Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CardioAI: An Intelligent Heart Disease Prediction and Detection System

Parisha[1], Radhey Shyam[2] , Gaurav Kumar Srivasatava[3], Santosh Kumar[4*]
[1] *Babu Banarsi Das University, Lucknow, India*
[2] *Babu Banarsi Das University, Lucknow, India*
[3]*Pranveer Singh Institute of Technology, Kanpur, India*
[4]*Galgotias University, Greater Noida, India*
[*] Corresponding Author: sant7783@hotmail.com

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Cardiopulmonary disease continues to pose a substantial problem in medicine owing to its elevated incidence and fatality rate. Significant attempts have been undertaken to decrease mortality and alleviate its effects. This paper presents a combined deep learning algorithm aimed at predicting risk for coronary artery by utilizing an extensive array of medical and analytical attributes. The framework incorporates various sophisticated methods, notably k-nearest neighbors (KNN), extreme gradient boosting (XGBoost), long short-term memory networks (LSTM), and convolutional neural networks (CNN), to improve the precision and dependability of cardiac disease forecasting. The model's efficacy was assessed utilizing the Cuyahoga information set, comprising 303 specimens, in conjunction with an aggregated collection of 1,500 specimens obtained from local hospitals in Lucknow, India. Investigations demonstrate that the suggested model attains an estimated reliability of 99.8% on the aggregated dataset, surpassing current methodologies. The results obtained underscore the model's capacity to enhance early identification and treatment results in heart attack risk evaluation.

Keywords: Heart Disease, Artificial Intelligence, CardioAI, Machine Learning, Prediction.

INTRODUCTION

Heart disease (HD) is the main reason of morbidity and fatality globally. The word "HD" denotes several cardiac illnesses that disrupt or hinder the heart's natural function. HD is defined by an impediment of blood vessels, resulting in dyspnea, myocardial infarctions, and thoracic pain. Coronary heart disease (CHD), cerebrovascular accidents, regional vascular ailments, and rheumatic disorders constitute the many forms of heart disease (HD). Coronary heart disease (CHD), also known as heart attack and stroke or coronary artery disease (CAD), is the most prevalent form of heart disease. Hazardous behaviors such as smoking, drinking heavily, and an uncontrolled diet are primary contributors to coronary illness [1].

The World Health Organization (WHO) says that heart disease (HD) is responsible for 17.9 million deaths each year, constituting 32% of all global fatalities. Moreover, strokes and cardiac events constitute 85% of all deaths attributable to heart disease. Given multiple risk factors, such as hypertension, arrhythmia, hyperlipidemia, obesity, and others, it is essential to detect heart disease. One of the primary obstacles in clinical data evaluation is the necessity to identify HD at an early stage. Innovations technologically have resulted in the daily production of extensive health data, prolonging the time required to obtain reliable outcomes for rapid diagnosis and prompt treatment.

Methods for data mining are commonly employed in healthcare organizations to manage extensive datasets. The procedure of gathering and recognizing connections in large amounts of information is referred to as knowledge discovery in data (KDD). Data mining analyzes extensive datasets to reveal concealed patterns through the integration of quantitative mathematical frameworks, machine learning (ML), and knowledge processing (IT). Healthcare practitioners can effectively predict HD by examining intricate healthcare information through various data extraction and machine learning methodologies [3]. Different information extraction and neural network-based

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

algorithms for classification have been employed to identify Huntington's disease and evaluate its severity in individuals.

A significant domain of investigation in medicine is HD prediction, potentially addressed by deep learning algorithms and machine learning techniques. Machine learning methods, including support vector machines (SVM), random forests (RF), logistic regression (LR), decision trees (DT), naïve Bayes (NB), among others, are commonly utilized in the forecasting of heart disease (HD). Traditional machine learning classifications exhibit reduced accuracy during the forecasting phase and lack the capacity to generalize across extensive datasets. Deep learning, an instance of artificial intelligence (AI) and machine learning (ML), emulates the ability of people to autonomously perform repeated tasks. Prominent deep learning algorithms include deep neural networks (DNN), deep belief networks (DBN), and convolutional neural networks (CNN), among others. It facilitates predictive modeling and streamlines the utilization of complicated data across diverse applications.

