

Article

Not peer-reviewed version

Heart Disease Prediction Redefined: Hybrid CNN & MobileNet Model Comparison

[Shubham Malhotra](#)*, Meenu Gupta, [Muhammad Saqib](#), Fnu Yashu

Posted Date: 9 May 2025

doi: 10.20944/preprints202505.0715.v1

Keywords: Classification; Deep Learning; Machine Learning; Heart Disease; Hyperparameter Optimization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Heart Disease Prediction Redefined: Hybrid CNN & MobileNet Model Comparison

Shubham Malhotra ^{1,*}, Meenu Gupta ², Muhammad Saqib ³ and Fnu Yashu ⁴

¹ Rochester Institute of Technology, Rochester, NY, USA, 14623

² Department of Computer Science & Engineering, Chandigarh University, Punjab, India

³ Department of Computer Science, Texas Tech University, Lubbock, TX, USA

⁴ Department of Computer Science, Stony Brook University, Stony Brook, NY, USA

* Correspondence: shubham.malhotra28@gmail.com

Abstract: As reported by the World Health Assembly, heart problems, the majority of mortality globally, accounted for around 17.9 million cases in 2019, or over 32% of death rates worldwide. The need for early heart illness detection drives research into establishing effective machine learning-based diagnostic systems. At the present, the growth of machine learning in healthcare largely supports the ability of analysts or physicians to carry out their jobs, spot healthcare trends, and generate illness prediction models. Deep learning, on the other hand, has advanced quickly and emerged as the most widely used technique in recent years, including for illness detection. Optimizing a combination of convolutional neural network (CNN) as well as Mobilenet approach for cardiac disease classification is the major objective of this project. Through utilizing sophisticated DL methods such as LSTM, SVC+Rf as Hybrid model and CNN+Mobilenet Algorithm, we were able to attain an impressive testing accuracy that reached 98.54%. CNN+Mobilenet By integrating the structured datasets with modern DL algorithms, this work unveils a unique solution to the challenge of expandable and accurate discoveries of the heart sickness, resulting in the power to enhance the clinical results while decreasing death rates.

Keywords: classification; deep learning; machine learning; heart disease; and hyperparameter optimization

I. Introduction

The World Health Organization indicated that in 2019, 179,000,000 hospitalizations globally were linked to CVD, contributing to 32% of all fatalities. Of those fatalities, 85% were triggered by strokes and cardiovascular disease. Over 75% of CVD-attributable mortality occurs in low- to moderate-socioeconomic nations. Furthermore, three-quarters of the 17 million premature deaths from non-communicable diseases in 2019 were linked to CVDs. The spiraling occurrence of heart-related diseases offers a noteworthy issue, needing precise therapy [1].

Biomarkers, characterized as all forms of biological indicators, serve a key role in patients' prompt diagnoses and survivability. However, countless data are established every day in the medical care business. In the lack of contemporary technologies, early diagnosis and recovery from sickness have become challenging jobs. Therefore, sophisticated strategies are required to make prompt assessments and illness survival utilizing this enormous volume of data. Cognitive approaches are in the area of DL [2,3].

In subsequent years, deep learning has played an imperative role in the recognizing CVDs. DL procedures are a sort of ANN that can acquire and observe sequences in information, rendering them ideally suited for processing big and complicated datasets typically encountered in medical imagery and diagnostic tests. Deep learning has confirmed important in finding and predicting coronary problems using diagnostic images such as cardiac echocardiograms, coronary arteries, and MRI scans. Deep learning algorithms may find out how to acknowledge certain habits and characteristics

in these pictures that may signal the existence of cardiac illness, notably anomalies in heart structure, circulation flow, and muscle texture [4,5].

