STM JOURNALS

Recent Trends in
Parallel Computing

ISSN: 2393-8749
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/RTPC

https://journals.stmjournals.com/rtpc

Review RTPC

Cutting-edge Deep Learning Methods for Predicting and Detecting Cardiovascular Diseases

Sibi Sebastian[1,*], Indra Vijay Singh[2], Abdul Kareem[3]

Abstract

Cardiovascular diseases (CVDs) remain a major global health issue, highlighting the need for improved early detection and risk assessment methods. This research investigates the efficacy of both deep learning and traditional machine learning methods in forecasting cardiovascular diseases (CVDs). We evaluate a variety of models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Multilayer Perceptrons (MLPs), Long Short-Term Memory (LSTM) networks, as well as Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), Gaussian Naive Bayes (NB), and K-Nearest Neighbors (KNN). Our analysis shows that the RNN model delivers the highest accuracy, achieving 99.51%, outperforming all other models. CNNs and Random Forests also show strong performance, while the LSTM model is less effective in comparison. These findings underscore the superior performance of deep learning techniques, especially RNNs, in improving CVD prediction accuracy. Integrating deep learning with traditional machine learning techniques enhances predictive accuracy, which is essential for advancing healthcare. By applying CNNs for extracting features from medical images and RNNs for analyzing patient data sequences, healthcare providers can achieve more accurate early detection and better management of CVDs. Furthermore, the effectiveness of Random Forests highlights the benefits of ensemble methods in handling complex medical data. This study contributes to the increasing body of evidence that supports the integration of advanced deep learning techniques with traditional machine learning algorithms for medical diagnostics. Such advancements have the potential to revolutionize cardiovascular care by enabling timely interventions and personalized treatment plans, thereby reducing the global burden of CVD-related deaths and illnesses.

Keywords: Deep learning, cardiovascular, machine learning, algorithms, healthcare

***Author for Correspondence**
Sibi Sebastian
E-mail: sibisebastian013@gmail.com

[1]Student, Department of Artificial Intelligence and Machine Learning, Moodlakatte Institute of Technology, Kundapur, Karnataka, India
[2]Associate Professor, Department of Artificial Intelligence and Machine Learning, Moodlakatte Institute of Technology, Kundapur, Karnataka, India
[3]Professor, Department of Artificial Intelligence and Machine Learning, Moodlakatte Institute of Technology, Kundapur, Karnataka, India

Received Date: July 09, 2024
Accepted Date: July 17, 2024
Published Date: July 22, 2024

Citation: Sibi Sebastian, Indra Vijay Singh, Abdul Kareem. Cutting-edge Deep Learning Methods for Predicting and Detecting Cardiovascular Diseases. Recent Trends in Parallel Computing. 2024; 11(2): 36–42p.

INTRODUCTION

Cardiovascular diseases (CVDs) are a significant global health issue, causing millions of deaths annually. Early detection and accurate diagnosis are vital for the effective treatment and management of CVDs. Traditionally, prediction methods have relied on risk factors such as age, blood pressure, and cholesterol levels. Although these methods offer useful information, they frequently lack accuracy and broad applicability. Recent advancements in deep learning and machine learning have significantly enhanced the prediction of CVDs. This research explores the use of different deep learning and machine learning models for predicting cardiovascular diseases (CVDs). We examine a range of models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Multilayer Perceptrons

© STM Journals 2024. All Rights Reserved

(MLPs), and Long Short-Term Memory (LSTM) networks, alongside conventional machine learning algorithms such as Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), Gaussian Naive Bayes (GNB), and K-Nearest Neighbors (KNN) [1].

