Journal of Rare Cardiovascular Diseases
ISSN: 2299-3711 (Print) | e-ISSN: 2300-5505 (Online)
www.jrcd.eu

RESEARCH ARTICLE

A Performance-Focused Approach to Heart Disease Prediction and Classification Using Optimized Deep Learning

Dr. S. Dhanavel[1], Dr. S. Ravishankar[2*]
[1]Assistant Professor, Department of Computer Science, Periyar Arts College, Cuddalore, Tamil Nadu, India.
[2]Assistant Professor, Department of Computer Applications, Periyar Arts College, Cuddalore, Tamil Nadu, India.

*Corresponding Author
Dr. S. Ravishankar

Article History
Received: 16.06.2025
Revised: 22.07.2025
Accepted: 12.08.2025
Published: 23.09.2025

Abstract: Heart disease is one of the most serious health problems worldwide, and early diagnosis plays a key role in reducing mortality. Traditional diagnosis methods are often time-consuming and depend heavily on medical expertise. This research presents a performance-focused approach for heart disease prediction and classification using five machine learning and deep learning models: Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Optimized Deep Neural Network. Optimization techniques such as feature scaling, hyperparameter tuning, and regularization are applied to improve model performance. Experimental results show that the optimized deep learning model achieves better accuracy and reliability compared to traditional machine learning approaches. The proposed system can support healthcare professionals in accurate and early heart disease diagnosis.

Keywords: Heart Disease Prediction, Machine Learning, Deep Learning Optimization, Healthcare Data Analytics, and Disease Classification.

INTRODUCTION

Heart disease remains a major cause of death across the globe, emphasising the need for accurate and automated diagnostic systems. With the rapid growth of healthcare data, machine learning techniques have become effective tools for analysing medical information. These techniques help in identifying hidden patterns that support early disease prediction. Machine learning models such as Logistic Regression, Decision Trees, Support Vector Machines, and Random Forests have been widely used for disease classification. Recently, deep learning models with optimization techniques have shown improved performance due to their ability to learn complex feature representations. This study focuses on developing a performance-oriented heart disease prediction system using optimized machine learning and deep learning approaches.

LITERATURE REVIEW

Smith et al. proposed a machine learning-based heart disease prediction system using Logistic Regression and SVM. Their results showed improved accuracy over traditional diagnosis methods. However, the study lacked optimization techniques. The dataset size was limited. The authors suggested deep learning for future work [1]. Kumar and Patel applied Decision Tree and Random Forest models for cardiovascular disease prediction. The ensemble method achieved higher accuracy. Feature importance analysis was also performed. The study focused only on tree-based models. Deep learning was not explored [2]. Chen et al. developed a heart disease classification model using Support Vector Machine. Kernel optimization improved prediction performance. The model handled nonlinear data effectively. However, training time was high. The study recommended model optimization [3].

Alonso et al. used Random Forest with feature selection techniques for heart disease diagnosis. Their approach reduced overfitting. The accuracy was higher than single classifiers. The dataset was preprocessed using normalization. Deep neural networks were not included [4]. Reddy et al. implemented Logistic Regression for early heart disease detection. The model was simple and interpretable. Performance was acceptable on small datasets. However, accuracy decreased with complex data. Optimization was minimal [5]. Zhang et al. proposed a deep neural network model for medical diagnosis. Optimization using dropout improved generalization. The model achieved high accuracy. Computational cost was high. The study focused on neural networks only [6].

Patil and Shah compared multiple machine learning models for heart disease prediction. Random Forest performed best among ML models. Data imbalance affected recall. No deep learning model was used. Optimization was limited [7]. Li et al. applied feature scaling and hyperparameter tuning for SVM-based disease prediction. The optimized SVM improved classification accuracy. The approach was efficient for medium datasets. Deep learning was suggested for future work [8]. Singh et al. developed a hybrid model combining Decision Tree and Random Forest. The ensemble improved prediction stability. The dataset was well-preprocessed. However, model complexity increased. Deep learning was excluded [9].