According to the Sustainable Development Goal (SDG) of the United Nations, which states that all people ought to remain well and happy, this study analyzes cardiovascular diseases. Cardiovascular illness is generally diagnosed through analyzing an individual's sensations and doing an exercise test. The manifestations attributed to CVD involve physical brittleness, difficulty of airflow, irritation in the feet, weakness, and other concomitant indications. Cardiovascular disease (CVD) is a catastrophic global health concern that may be attached to many put-at-risk factors, including hypertension, excess lipid levels, smoking, sluggish habits, and obesity [6,7].

Machine learning is substantially employed in healthcare diagnoses and the medical business. Machine learning includes numerous uses in the medical industry, spanning medication discovery, therapeutic diagnostics, widespread estimation, and heart failure forecasts. Machine learning algorithms could find out characteristics from huge biomedical data and do forecast analysis. Machine learning delivers various value contrasts to classical medical procedures, ranging from reducing time and money, which helps strengthen prognosis [8–10].

The prompt identification of heart illness attempts to lower death rates by examining the person's existing cardiac status together with risk factors. Contemporary strategies usually rely on initially processed information with constrained characteristics, making the diagnosis and categorization of heart illness an especially difficult issue. A full approach important owing to the condition's intricacy, encompassing holistic health assessments, systematic symptom inspections, threat assessment for factors, and detailed screening tests. Finally, the restricted relevance of routine data sets in situations that are real considerably hampers successful detection and identification. Further, constrained use of conventional data sets actual scenarios worsens the difficulty of exact detection and categorization [11,12].

Considering combining elements of cardiopulmonary diagnosis, categorization, and even prospects, the research's unique approach enables individualized illness categorization and prognosis. A challenging model employing detectors from CNN+Mobilenet, CNN, and SVC+RF, coupled with producing certain data sets like CVD, illustrates the feasibility of reliable forecasts.

In order to summarize the work we did in this paper:

- We acquired CVD datasets from the Mendeley Data and grouped them into several classes in this research.
- We compared the efficacy of our suggested techniques.
- To compare our suggested models to those that presently exist.

II. Literature Review

Manasaleh Al Reshan et al. [13] offers an upgraded HDNN technique for heart disease (HD) forecasting combining deep ANN, LSTM, CNN, as well as hybrid CNN-LSTM models. The additional tree encoder assists in choosing characteristics, and data imputed assures the integrity of the information. The algorithm had been refined on Cleveland and a complete HD dataset. CNN-LSTM obtained 97.75% effectiveness on Cleveland and 98.86% on the whole dataset. The model suggested improves standard ML approaches in precision, reliability, sensitivity, MCC, specificity, then F1-score, and AUC.

Senthilkumar Mohan et al. [14] utilized many classification algorithms on the Institution of California Irvine (UCI) repository's CVD datasets. The researchers have picked characteristics in the solution they suggested using an inward selection approach. Eleven crucial traits were picked utilizing this procedure, and those characteristics were then utilized to train algorithms like SVM, prune Decision Tree (DT) methods, and logistic regression. The highest overall effectiveness across all the scenarios, 87.1%, was reached by applying logistic regression.

Numerous research studies focused on heart disease prediction employing ML techniques. DBSCAN is being requested for identifying outliers, while SMOTE-ENN assists in the balancing data

sets that are imbalanced, enhancing a accuracy of models. XGBoost-based systems have exhibited outstanding efficiency, obtaining up to 98.40% reliability. Previous work has also included cardiovascular condition forecasting techniques into clinical decision-making systems (HDCDSS), leveraging MongoDB used for effective health care information administration. The next phase involves evaluating various data-gathering methodologies like outlier detection gets nearer, and cutting-edge technology for better medical decision-making [15].

Jian Ping Li et al. [16] created ML-based cardiac disorder detection systems utilizing a classifications including LR, K-NN, ANN, SVM, NB, and DT. The choice of the feature tactics, notably Relief, MRMR, LASSO, AN LLBFS, as well as novel FCMIM, has been built to boost the rate of classification. The Cleveland heart disease data had been frequently utilized for modeling assessment, with SVM achieving 92.37% performance using FCMIM. The study underlines the relevance of choosing key parameters like thallium, as well as scan and exercise-induced angina, while removing less crucial ones like FBS.