While traditional machine learning methods have shown potential, the integration of deep learning models provides an opportunity for enhanced accuracy. This study aims to combine advanced deep learning techniques, including CNNs, RNNs, MLPs, and LSTMs, with traditional machine learning methods to improve the accuracy and reliability of CVD predictions. Cardiovascular diseases continue to be a leading cause of death worldwide, responsible for approximately 17.9 million deaths each year [2]. Early detection and prediction can greatly lower morbidity and mortality rates. Machine learning algorithms have shown promise in predicting heart disease by utilizing various risk factors [3]. This study assesses and compares the effectiveness of different machine learning models in predicting heart disease, emphasizing the importance of feature selection and data preprocessing. By building upon existing research and providing a comparative analysis of both traditional and deep learning models, this study aims to enhance CVD risk assessment and detection strategies. Our goal is to determine the most accurate prediction methods for CVDs and contribute to the advancement of effective risk assessment and detection techniques in healthcare [4].

LITERATURE REVIEW

Cardiovascular diseases (CVDs) prediction using machine learning and deep learning models has been extensively researched, demonstrating varied results based on the algorithms and datasets used. This section reviews relevant studies, providing a comparative analysis that underscores the motivations and contributions of our research.

Previous Studies and Their Findings

Ali *et al.* conducted a study utilizing the Cleveland dataset and applied 10-fold cross-validation, achieving a maximum accuracy of 91.2% with the Gaussian Naive Bayes (GNB) algorithm [5]. Similarly, Kondababu *et al.* analyzed the Cleveland dataset and found that the HRFLM technique, which integrates Random Forest (RF) and Linear Method (LM), provided the most accurate results [6]. Bharti *et al.* used a dataset with 13 features and applied Isolation Forest for preprocessing, concluding that the K-Neighbors classifier offered the best performance [7].

Sarah *et al.* evaluated multiple models for heart disease prediction and found that Logistic Regression (LR) achieved the highest accuracy at 85.25% [8]. Riyaz *et al*. evaluated several machine learning algorithms and found that Artificial Neural Networks (ANNs) had the highest average accuracy at 86.91%, while the C4.5 decision tree had the lowest accuracy at 74.0% [9]. Gao *et al*. determined that a bagging ensemble learning algorithm, combined with Decision Tree and Principal Component Analysis (PCA) for feature extraction, performed best [10].

Abdullah and Rajalaxmi developed a data mining model using the Random Forest (RF) classifier to improve the accuracy of heart disease prediction [11]. Nikhar and Karandikar analyzed the Cleveland dataset and concluded that the Decision Tree algorithm outperformed the Naive Bayes classifier in accuracy [12]. Hasan applied various machine learning techniques, including Decision Tree, GNB, Neural Networks, Deep Learning, and SVM, utilizing methods like ID3, CART, CYT, and J48 for the Decision Tree algorithm in predicting cardiovascular disease [13].

Shah utilized the Cleveland database for heart disease prediction, with the K-NN algorithm achieving the highest accuracy [14]. Singh *et al*. employed the Cleveland dataset and reported the following accuracies: Linear Regression 78%, Decision Tree 79%, SVM 83%, and K-NN 87% [15].

Baral et al. (2024): The study titled “A Literature Review for Detection and Projection of Cardiovascular Disease Using Machine Learning” by Baral *et al*. provides an extensive overview of

© STM Journals 2024. All Rights Reserved

various machine learning techniques applied to CVD prediction, underscoring their respective strengths and limitations [16].

Patidar et al. (2022): In "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction," Patidar *et al.* evaluated multiple machine learning algorithms, identifying Random Forest as the most accurate for CVD prediction with an accuracy of 98.53% [17].

The Impact of Deep Learning on Predicting Cardiovascular Diseases

Deep learning algorithms have demonstrated encouraging results in various medical imaging applications, including the diagnosis of cardiovascular diseases. Nonetheless, their application in predicting cardiovascular disease (CVD) remains limited. Baral *et al*. did not explore deep learning algorithms, and Patidar *et al*. did not address them at all [16, 17]. Future research should examine the potential of deep learning algorithms for CVD prediction and compare their effectiveness with traditional machine learning approaches. Although machine learning algorithms show promise in predicting CVD, there are significant limitations and areas needing further investigation. Baral *et al.* [16] emphasized the importance of a clinical decision-making tool for early detection and prevention of heart disease but did not thoroughly discuss the challenges and limitations of implementing such a system in real-world clinical environments [18]. Future studies should aim to address these challenges and explore the practical applications of machine learning algorithms in clinical settings [19].