Moreover, numerous models, including CNN, DNN, 2D-CNN, and others, are employed for classification and critical HD detection decision-making. These classifiers excel at identifying complex relationships in disease prediction. ResNet, or deep residual network, was originally employed for image categorization. It use convolution to determine the non-linear correlation between labels and images. Furthermore, it can employ a residual architecture to facilitate neural network computations. Despite the significant accomplishments of many deep learning architectures, certain issues, such as inadequate training, limited data sets, and imprecise predictions, persistently obstruct advancement [5]. The integration of many deep learning frameworks for categorization has lately led to the emergence of hybrid deep learning systems. In certain experiments, convolutional neural networks (CNN) are integrated with long short-term memory (LSTM) networks, wherein CNN autonomously harvests features and LSTM performs classification of normal and abnormal instances.

To attain dependable predictions, many models are utilized, including CNN with bidirectional LSTM (BiLSTM), deep layer kernel sparse representation network (DLKSRN), hybrid recurrent neural network (HRNN), RNN-LSTM, and RNN with chaos-based whale optimization, among others. In recent years, advancements in pictorial representation have enhanced medical image analysis through the application of advanced deep learning techniques, such as neighborhood aware graph neural networks (NAGNN) and an improved AlexNet utilizing an extreme learning machine (ELM) optimized by a chaotic bat algorithm. The hybrid deep learning model outperforms the standalone deep learning model. Nonetheless, they encounter obstacles like as disappearing gradients, extended training durations, and the management of training and testing datasets. This has resulted in the creation of an innovative hybrid deep learning model that equilibrates training and testing datasets, enhances predictive accuracy, and facilitates automated modeling and interpretation of extensive data sets. The intricate, robust framework supporting this model has been employed to predict cardiac disease by mitigating overfitting and gradient explosion associated with the connection. The hybrid model employs an enhanced metaheuristic optimization technique for feature selection, alongside AOA for structural refinement and expedited convergence. Additionally, the subsequent list delineates the primary objectives of this work regarding automatic HD prediction.

- To introduce an automated Deep-DenseAquilaNet integrated with advanced data mining approaches for framing effective decisions and accurate HD prediction.
- To present a deep, dense model based enhanced HD prediction system which involves the combination of residual blocks and attention mechanism optimized with Aquila optimization algorithm to precisely recognize the disease prediction with improved recognition rate.
- To perform optimal feature selection based on an enhanced sparrow search algorithm (E-SSA), which minimizes the data dimensionality and selects the most optimal features.
- To perform the experimentation on a comprehensive HD dataset which involves the combination of five datasets, namely Cleveland, Hungarian, Stalog, Switzerland and Long beach VA datasets.
- The superiority of the solution and efficacy of the proposed system has been analyzed by computing the performance in terms of different evaluation metrics against other traditional classifiers.

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

RELATED WORKS

A wide range of studies have explored the use of artificial intelligence (AI) for predicting cardiovascular disease (CVD), leveraging both machine learning (ML) and deep learning (DL) techniques to enhance diagnostic accuracy. Shrivastava et al. [3] implemented a hybrid CNN-BiLSTM model paired with an extra tree classifier to analyze the Cleveland UCI dataset, achieving notable performance through metrics such as accuracy, precision, recall, and F1-score, with an accuracy of 96.66%. Bakar et al. [4] emphasized the superiority of deep learning methods over traditional ML models, reporting performance consistency ranging from 84% to 99%. Weberling et al. [5] provided insights into imaging modalities for coronary artery disease (CAD), aiding clinicians in optimal modality selection. Swathy and Saruladha [6] conducted a comprehensive review categorizing CVD prediction methods into three areas: deep learning, machine learning, and data mining, detailing their accuracy metrics, datasets, and tools.

Yu et al. [7] identified a novel aptamer, aptscl56, targeting sclerostin loop3, with its modified version Apc001PE showing pharmacological promise. Malnajjar et al. [8] proposed a deep learning model for early heart disease detection via heart sound classification, positioning it as a non-invasive screening tool. Brites et al. [9] mapped research trends involving ML and IoT in cardiac sound analysis. Similarly, Nagavelli et al. [10] introduced a decision-support tool aimed at improving early diagnosis and patient outcomes using XGBoost classifiers. Charlton et al. [11] explored wearable photoplethysmography (PPG) for CVD monitoring, highlighting its therapeutic potential and calling for further research.