Senthilkumar Mohan et al. [17] underlines the necessity to analyze raw medical information to accelerate cardiovascular disease identification and decrease the mortality. Machine learning approaches had become extensively employed, with a combination of HRFLM, which blends a Random Forest (RF) and Linear Classifier (LM), showing highly precise forecasting. Research demonstrates utilizing actual data sets rather than simulated boosts a practical applicability. The next phase includes investigating varied ML combos and creating sophisticated algorithms for the feature selection to increase prediction accuracy. These advances swear to give greater insight into critical aspects for improved cardiovascular disease detection.

Abdallah Abdellatif et al. [18] concentrated on determining to increase the RF classifier's learning time and performance. They depend on the following techniques: distinctly segregating the training data sets; producing RF base decision trees employing split measures or numerous characteristic assessments and using corresponding voting instead of a significant amount of voting; generating the most widely spread classifiers using an extensive range of bootstrap datasets; and employing the dynamic programming method to identify the most prominent part of Random Forest (RF).

The literature demonstrates innovations in cardiac disease prognosis using machine learning as well as deep learning, increasing reliability as well as accuracy. Hybrid models like CNN-LSTM and HRFLM surpass standard approaches, obtaining over 98% accuracy. Feature choice approaches (Relief, MRMR, FCMIM) promote categorization, where as SMOTE-ENN and DBSCAN optimize the quality of the data. Incorporating modeling into a medical decision-making system (HDCDSS) promotes implementation in healthcare.

III. Proposed Methodology

"Figure 1" illustrates a detailed depiction of the creation of the suggested research structure as a method. The image below depicts a comprehensive grasp of the foundation and components of the proposed structure. The accompanying graphic provides a comprehensive elucidation of the many elements and architecture of the proposed framework.

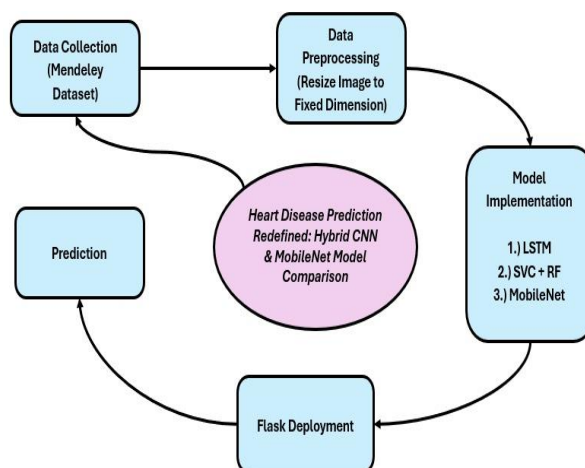


Figure 1. System Diagram for Heart Disease.

A. Dataset Gathering

The ECG data utilized in present research came from the Mendeley Data, which includes an electrocardiogram (ECG) sample grouped by three unique categories: normal, anomalous, and disease-specific ventricular signals. The dataset was assembled using an assortment of individuals, including both healthy people and patients with varied cardiac problems. The data-collecting strategy required getting reliable ECG recordings using regular medical-grade ECG equipment. The recordings are cleaned and annotated by the healthcare personnel to ensure accuracy and reliability. The dataset offers similar amount of ECG pictures in all three groups, which makes it a viable candidate for constructing machine learning algorithms for continual health monitoring and forecasting cardiac illness. The data collecting is a vital resource for the construction and testing of AI-driven medical diagnostics for medical reasons, enabling the early evaluation and classification of cardiac abnormalities. The impacted and non-affected individuals shown in “Figure 2” have been included in the dataset. The ethical oversight Committee of each institution issued legal authority for the data extraction. “Figure 3” illustrates the heatmap of our Mendeley Dataset.

