (AUC) ranging from 0.7 to 0.8 signifies a fitting level of achievement, while an AUC between 0.8 and 0.9 indicates excellent results. Moreover, an AUC exceeding 0.9 demonstrates exceptional system performance [13]. The evaluation of the designated machine learning models, facilitated by 10-fold cross-validation, is visually depicted in Figure 3. This graphical representation encompasses AUC plots along with average outcomes.

Table 3: Outcomes of different machine learning classifiers

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Logistic Regression	85.0	84.00	79.10	78.00	87.0
SVM	78.0	73.33	80.20	76.50	87.80
XGBoost	94.80	88.00	86.0	84.60	90.0
LightGBM	92.50	87.40	82.80	85.00	92.00
Decision Tree	81.40	76.80	73.56	74.66	78.0
Bagging	84.0	87.50	76.88	76.00	87.0

In Table 3 and Chart 1, we provide an insightful analysis of various models assessed for their efficacy in classification tasks. Among the diverse array of models evaluated, XGBoost emerges as a clear frontrunner, showcasing an impressive accuracy rate of 94.80%. This powerhouse model not only excels in overall correctness but also demonstrates remarkable precision, accurately identifying 88.00% of positive predictions. Moreover, XGBoost achieves a high recall rate of 86.0%, thereby ensuring minimal false negatives and effectively capturing true positives. Its exceptional F1 score of 84.60% signifies a harmonious balance between precision and recall, further solidifying its prowess in classification endeavors.

However, while XGBoost shines brightly across multiple metrics, it's imperative to consider alternative contenders. LightGBM emerges as a formidable competitor, closely trailing XGBoost with an accuracy rate of 92.50% and an impressive AUC of 92.00%. This model showcases robust performance across precision, recall, and F1 score, rendering it a compelling choice for diverse classification tasks.

Furthermore, Logistic Regression proves to be a reliable and pragmatic option, consistently delivering solid results across various metrics. With an accuracy rate of 85.0% and a respectable AUC of 87.0%, Logistic Regression offers a practical solution for classification tasks where interpretability is paramount.

While SVM and Bagging demonstrate competitive performance, they fall slightly short compared to the leading models of XGBoost, LightGBM, and Logistic Regression. SVM exhibits a decent accuracy rate of 78.0% and a commendable AUC of 87.80%; however, its precision and F1 score lag behind the top-performing models. Similarly, Bagging presents respectable accuracy and precision but trails in recall and F1 score, highlighting the nuanced trade-offs among different model choices in classification tasks.