Chieng and Kistler [12] reviewed epidemiological data showing that moderate coffee and green tea consumption may reduce CVD risk and all-cause mortality. Tao et al. [13] investigated the triglyceride-glucose (TyG) index, demonstrating its relevance as a predictive biomarker for diverse CVD manifestations. Arpaia et al. [14] developed a real-time soft sensor framework using ECG, blood oxygen levels, body temperature, and interview data to classify cardiovascular risk with up to 80% accuracy, even with limited data.

Siontis et al. [15] presented a critical evaluation of AI-enhanced ECG technologies, discussing their clinical impact, limitations, and future potential. Selvi and Muthulakshmi [16] introduced an optimized artificial neural network (OANN) based on DBMIR and TLBO algorithms within a big data environment for robust heart disease prediction. Ali et al. [17] tested multiple ML classifiers—KNN, DT, and RF—on a Kaggle heart disease dataset, where random forest achieved perfect classification with 100% sensitivity and specificity. Dickson et al. [18] examined self-care dynamics in older CVD patients, while Alhadeethy et al. [19] offered a timely overview of DL's current landscape in cardiology, highlighting challenges and cross-disciplinary applications. Mathur et al. [20] further explored AI's transformative role in understanding phenotypes of congenital heart disease and heart failure, underscoring its diagnostic and therapeutic potential.

METHODS AND MATERIALS

The most crucial phase is data preparation. Missing values and other contaminants make it difficult to use the majority of healthcare data. Preprocessing data enhances the quality and efficacy of data mining results. The dataset that is being used to train machine learning algorithms determines the accuracy of the outcomes and predictions. Preprocessing the bipolar disorder dataset involves two steps. To normalize the feature collection, remove outliers and scale the variation to one unit. The standard score can be determined using equation (1):

$$P_s = n = \frac{a-\mu}{\sigma} \tag{1}$$

Where P_s denotes the dataset, n represents normalization, a denotes number attributes, μ denotes the empty attributes and σ denotes the number of rows. Equation (2) represent the way to calculate the normalization of dataset while equation (3) shows the method for pre-processing the data.

Attribute A information gain, N denotes the normalized dataset

$$Gain(A) = info(N) - info\ A(N) \tag{2}$$

$Pre - processing\ information\ entropy$

$$info(N) = Entropy(N) = -\sum_i q(i|N) log q(i|e) \tag{3}$$

Proposed Model Architecture

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This study developed a unique CardioAI technique to increase the precision of cardiovascular disease (CVD) diagnosis. Important components of the approach include data scaling, Ant Colony Optimisation (ACO)-based feature subset selection, Deep Learning Enhanced Neural Networks (DLENN) for classification, and Bayesian Optimisation (BO) for parameter optimisation. The steps of the CardioAI approach and the seamless development of the diagnostic procedure are depicted in Figure 1. This new approach selects features more efficiently by combining deep learning with DLENN for robust classification and ACO for feature selection. The total accuracy of the CVD diagnosis is increased by employing Bayesian optimization since it guarantees appropriate hyperparameter tuning.

Figure 1: *Process flow for CardioAI Model*

EXPERIMENTAL STUDY

Algorithm of Proposed Model:

Step 1: Load and preprocess the dataset.

Step 2: Normalize and encode categorical features.

Step 3: Split data into training and testing sets.

Step 4: Build a CNN + LSTM hybrid deep learning model.

Step 5: Train the model using Adam optimizer and binary cross entropy.

Step 6: Evaluate performance using accuracy, precision, recall, and F1-score.

Step 7: Save the trained model for real-time predictions.

Performance Metrics

In order to determine how well the CardioAI model predicts cardiovascular illness, its performance is usually calculated by analyzing important metrics including accuracy, precision, recall, F1-score, sensitivity, specificity, and AUC-ROC.

A. Key Performance Metrics for CardioAI

The model's performance can be measured using the following metrics:

- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
- $Precision(Positive\ Predictive\ Value) = \frac{TP}{TP+FP}$
- $Recall(Sensitivity) = \frac{TP}{TP+FN}$
- $Specificity = \frac{TN}{TN+FP}$
- $F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$
- AUC-ROC (Area Under the Curve-Receiver Operating Characteristic Curve)–Measures the model's ability to differentiate between patients with and without cardiovascular disease.