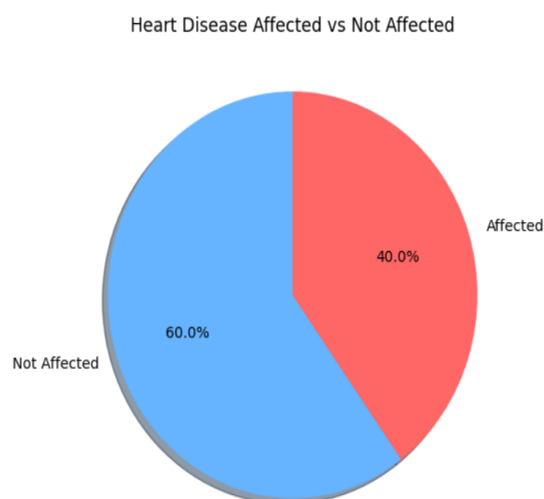


Figure 2. Classification between HD Affected VS. Not Affected.

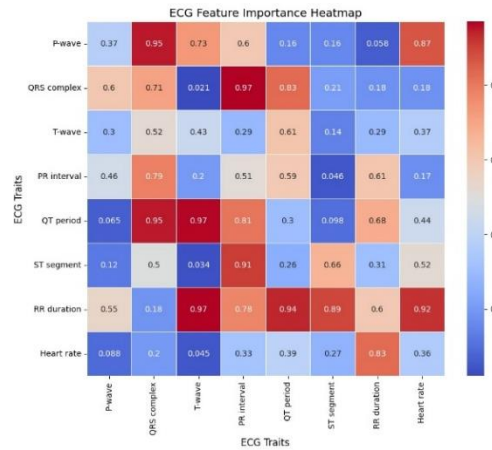


Figure 3. Heatmap for Mendeley Dataset.

B. Feature Selection

ECG traits are essential to foresee cardiac problems. The P-wave helps diagnose heart rate variability, whereas the QRS complex reveals ventricular irregularities. The T-wave, PR interval, QT period, and ST segment suggest ischemia, AV blockages, arrhythmias, and a myocardial infarction. The RR duration and heart rate assist measure abnormalities in heart rhythms and general cardiac health.

C. ML Model Selection

A variety of ML algorithms were used in our experimental study, including CNN+MobileNet, LSTM, SVC+RF.

1) *CNN+MobileNet*: The CNN+MobileNet combo which integrates CNN's deep extraction of the features with MobileNet's compact performance of allowing speedy and exact classification. CNN encodes spatial connections using convolutional layers, while Mobile Net minimizes computing cost utilizing depth wise segmented convolutions. The combination strategy enhances a efficacy while it boosting a speed and resource efficiency. The model learns hierarchical feature illustrations, ensuring good classification of new data. The "Equation (1)" illustrates the entire procedure for this convolution process is stated as, where W gathers attributes and adjusts activation.

$$y = f(W * x + b) \quad (1)$$

"Figure 4" explain the training as well as validation accuracy of the proposed hybrid model.

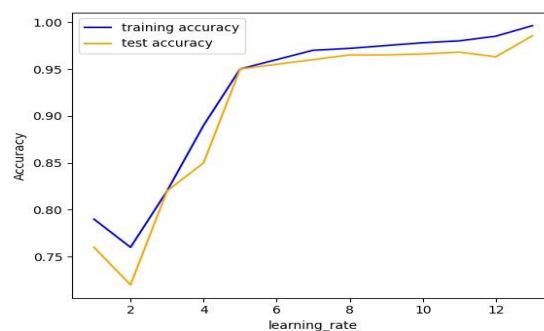


Figure 4. Training and Validation Accuracy of the Proposed Hybrid Model.