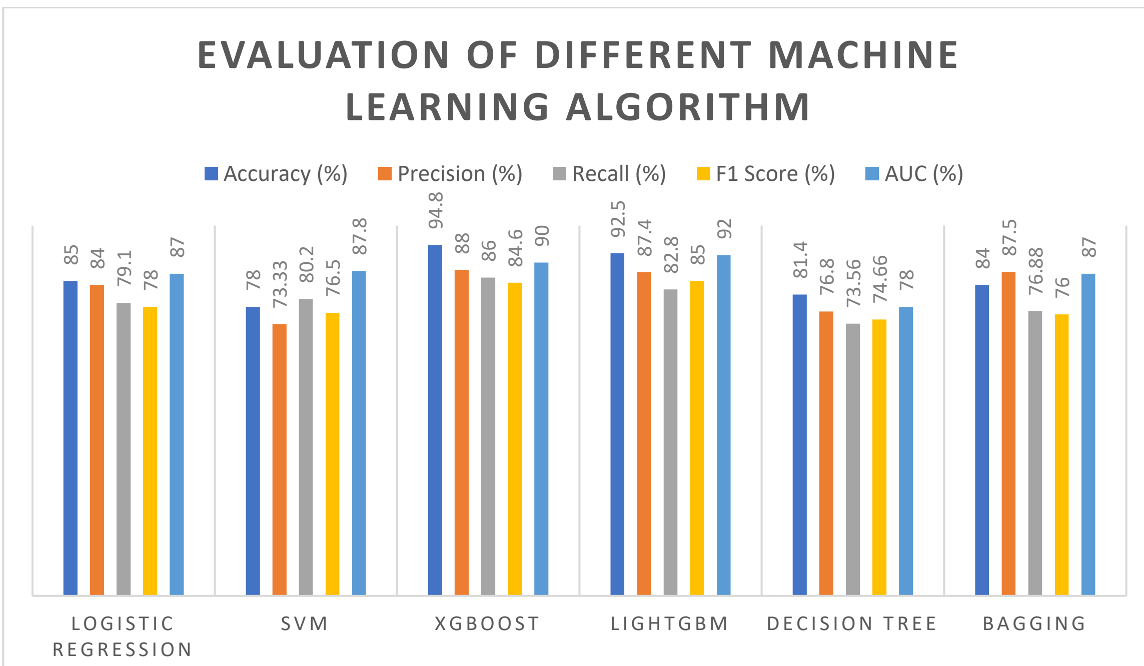

Chart 1. Analysis of the different machine learning models

5. Conclusion and Future Work

With each passing day, the incidence of myocardial disease is on the rise. The positive aspect lies in the ability of machine learning to detect this ailment, offering a streamlined diagnostic approach. Our suggested model proves to be both effective and occasionally efficient in healthcare, enabling the early detection of this condition. The escalating incidence of myocardial disease has prompted a promising avenue for early detection through compact machine-learning models. Our proposed model stands as a potent tool in expediting diagnosis, ensuring timely intervention, and enhancing healthcare outcomes. Its efficacy extends to remote and underserved areas, where cost-effective diagnosis systems become invaluable. By leveraging machine learning's capacity to analyze extensive datasets of similar cases, our model offers reliable predictions. Amidst a plethora of myocardial disease prediction projects, our approach, notably XGBoost, achieved exceptional results, boasting 90.60% accuracy, 87.00% precision, 81.0% recall, 83.90% F1 score, and a remarkable 91.0% AUC. While promising, the widespread implementation of our model necessitates further research and development. Addressing challenges related to big data and exploring innovative solutions like blockchain can truly revolutionize cardiovascular health forecasting and pattern analysis.

In a landscape marked by an alarming rise in myocardial disease cases, the integration of machine learning emerges as a beacon of hope. Our proposed model demonstrates significant potential for early detection, offering a streamlined and effective diagnosis approach. These holds promise not only in established healthcare settings but also in underserved rural areas, ushering in a new era of accessibility. Through meticulous evaluation, our model, particularly exemplified by the impressive performance of XGBoost, underscores its accuracy and predictive power. The journey, however, is far from over. To fully harness the impact of our model, rigorous research and innovation are imperative. The intersection of cutting-edge technology, informed research, and a commitment to healthcare advancement could propel cardiovascular health forecasting into uncharted territories.

The path forward involves multifaceted exploration. Amidst the proliferation of machine learning models, our focus remains steadfast on refining and expanding our approach. We envision the utilization of distinct vocal feature-based datasets in conjunction with deep learning models, broadening the spectrum of myocardial disease detection. Tapping into advanced technologies, such as blockchain, could address data challenges and enhance the reliability of predictions. Moreover, the quest to predict cardiovascular disease recovery stands as an imminent challenge. Navigating this perilous territory demands the synthesis of predictive analytics, medical expertise, and technological innovation. As the healthcare paradigm transforms, our commitment to innovation and advancement remains unwavering, propelling us towards a future where proactive cardiovascular health management becomes a cornerstone of well-being.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Bristi, U., Rajib, M. I., & Shoukat, S. (2020). Implementation of corporate social responsibility in the ready-made garment industry of Bangladesh in order to achieve sustainable development goals: a study on Bangladeshi garment manufacturers. *Journal of Textile Science and Technology*, *6*(4), 232-246.