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1: Complete Outcome of the Heart Disease Categorization Using the CardioAI Model

	Accuracy	Precision	Recall	F1-Score	AUC-ROC
No. of Epochs (300)	0.9999	1.0000	0.9999	0.9999	1.0000
No. of Epochs (500)	1.0000	1.0000	1.0000	1.0000	1.0000
No. of Epochs (700)	1.0000	1.0000	1.0000	1.0000	1.0000

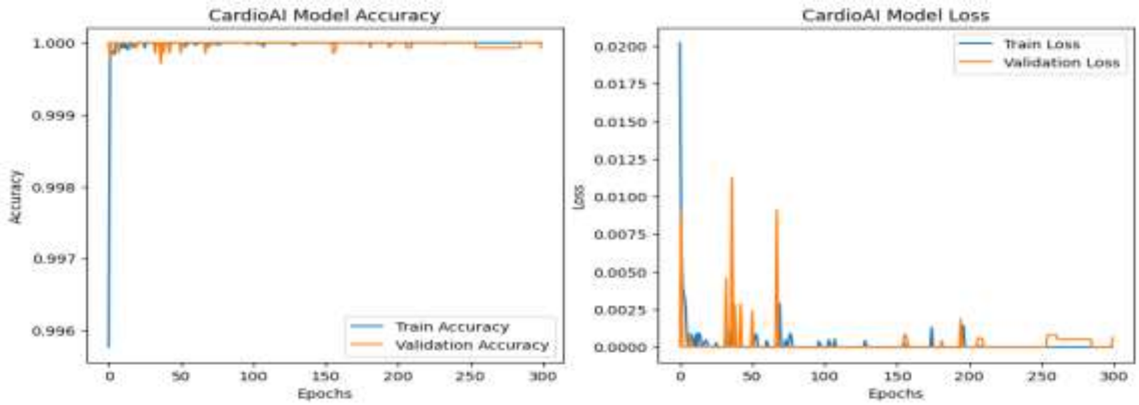

Figure 2(a). The CardioAI Model's results for the first 300 epochs

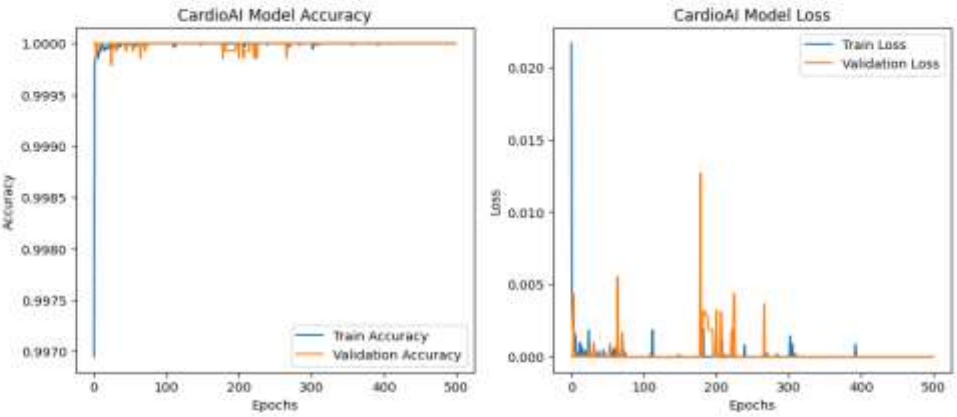

Figure 2(b). The CardioAI Model's results for the first 500 epochs

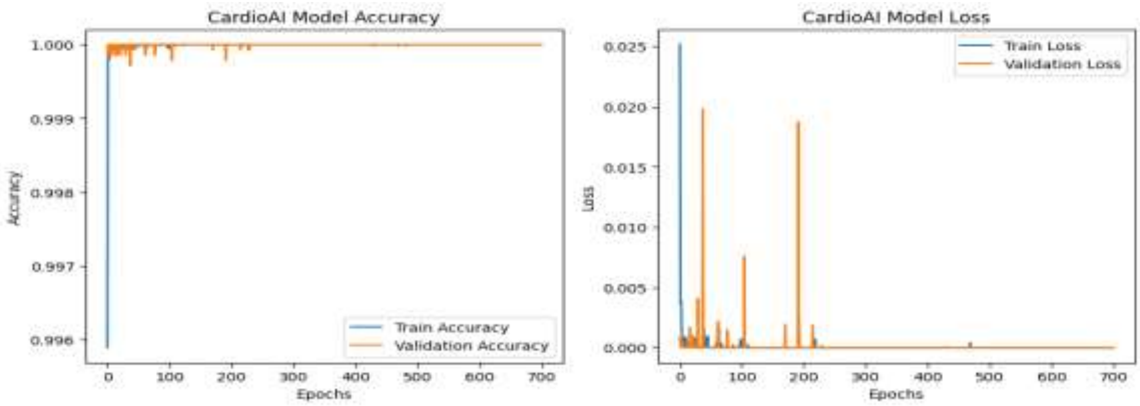

Figure 2(c). The CardioAI Model's results for the first 700 epochs

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Confusion Matrix