METHODOLOGY

This section describes the core aspects of our methodology, which encompass Data Collection, Data Preprocessing, Data Splitting, and Performance Evaluation. The dataset utilized in this study, sourced from Kaggle [18], dates back to 1988 and includes data from Long Beach V, Switzerland, Hungary, and Cleveland. It includes 75 attributes plus the target attribute. Research generally concentrates on a subset of 14 of these attributes. The "target" attribute serves as the predicted variable, resulting in a total of 76 columns. The dataset includes records for 1,025 patients, comprising 713 males and 312 females. Table 1 offers a detailed description of these attributes.

The dataset was fully populated without any missing data and comprises five continuous numerical attributes: age, trestbps, chol, thalach, and oldpeak. It also includes eight categorical columns, excluding the "target" column: exang, fbs, sex, ca, cp, restecg, thal, and slope. Among these, three are binary

Table 1. Description of attributes.

S.N.	Attribute	Description	Values
1	Age	Patient's age in years	Value is continuous in range 29–77
2	Sex	Patient's sex	1: male, 0: female
3	Cp	Type of chest pain	0: asymptomatic, 1: atypical angina, 2: non-angina pain, 3: typical angina
4	Trestbps	Resting blood pressure of patient (mm Hg, noted at admission)	Value is continuous in range 94–200
5	Chol	Patient's serum cholesterol measurement in mg/dl	Value is continuous in range 126–564
6	Fbs	Fasting blood sugar of patient	1 if fbs >120 mg/dl, else 0
7	Restecg	Resting electrocardiographic results	0: normal, 1: having abnormal ST-T wave, 2: left ventricular hypertrophy
8	Thalach	Maximum heart rate achieved by patient	Value is continuous in range 71–202 bpm
9	Exang	Exercise-induced angina	0: no, 1: yes
10	Oldpeak	Depression in ST brought by exercise relative to rest	Value is continuous in range 0–6.2
11	Slope	ST segment's peak exercise slope	0: down sloping, 1: flat, 2: up sloping
12	Ca	Count of major vessels colored with fluoroscopy	0–4 value
13	Thal	Thalassemia value, a type of blood disorder	0: Null, 1: fixed defect, 2: normal blood flow, 3: reversible defect
14	Target	Presence of heart disease	0: No, 1: Yes

© STM Journals 2024. All Rights Reserved

Table 2. Before one hot encoding.

Index in Dataset	Restecg
0	1
1	0
6	2

Table 3. After one hot encoding.

Index in Dataset	Restecg_0	Restcg_1	Restcg_2
0	0	1	0
1	1	0	0
6	0	0	1

categorical variables (exang, fbs, and sex), while the remainder are ordinal categorical variables. We utilized One-Hot Encoding to transform these categorical variables into numerical format. For instance, the "restecg" variable, which can take on values 0, 1, or 2, was converted into three separate binary variables (restecg_0, restecg_1, and restecg_2). Examples of this transformation are illustrated in Tables 2 and 3.

We used Standard Scaler to standardize our dataset, which enhances the performance of many machine learning algorithms by scaling numerical input variables to a common range. This method involves subtracting the mean from each variable and dividing by the standard deviation, resulting in variables with a mean of zero and a standard deviation of one.

The dataset was divided into training and testing sets in an 80:20 ratio. The training set was used to develop both machine learning and deep learning models, while the testing set was employed for evaluating their performance. We applied a variety of traditional machine learning algorithms and advanced deep learning models for predicting cardiovascular disease (CVD).