2) *Long-Short Term Memory (LSTM)*: A LSTM networks, which is a form of recurrent neural network (RNN) intended to handle sequential input by collecting a long-term dependency. When used to the ECG picture datasets, ECG signals are the first translated into a sequence of pixel-based data. The LSTM model it learns temporal patterns by examining fluctuations in pixel intensity over time, detecting characteristics such as P-waves, QRS complexes, and T-waves. By analyzing ECG pictures as a time-series data, LSTM successfully identifies anomalies such arrhythmias and myocardial infarction. This is strategy enhances accuracy in ECG classification by utilizing a sequential dependencies in image-based signal.

$$y = \text{Softmax}(W \cdot \text{LSTM}(\text{Seq}(X), \theta)^T + b) \quad (2)$$

The "Equation (2)" illustrates the entire procedure of transforming ECG pictures into consecutive information, analyzing them with an LSTM, and categorizing the data for anomaly detection.

3) *SVC + RF*: The SVC+RF hybrid method combines Random Forest (RF) as well as Support Vector Classifier (SVC) enhance the classification accuracy. RF, is a ensemble learning method, builds multiple decision trees as well as aggregates their outputs to reduce overfitting and also it improves a generalization. SVC, on the other hand, finds the optimal hyperplane that maximizes margin between different classes, making it effective for a high-dimensional data. In this hybrid approach, RF is used for feature selection or initial classification, and its refined feature set is then passed to SVC for final decision-making. This combination leverages a RF's robustness in the handling complex data distributions and SVC's efficiency in finding precise decision boundaries, leading improved classification performance. The decision function is defined as:

$$y = \text{sgn}(\lambda \cdot f_{\text{SVC}}(x) + (1 - \lambda) \cdot f_{\text{RF}}(x)) \quad (3)$$

"Equation (3)" is the kernels functions in SVC, and α_i are the Varangian multiplier that determine the support vectors. The combined RF+SVC model is frequently used in health care diagnosis, picture acknowledgment, and other categorization applications demanding high precision as well as durability.

The resulting algorithms are included into an intuitive web-based app developed with Streamlit cloud, facilitating users to enter pertinent health parameters and obtain the immediate estimates. The program is hosted on Streamlit Community Cloud, rendering it is a readily available without the necessity for local installations. Real-time benchmarks for performance between the Hybrid CNN and MobileNet models are presented, emphasising disparities in precision, rapidity, and utilisation of resources. This implementation transforms the simplicity and openness in a predictive medical software.

IV. Result and Discussion

This work's findings of the experiment were acquired utilizing a CVD dataset. The research effort focuses analyzing the usefulness of multiple DL algorithms in classifying and forecasting results throughout this area. A hybrid approach comprising numerous ML and DL techniques was built, and the achievement rates of these techniques are given in "Figure 5". The ML and DL algorithms are utilized for a research were CNN+Mobilenet, LSTM, and SVC+RF. "Table I" offers in the depth assessment of the accuracy rates attained by a each strategy, clearly it revealing greater efficiency of the CNN+Mobilenet. This study highlights the potential for the hybrid optimization tactics in enhancing the dependability and precision of computational models in crucial areas such as CVD prediction.

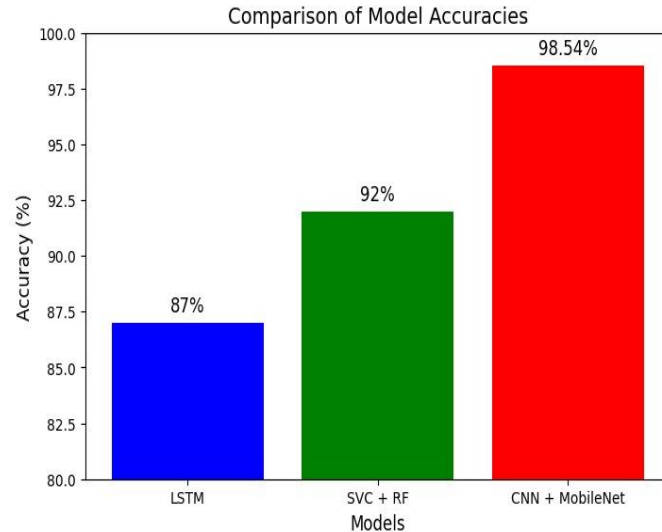


Figure 5. Algorithm Comparison for Heart Disease.