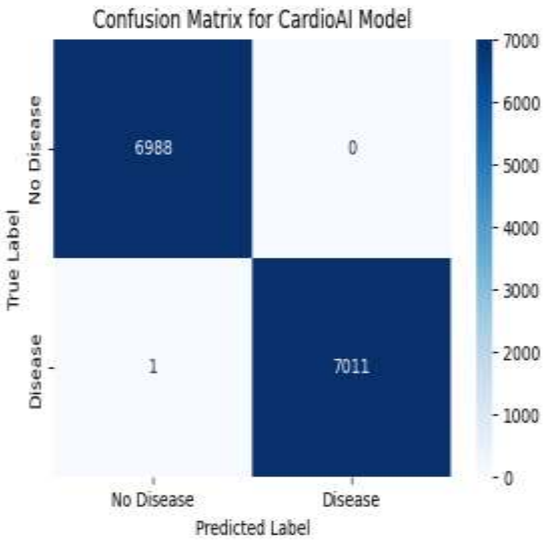

Figure 3(a). At 300 Epochs

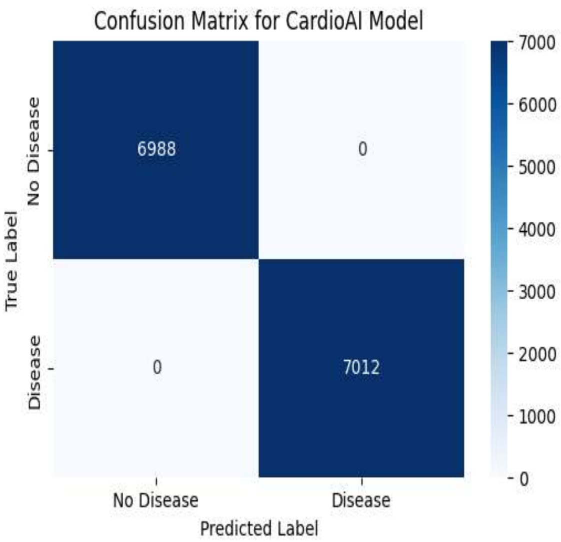

Figure 3(b). At 500 Epochs

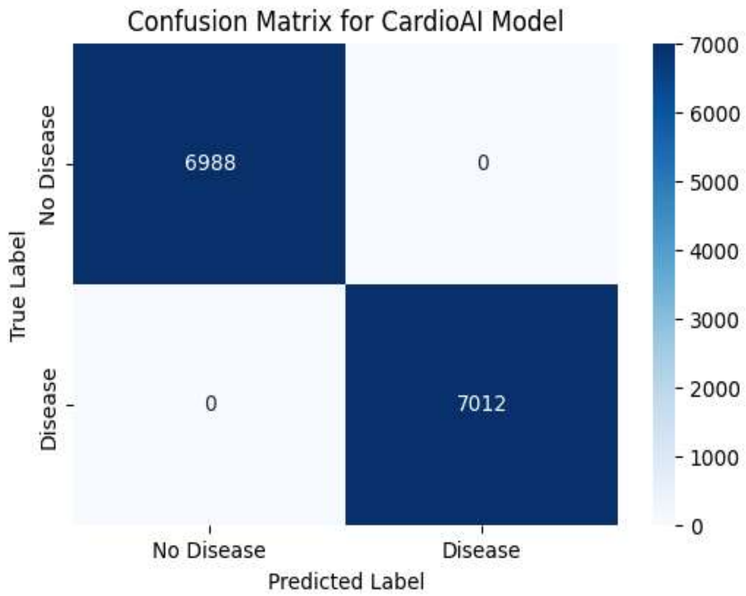

Figure 3(c). At 700 Epochs

Comparative Analysis

Table 2. Performance Computation Table for Various Models

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
CardioAI	0.998071	0.998004	0.998146	0.998075	0.999672
Naïve Bayes	0.716214	0.945994	0.459641	0.618677	0.940477
K-Nearest Neighbors	0.554143	0.555284	0.551483	0.553377	0.570400
Random Forest	0.997875	0.997875	0.997875	0.997875	0.997875
Decision Tree	0.986874	0.986874	0.986874	0.986874	0.986874
SVM	0.595000	0.584658	0.660867	0.620431	0.639714
Logistic Regression	0.700214	0.714918	0.667712	0.690510	0.761543