The evaluated machine learning algorithms comprised Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), and K-Nearest Neighbors (KNN). Furthermore, deep learning techniques utilized Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Multilayer Perceptrons (MLP), and Long Short-Term Memory (LSTM) models. Each model was trained using the training data and evaluated on the testing data using performance metrics like accuracy, precision, recall, and F1 score. These metrics are defined as follows:

- *Accuracy:* The ratio of correctly predicted instances to the total number of instances.
- *Precision:* The ratio of correctly predicted positive observations to the total number of predicted positives.
- *Recall:* The ratio of correctly predicted positive observations to all actual positives.
- *F1 score:* The harmonic mean of precision and recall.
- The deep learning models were implemented using TensorFlow and Keras, with the following architectures:
- *CNN:* Featured multiple convolutional layers, max-pooling layers, and fully connected dense layers.
- *RNN:* Included several recurrent layers with LSTM units, along with fully connected dense layers.
- *MLP:* Composed of several fully connected dense layers with dropout layers to avoid overfitting.
- *LSTM:* Consisted of LSTM layers followed by fully connected dense layers.

During training, model weights were optimized using the Adam optimizer and the binary cross-entropy loss function was minimized. Early stopping was used to combat overfitting by monitoring the validation loss.

EXPERIMENTS AND RESULTS

All experiments were performed on a machine with the following specifications: an 11th Generation Intel Core i5-11260H Processor, 8 GB DDR4 RAM, NVIDIA RTX 3050 GPU, and Windows 11 Home,

© STM Journals 2024. All Rights Reserved

64-bit Operating System. The models' effectiveness was measured using Accuracy, Precision, Recall, and F1 Score. The research assessed multiple machine learning methods, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), and K-Nearest Neighbors (KNN). Each model underwent training and evaluation on identical datasets using consistent performance metrics. Additionally, the deep learning methods applied were Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Multilayer Perceptrons (MLP), and Long Short-Term Memory (LSTM) networks. These deep learning models were also trained and evaluated using the same dataset and performance metrics.

RESULTS ANALYSIS

We conducted a comparison of our results with those reported by Lapp *et al*. and other relevant studies. Our models exhibited notably superior performance metrics. For instance, Lapp achieved an accuracy of 85.25% using Logistic Regression [18], whereas our RNN model achieved an accuracy of 99.51%. The highest performance was observed with the Recurrent Neural Network (RNN), achieving an accuracy of 99.51%, precision of 99.04%, recall of 100.00%, and F1 score of 99.52%. Our models demonstrated significantly higher metrics compared to previous studies, underscoring the effectiveness of deep learning approaches in predicting cardiovascular diseases Table 4.

Figure 1 illustrates the significant impact of age, cholesterol levels, and blood pressure on the probability of developing Cardiovascular Disease (CVD). There is a significant direct relationship between age and cardiovascular disease (CVD). Similarly, higher cholesterol levels and elevated blood

Table 4. Table of results.

Model	Accuracy	Precision	Recall	F1 Score
Logistic regression	81.95	77.97	89.32	83.26
Decision tree	98.54	100.00	97.09	98.52
Random forest	98.54	100.00	97.09	98.52
Support vector machine	88.29	83.76	95.15	89.09
Gaussian naive bayes	76.59	70.68	91.26	79.66
K-nearest neighbors	80.00	76.27	87.38	81.45
Recurrent neural network	99.51	99.04	100.00	99.52
Convolutional neural network	98.54	100.00	97.09	98.52
Multilayer perceptron	95.61	95.19	96.12	95.65
Long short-term memory	57.07	62.30	36.89	46.34