Our key contribution to the initiative is developing the real-world cardiovascular disease datasets, notably the CVD Dataset. The CVD dataset, built for detecting patterns CVD categories, shows high accuracy in CNN+Mobilenet. The CVD dataset, including unaffected persons, provides binary classification. CNN+Mobilenet, apparently owing to increasing complexity from undamaged occurrences. The CVD information set, comprising afflicted vs unaffected individuals, provided the maximum efficiency in CNN+Mobilenet 98.54% and Lstm at 87%. Its proportional modeling helps to capture delicate correlations, making it suited for the two separate binary and multi-class scenarios. Amongst all datasets, Rf+Svc did well with structured information; nevertheless CNN+Mobilenet excelled in managing complexity, which underscores the relevance of tailoring models to dataset characteristics.

Table I. Heart disease Accuracy classification.

Techniques	Accuracy (%)
CNN + MobileNet	98.54%
LSTM	87.00%
SVC + RF	92.00%

A. Performance Evaluation Using Confusion Matrix

The confusion matrix illustrates the variance among real and expected values in a dataset. As seen in "Figure 5", it acts as a fundamental tool for analyzing the performance of a classification techniques in ML. The sections of the confusion matrix include.

- **Correct Positive:** The model properly detects a positive case when the real value is positive.
- **Incorrect Positive:** A erroneous prediction when the model labels a negative situation as positive.
- **Incorrect Negative:** A misclassification when a positive case is incorrectly classified as negative.
- **Correct Negative:** The model successfully predicts a negative result when the actual value is negative.

Additionally, critical evaluations of metrics, including accuracy, memory, recall, as well as the F1 score, are created using a confusion matrix to determine an efficacy of the model. Efficacy is one

of the most regularly utilized metrics, measuring a proportion of the right predictions across every category. It is calculated using the following "Equation (4)":

$$accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

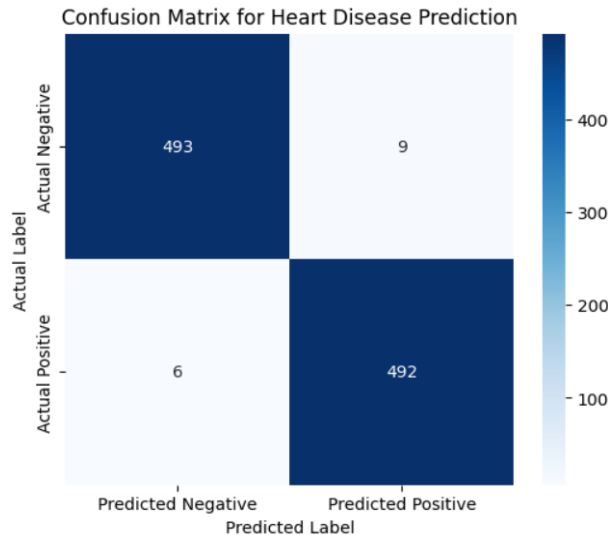


Figure 6. Confusion Matrix for Heart Disease.

V. Conclusion

The severity of cardiac illness may be assessed using a hybrid model using deep learning technologies. The results of our experiments demonstrate that CNN+Mobilenet achieved superior accuracy (98.54%) compared to ML classifiers like SVC+RF (92%) and other DL models such as LSTM (87%). As the dataset evolves, the algorithm has effectively evaluated the acquired content and accurately classified it. This study involves the customization, training, and assessment of three algorithms using three distinct models: CNN+Mobilenet, LSTM, and SVC+RF, both with and without hyperparameter optimization techniques. Their accuracies are juxtaposed with the existing methodologies. Substantially altering the specs of CNN+MobileNet resulted in a testing accuracy of 98.54%. The CNN classifier with MobileNet achieves the highest testing accuracy of the three approaches, at 98.54%. The CNN classifier achieves a testing accuracy of 89% in the absence of MobileNet. The CNN with MobileNet is the best hyperparameter for accuracy. Future research will focus on using diverse ML techniques as well as improved feature selection methods to boost the accuracy and efficacy of heart disease prediction. To make prediction systems work better, optimization methods will be used. The projects will also include parts for after the diagnosis, like managing and treating cardiovascular diseases.