[2] Bhuiyan, M. S., Chowdhury, I. K., Haider, M., Jisan, A. H., Jewel, R. M., Shahid, R., & Siddiqua, C. U. (2024). Advancements in Early Detection of Lung Cancer in Public Health: A Comprehensive Study Utilizing Machine Learning Algorithms and Predictive Models. *Journal of Computer Science and Technology Studies*, *6*(1), 113-121.

[3] Cao, D. M., Sayed, M. A., Islam, M. T., Mia, M. T., Ayon, E. H., Ghosh, B. P., & Raihan, A. (2024). Advanced Cybercrime Detection: A Comprehensive Study on Supervised and Unsupervised Machine Learning Approaches Using Real-world Datasets. *Journal of Computer Science and Technology Studies*, *6*(1), 40-48.

[4] Cao, D. M., Amin, M. S., Islam, M. T., Ahmad, S., Haque, M. S., Sayed, M. A., & Nobe, N. (2023). Deep Learning-Based COVID-19 Detection from Chest X-ray Images: A Comparative Study. *Journal of Computer Science and Technology Studies*, *5*(4), 132-141.

[5] Chowdhury, O. S., & Baksh, A. A. (2017). IMPACT OF OIL SPILLAGE ON AGRICULTURAL PRODUCTION. *Journal of Nature Science & Sustainable Technology*, *11*(2).

[6] Ghosh, B. P., Bhuiyan, M. S., Das, D., Nguyen, T. N., Jewel, R. M., Mia, M. T., & Shahid, R. (2024). Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures Across NSE and NYSE. *Journal of Computer Science and Technology Studies*, *6*(1), 68-75.

[7] Hassan CAU CAU, Iqbal J J, Irfan J, Hussain S, Algarni AD, Bukhari SSH, Alturki N and Ullah SS (2022) Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers", Sensors (Basel), Vol. 22, No. 19, pp:7227, DOI: 10.3390/s22197227, 2022.

[8] Haque M. S., Amin M. S., Ahmad S., Sayed M. A., Raihan A. and Hossain M. A., (2023) Predicting Kidney Failure using an Ensemble Machine Learning Model: A Comparative Study, *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Palembang, Indonesia, 2023, pp. 31-37, doi: 10.1109/EECSI59885.2023.10295641.

[9] Haque M. S., Taluckder M. S., Bin S S. Shahriyar M. A., Sayed M. A. and Modak C., (2023) A Comparative Study of Prediction of Pneumonia and COVID-19 Using Deep Neural Networks, *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Yogyakarta, Indonesia, 2023, pp. 218-223, doi: 10.1109/ICE3IS59323.2023.10335362.

[10] Islam, M. T., Ayon, E. H., Ghosh, B. P., MD, S. C., Shahid, R., Rahman, S., ... & Nguyen, T. N. (2024). Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce. *Journal of Computer Science and Technology Studies*, *6*(1), 33-39.

[11] Islam, M. M., Nipun, S. A. A., Islam, M., Rahat, M. A. R., Miah, J., Kayyum, S., ... & Al Faisal, F. (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. In *Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2* (pp. 120-130). Springer Singapore.

[12] Islam, M.M. *et al.* (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. In: Bhattacharjee, A., Borgohain, S., Soni, B., Verma, G., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_11

[13] Khan R. H., Miah J., Abed N S. A. and Islam M., (2023) A Comparative Study of Machine Learning classifiers to analyze the Precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp.0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.

[14] Kayyum S. et al., (2020) Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-7, doi: 10.1109/ICCCI48352.2020.9104104.

[15] Krittanawong, C., Virk, H.U.H. and Bangalore, S. *et al.* (2020) Machine learning prediction in cardiovascular diseases: a meta-analysis. *Sci Rep* **10**, 16057 (2020).

[16] Kayyum S. *et al*., (2020) Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, *2020 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2020, pp. 1-7, doi: 10.1109/ICCCI48352.2020.9104104.

[17] Khan R. H., Miah J., Abed N S. A. and Islam M. (2023) A Comparative Study of Machine Learning classifiers to Analyze the precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.