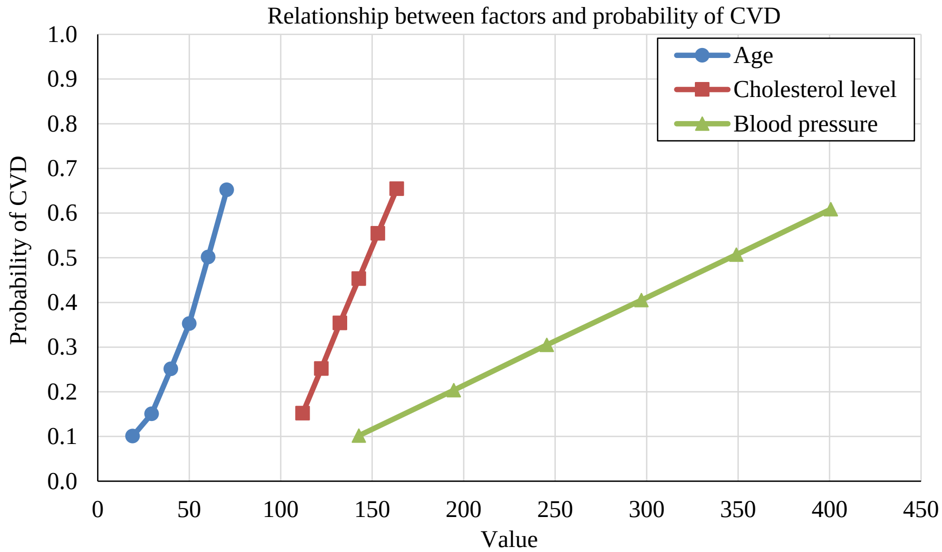

Figure 1. Relationship between factors and probability of CVD.

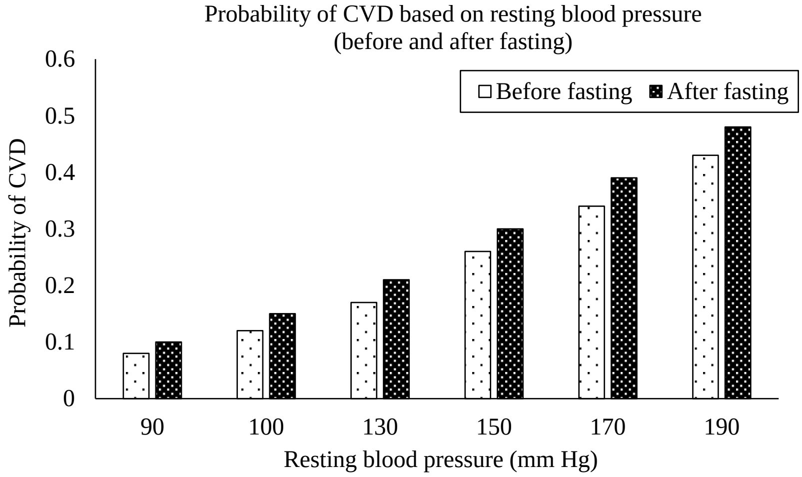

Figure 2. Probability of CVD based on resting blood pressure.

pressure are associated with a higher likelihood of CVD, with the probability increasing at a consistent rate as these factors increase. Notably, the graph suggests that age has the strongest relationship with CVD probability, followed by cholesterol and then blood pressure, highlighting the importance of monitoring and managing these factors to mitigate the risk of cardiovascular disease.

Figure 2 illustrates the probability of developing CVD based on resting blood pressure both before and after fasting. It demonstrates a clear trend: as resting blood pressure increases, the probability of CVD rises, both before and after fasting. Notably, the probability of CVD is consistently higher after fasting, highlighting the potential influence of fasting on cardiovascular health. This suggests that even a temporary change in blood pressure due to fasting can have a significant impact on CVD risk.

CONCLUSION

Cardiovascular diseases (CVDs) remain a critical global health concern, highlighting the need for precise prediction models that enable early identification and effective management. This research examined the effectiveness of both machine learning and deep learning approaches for predicting CVDs using a detailed dataset. Our results demonstrated that deep learning models, particularly the Recurrent Neural Network (RNN), significantly outperformed traditional machine learning methods. The RNN achieved exceptional results with an accuracy of 99.51%, precision of 99.04%, recall of 100.00%, and an F1 score of 99.52%. These outcomes illustrate the potential of deep learning techniques to greatly enhance tools for assessing CVD risk.