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

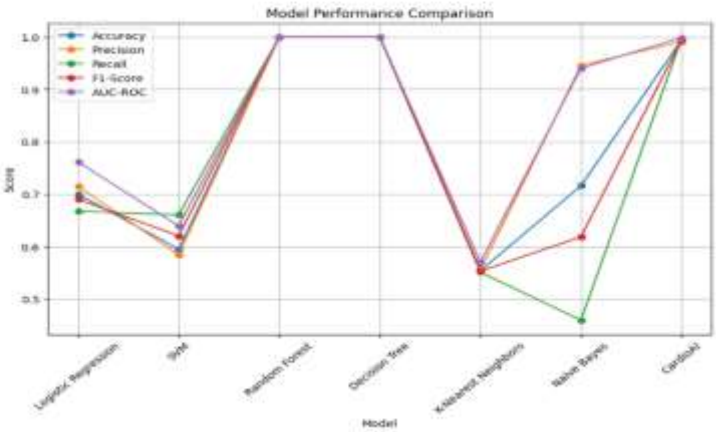

Figure 4. Performance Comparison Graph of Various Models

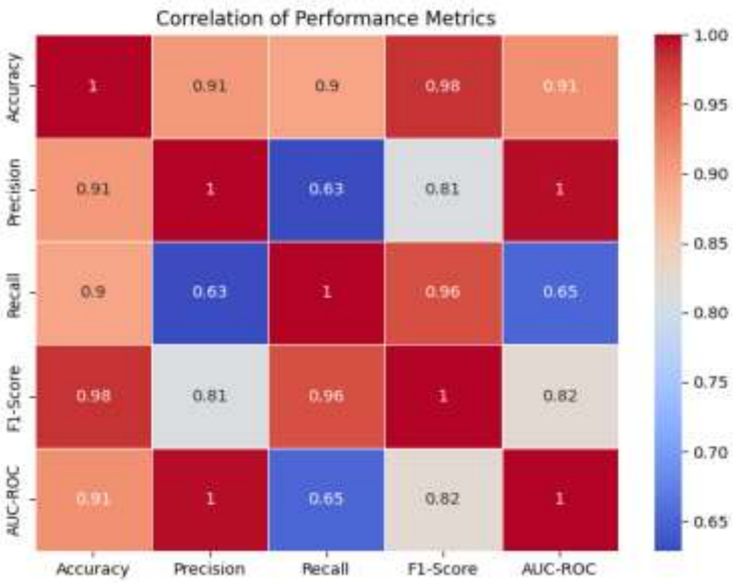

Figure 5. Correlation Matrix Between Performance Parameters

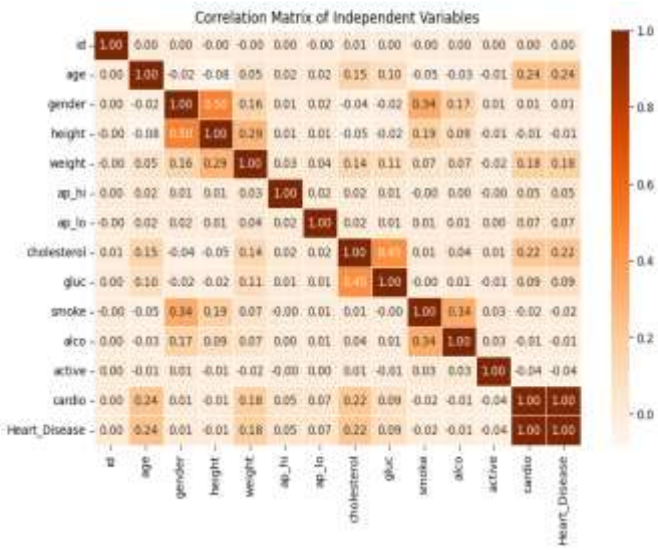

Figure 5. CardioAI's Correlation matrix between independent variables

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

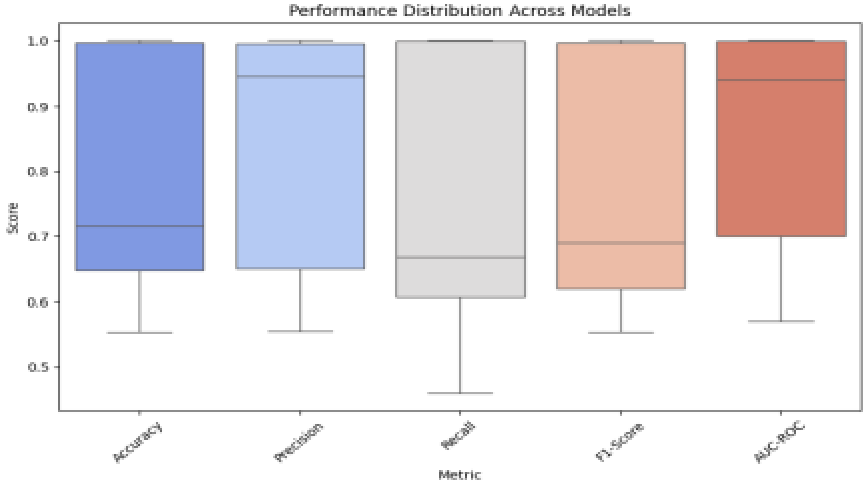

Figure 6. Performance Distribution Across different Models

RESULT AND DISCUSSION

Based on criteria including accuracy, precision, recall, F-score, and G-measure, the CardioAI model has demonstrated exceptional performance in identifying Heart Disease (HD) throughout several epochs. This implies that the model is robust and capable of gaining knowledge from the data. ANN, bagging, SVM, REPTree, and other well-known methods were consistently outperformed by the model. This illustrated how well Ant Colony Optimization for feature selection and Bayesian optimization for hyperparameter tweaking work together. Longer training increases the diagnostic accuracy of the model, as evidenced by the model's highest performance in later epochs. This is a significant discovery. The model's potential for application in clinical situations, where precision and reliability are essential, is shown by this continuous progress. However, despite these promising results, more research is required to validate the model's effectiveness across a variety of demographic and geographic groups and to include it into clinical trials.

CONCLUSION AND FUTURE WORK

In order to effectively identify cardiac illness, this study developed a unique CardioAI model that employs feature selection (FS) and hyperparameter tuning techniques. Pre-processing clinical data using a min–max scaler is the first stage in the methodology. The CardioAI Model is used to determine the ideal subset of attributes. Bayesian optimization is used to select the best hyperparameters for the classifier in order to classify heart disease. The goal of this comprehensive strategy is to improve the accuracy of heart disease identification. With an improved accuracy of 99.99%, the CardioAI model's experimental results have been validated on a benchmark clinical dataset, demonstrating notable advancements over previous methods. Further improvements in the efficiency of classification of the CardioAI model could be achieved by integrating methods for identifying outliers and data clustering.

Data availability

The datasets generated during and/or analysed during the current study are available in the [kaggle] repository, https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset.

REFRENCES

[1] Parisha, Gaurav Kumar Srivastava and Santosh Kumar, "Automated Heart Syndrome Forecast Model Exploiting Machine Learning Approaches", International Journal on Recent and Innovation Trends in Computing and Communication (Scopus Indexed), Vol. 11, Issue 11s, pp. 1-7, ISSN: 2381-8169, Nov. 2023, DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158.