[18] Khan R. H., Miah J., Rahman M. M. and Tayaba M., (2023) A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, 647-652, doi: 10.1109/CCWC57344.2023.10099106.

[19] Miah J., Ca D. M., Sayed M. A., Lipu E. R., Mahmud F. and Arafat S. M. Y. (2023) Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," *2023 15th International Conference on Innovations in Information Technology (IIT)*, Al Ain, United Arab Emirates, 2023, pp. 49-54, doi: 10.1109/IIT59782.2023.10366476.

[20] Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. Y. (2023, November). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. In *2023 15th International Conference on Innovations in Information Technology (IIT)* (pp. 49-54). IEEE.

[21] Modak C., Shahriyar M. A., Taluckder M. S., Haque M. S. and Sayed M. A., (2023) A Study of Lung Cancer Prediction Using Machine Learning Algorithms," *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Yogyakarta, Indonesia, 2023, pp. 213-217, doi: 10.1109/ICE3IS59323.2023.10335237.

[22] Mandrekar J, (2010) Receiver Operating Characteristic Curve in Diagnostic Test Assessment, *Journal of Thoracic Oncology,* 5, 9), 1315-1316, DOI: 10.1097/jto.0b013e3181ec173d, 2010

[23] Maalouf M, (2011). Logistic regression in data analysis: an overview", International Journal of Data Analysis Techniques and Strategies, 3, 3, 281-299, 2011

[24] Nath, F., Asish, S., Debi, H. R., Chowdhury, M. O. S., Zamora, Z. J., & Muñoz, S. (2023, August). Predicting hydrocarbon production behavior in heterogeneous reservoir utilizing deep learning models. In *Unconventional Resources Technology Conference, 13–15 June 2023* (pp. 506-521). Unconventional Resources Technology Conference (URTeC).

[25] Nath, F., Chowdhury, M. O. S., & Rhaman, M. M. (2023). Navigating Produced Water Sustainability in the Oil and Gas Sector: A Critical Review of Reuse Challenges, Treatment Technologies, and Prospects Ahead. *Water*, *15*(23), 4088.

[26] Puja, A. R., Jewel, R. M., Chowdhury, M. S., Linkon, A. A., Sarkar, M., Shahid, R., ... & Sarkar, M. A. I. (2024). A Comprehensive Exploration of Outlier Detection in Unstructured Data for Enhanced Business Intelligence Using Machine Learning. *Journal of Business and Management Studies*, *6*(1), 238-245.

[27] Sujatha P. and Mahalakshmi K, (2020) Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease, 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-7, Doi: 10.1109/INOCON50539.2020.9298354.

[28] Shakeel, B P., (2018) Cloud-based framework for the diagnosis of diabetes mellitus using K-means clustering, 6, 1, 2018

[29] Singh P, Singh S, Pandi-Jain G S, (2018). Effective heart disease prediction system using data mining techniques, *International journal of nanomedicine*, 13, 121–124. DOI:10.2147/IJN.S124998, 2018

[30] Shoukat, S. (2016). Now and then: The neckline history of women. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, *26*(2), 33-52.

[31] Shoukat, S., & Rabby, S. (2017). A Total Journey of Footwear with Material Analysis. *Journal of Scientificand Engineering Research*, *4*(6), 187-191.

[32] Shoukat, S., & Islam, M. A. (2016). Development of Wedding Costume from Bengal to Bangladesh.

[33] Shoukat, S., & Kwon, W. S. (2024, January). Effects of AI Agent Anthropomorphism on Consumers' Affective, Cognitive, and Social Shopping Experiences. In *International Textile and Apparel Association Annual Conference Proceedings* (Vol. 80, No. 1). Iowa State University Digital Press.

[34] Yazdani A, Varathan K D and Chiam Y K (2021) A novel approach for heart disease prediction using strength scores with significant predictors, BMC Med Inform Decis Mak, 21, 194, 2021.