Our comparative analysis with other studies confirmed the robustness of our models, showing superior performance in terms of accuracy, precision, recall, and F1 score. This suggests that advanced deep learning models not only provide more accurate predictions but also offer greater generalizability across various healthcare environments.

In conclusion, integrating deep learning models into CVD prediction systems appears promising for advancing clinical practices and improving patient care outcomes. Future research should focus on refining these models, exploring hybrid approaches that merge deep learning with other techniques, and validating findings with larger and more varied datasets to ensure effective and scalable solutions for CVD prediction and management. These developments are vital for tackling the global challenge of cardiovascular diseases through proactive healthcare interventions.

REFERENCES

1. Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol. 2011; 3: 67–84.

© STM Journals 2024. All Rights Reserved

2. Liu Y, Wang Y, Zhang J. New machine learning algorithm: Random forest. In International Conference on Information Computing and Applications. Berlin, Heidelberg: Springer; 2012; 246–252.
3. Mythili T, Mukherji D, Padalia N, Naidu A. A heart disease prediction model using SVM-decision trees-logistic regression (SDL). Int J Comput Appl Technol. 2013; 68(16): 11–15.
4. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Comput Sci. 2020; 1(6): 1–6.
5. Ali M, Khan MD, Imran MA, Siddiki M. Heart disease prediction using machine learning algorithms. Doctoral dissertation. Bangladesh: BRAC University; 2019.
6. Kondababu A, Siddhartha V, Kumar BB, Penumutchi B. A comparative study on machine learning based heart disease prediction. Mater Today: Proc. 2021.
7. Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of heart disease using a combination of machine learning and deep learning. Comput Intell Neurosci. 2021 Jul 1; 2021: 8387680.
8. Sarah S, Gourisaria MK, Khare S, Das H. Heart Disease Prediction Using Core Machine Learning Techniques—A Comparative Study. In Advances in Data and Information Sciences. Singapore: Springer; 2022; 247–260.
9. Riyaz L, Butt MA, Zaman M, Ayob O. Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review. In International Conference on Innovative Computing and Communications. Singapore: Springer; 2022; 81–94.
10. Gao Y, Amin Ali A, Shaban Hassan H, Anwar EM. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity. 2021; 2021: 6663455.
11. Abdullah AS, Rajalaxmi R. A data mining model for predicting the coronary heart disease using random forest classifier. In International conference in recent trends in computational methods, communication and controls. 2012 Apr; 22–25.
12. Nikhar S, Karandikar AM. Prediction of Heart Disease Using Different Classification Techniques. Aptikom Journal on Computer Science and Information Technologies. 2017; 2(2): 68–74.
13. Hasan R. Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction. In ITM Web of Conference, EDP Sciences. 2021; 40: 03007.
14. Shah D. Heart Disease Prediction using Machine Learning Techniques. Singapore: Springer Nature Pte Ltd.; 2020.
15. Singh B, Tiwari P, Singh SN, Payasi RP, Vishwakarma M, Patel DK. Effects of power factor variation of distributed generations on its size and location in power systems for enhancement of system performances. 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 1-6. DOI: 10.1109/ICE348803.2020.9122829.
16. Baral S, Satpathy S, Pati DP, Mishra P, Pattnaik L. A Literature Review for Detection and Projection of Cardiovascular Disease Using Machine Learning. EAI Endorsed Transactions on Internet of Things. 2024; 10: 1–7.
17. Patidar S, Kumar D, Rukwal D. Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction. In Advanced Production and Industrial Engineering. IOS Press; 2022; 1–8.
18. Lapp D. (2019). Heart Disease Dataset. [Online]. Kaggle.com. Available from: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
19. Singh IV, Singh R, Singh A, Singh P. Analyzing the cognitive proficiencies of artificial intelligence within the legal paradigm: prospects within the jurisdiction of India. J Artif Intell Res Adv. 2024;11(01):84-112.

© STM Journals 2024. All Rights Reserved