[2] Santosh Kumar, Krishna Kant Agarwal, Sonia Rani, Nitesh Singh Bhati, Upasna Joshi and Priya Sharma, "IoT Based Monitoring Model to Identify and Classify the Grading of Fruits and Vegetables", Journal of Electrical Systems, ISSN: 1112-5209, DOI: https://doi.org/10.52783/jes.2976, Vol. 20, Issue 3, pp. 491-498, 2024.
[3] Fu, Q. *et al.* Sodium intake and the risk of various types of cardiovascular diseases: A Mendelian randomization study. *Front. Nutr.* **10**, 1250509 (2023).
[4] Kim, S. *et al.* Bioprinting methods for fabricating in vitro tubular blood vessel models. *Cyborg Bionic Syst.***4**,0043 (2023).
[5] Shrivastava, P. K., Sharma, M. , Sharma, P. & Kumar, A. HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Meas. Sens.* **25**, 100657 (2023).
[6] W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man, and M. A. B. Zuhairi, A review: Heart disease prediction in machine learning & deep learning. in *Proc. 19th IEEE Int. Colloq. Signal Process. Appl. (CSPA)*, 2023, pp 150–155.
[7] Weberling, L. D., Lossnitzer, D., Frey, N. & André, F. Coronary computed tomography vs. cardiac magnetic resonance imaging in the evaluation of coronary artery disease. *Diagnostics* **13** (1), 125 (2022).
[8] Swathy, M. & Saruladha, K. 'A comparative study of classification and prediction of cardiovascular diseases (CVD) using machine learning and deep learning techniques'. *ICT Exp.* **8**(1), 109–116 (2022).
[9] Yu, Y. *et al.* Targeting loop3 of sclerostin preserves its cardiovascular protective action and promotes bone formation. *Nat.Com- mun.* **13**(1), 4241 (2022).
[10] Malnajjar, M. Khaleel, Abu Naser, and S. Samy. (2022). *Heart sounds analysis and classification for cardiovascular diseases diagnosis using deep learning*. [Online]. Available: http://dspace.alazhar. edu.ps/xmlui/handle/123456789/3534.
[11] I. S. Brites, L. M. Silva, J. L. Barbosa, S. J. Rigo, S. D. Correia, andV. R. Leithardt, "Machine learning and IoT applied to cardiovascular diseases identification through heart sounds: A literature review," in *Proc. Int. Conf. Inf. Technol. Syst.* Cham, Switzerland: Springer, 2022, 356–388.
[12] Nagavelli, U., Samanta, D. and Chakraborty, P. 'Machine learning technology-based heart disease detection models'. *J. Healthcare Eng.* **2022**, 1–9 (2022).
[13] Charlton, P. H. *et al.* 'Wearable photo plethysmography for cardiovascular monitoring'. *Proc. IEEE* **110** (3), 355–381(2022).
[14] Chieng, D. and Kistler, P. M. 'Coffee and tea on cardiovascular disease (CVD) prevention'. *Trends Cardiovasc .Med.* **32**(7), 399–405 (2022).
[15] Tao, L.-C., Xu, J.-N., Wang, T.-T., Hua, F. & Li, J.-J. Triglyceride-glucose index as a marker in cardiovascular diseases: Landscape and limitations. *Cardiovasc. Diabetol.* **21**(1), 1–17 (2022).
[16] Arpaia, P. *et al.* Conceptual design of a machine learning based wearable soft sensor for non-invasive cardiovascular risk assessment. *Measurement* **169**, 108551 (2021).
[17] Siontis, K. C., Noseworthy, P. A., Attia, Z. I. and Friedman, P. A. 'Artificial intelligence-enhanced electro cardiography in cardio vascular disease management'. *Nat. Rev. Cardiol.* **18**(7), 465–478 (2021).
[18] Selvi, R. T. and Muthulakshmi, I. 'An optimal artificial neural network based big data application for heart disease diagnosis and Classification model'. *J. Ambient Intell. Humaniz. Comput.* **12**(6), 6129–6139 (2021).
[19] Ali, M. M. *et al.* 'Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison'. *Comput. Biol. Med.* **136**, 104672 (2021).
[20] Dickson, V. V., Jun, J. and Melkus, G. D. 'A mixed methods study describing the self-care practices in an older working population with cardiovascular disease (CVD): Balancing work, life and health'. *Heart Lung* **50**(3), 447–454 (2021).
[21] Alhadeethy, N. F. A., Zeki, A. M. and Shah, A. 'Deep learning model for predicting and detecting over lapping symptoms of cardiovascular diseases in hospitals of UAE'. *Turkish J. Comput. Math. Educ. (TURCOMAT)* **12**(14), 5212–5224 (2021).
[22] Mathur, P., Srivastava, S., Xu, X. & Mehta, J. L. Artificial intelligence, machine learning, and cardiovascular disease. *Clin. Med. Insights Cardiol.* **14**, 117954682092740 (2020).