References

1. Gabriel, J. Jasmine, and L. Jani Anbarasi. „Accurate Cardiovascular Disease Prediction: Leveraging Opt_hpLGBM With Dual-Tier Feature Selection.” *IEEE Access* (2024).
2. Ay, Şevket, Ekin Ekinci, and Zeynep Garip. „A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases.” *The Journal of Supercomputing* 79.11 (2023): 11797-11826.
3. Jaiswal, Arunima, Monika Singh, and Nitin Sachdeva. „Empirical analysis of heart disease prediction using deep learning.” *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. IEEE, 2023.

4. Ghorashi, S., Rehman, K., Riaz, A., Alkahtani, H. K., Samak, A. H., Cherrez-Ojeda, I., & Parveen, A. (2023). Leveraging regression analysis to predict overlapping symptoms of cardiovascular diseases. *IEEE Access*, *11*, 60254-60266.
5. Obayya, M., Alsamri, J. M., Al-Hagery, M. A., Mohammed, A., & Hamza, M. A. (2023). Automated cardiovascular disease diagnosis using Honey Badger Optimization with modified deep learning model. *IEEE Access*, *11*, 64272-64281.
6. Ullah, T., Ullah, S. I., Ullah, K., Ishaq, M., Khan, A., Ghadi, Y. Y., & Algarni, A. (2024). Machine learning-based cardiovascular disease detection using optimal feature selection. *IEEE Access*, *12*, 16431-16446.
7. El-Sofany, Hosam F. „Predicting heart diseases using machine learning and different data classification techniques.” *IEEE Access* (2024).
8. Qadri, A. M., Raza, A., Munir, K., & Almutairi, M. S. (2023). Effective feature engineering technique for heart disease prediction with machine learning. *IEEE Access*, *11*, 56214-56224.
9. Gupta, A., Kumar, R., Arora, H. S., & Raman, B. (2019). MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE access*, *8*, 14659-14674.
10. Haque, M., Miah, A. S. M., Gupta, D., Prince, M. M. A. H., Alam, T., Sharmin, N., ... & Shin, J. (2024). Multi-class heart disease Detection, Classification, and Prediction using Machine Learning Models. *arXiv preprint arXiv:2412.04792*.
11. Kumar, A., Singh, K. U., & Kumar, M. (2023). A clinical data analysis based diagnostic systems for heart disease prediction using ensemble method. *Big Data Mining and Analytics*, *6*(4), 513-525.
12. Almazroi, A. A., Aldahhri, E. A., Bashir, S., & Ashfaq, S. (2023). A clinical decision support system for heart disease prediction using deep learning. *IEEE Access*, *11*, 61646-61659.
13. Al Reshan, M. S., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., & Shaikh, A. (2023). A robust heart disease prediction system using hybrid deep neural networks. *IEEE Access*, *11*, 121574-121591.
14. S. Mohan, C. Thirumalai, and G. Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” *IEEE Access*, vol. 7, pp. 81542–81554, Jan. 2019, doi: 10.1109/access.2019.2923707.
15. Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2020). HDPM: an effective heart disease prediction model for a clinical decision support system. *Ieee Access*, *8*, 133034-133050.
16. Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE access*, *8*, 107562-107582.
17. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
18. A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Ghenni, “Improving the Heart Disease Detection and Patients’ Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest,” *IEEE Access*, vol. 10, pp. 67363–67372, Jan. 2022, doi: 10.1109/access.2022.3185129.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.