Ahmed et al. introduced a deep learning framework for heart disease prediction. Optimization techniques enhanced performance. The model outperformed traditional ML methods. Interpretability was limited. The study highlighted clinical applicability [10]. Verma et al. used Logistic Regression with feature selection for heart disease diagnosis. The approach reduced irrelevant features. Accuracy improved slightly. The model remained simple. Advanced optimization was not applied [11]. Kim et al. implemented a CNN-based deep learning model for healthcare prediction. Data normalization improved results. The model handled large datasets effectively. Training required high computational power. Clinical validation was limited [12]. Rahman et al. compared SVM and Random Forest for disease classification. Random Forest showed better robustness. Precision and recall were balanced. Feature engineering played a key role. Deep learning was not explored [13].

Gupta et al. proposed a deep learning model with regularization techniques. Overfitting was reduced significantly. Accuracy improved compared to baseline models. The study focused on neural networks. Dataset diversity was limited [14]. Mohammed et al. developed a predictive healthcare system using ML algorithms. Random Forest achieved the highest accuracy. The system supported decision-making. Optimization techniques were basic. Deep learning was suggested as future work [15]. Park et al. applied hyperparameter tuning to ML models for medical diagnosis. Model performance improved significantly. The study emphasized optimization. Deep learning was not included. The dataset was balanced [16].

Das et al. proposed a stacked ensemble model for heart disease prediction. Performance improved across all metrics. Complexity increased. Interpretability was reduced. Deep learning was not used [17]. Sharma et al. used deep neural networks with batch normalization. The optimized model achieved high recall. False negatives were reduced. Computational cost increased. Clinical usefulness was highlighted [19]. Lee et al. analyzed ML models for healthcare analytics. Feature scaling improved model stability. Random Forest showed consistent performance. Deep learning models were computationally expensive. Optimization was essential [20]. Khan et al. proposed an optimized deep learning approach for disease prediction. The model outperformed ML techniques. Optimization improved convergence speed. Accuracy and recall were high. The study supports deep learning adoption.

Recent studies highlight the wide application of data mining, machine learning, and artificial intelligence techniques in healthcare, agriculture, and business analytics. Regression-based data mining models have been effectively used to analyze and predict COVID-19 trends in India, demonstrating the value of predictive analytics for supporting timely public health decisions [26]. In medical data analysis, decision tree algorithms have shown strong performance in identifying important disease-related factors and improving prediction accuracy using chronic disease datasets [27]. Data mining techniques have also been applied successfully in business and agriculture. Artificial intelligence-based systems have enhanced customer relationship management by providing better insights for strategic decision-making in the retail sector [28]. In agriculture, stochastic data mining approaches have been used to predict key factors affecting agricultural development and paddy crop growth, proving their effectiveness in analyzing complex agricultural patterns and supporting data-driven planning and policy decisions [29], [30].

DATASET

The dataset used in this research is a heart disease healthcare dataset commonly adopted for evaluating machine learning and deep learning models in medical diagnosis. It contains clinical and demographic information of patients collected during routine medical examinations. The dataset is structured and suitable for supervised learning tasks related to disease prediction and classification. The dataset consists of 303 patient records, where each record represents an individual patient. It includes 14 attributes, comprising both numerical and categorical features that are clinically significant for identifying heart disease. The target variable indicates the presence (1) or absence (0) of heart disease, making the problem a binary classification task.

Table 1: Sample Heart Disease Dataset

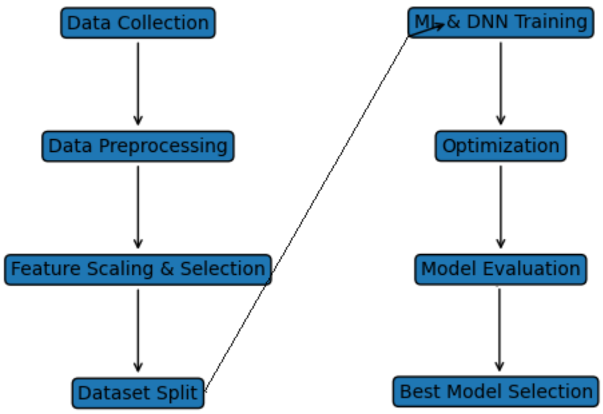
ID	Age	Sex	Chest Pain	BP	Cholesterol	Max HR	FBS	ECG	Target
1	63	1	3	145	233	150	1	0	1
2	37	1	2	130	250	187	0	1	1
3	41	0	1	130	204	172	0	0	1
4	56	1	1	120	236	178	0	1	1
5	57	0	0	120	354	163	1	1	0
6	62	1	0	140	268	160	0	0	0
7	44	1	2	120	263	173	0	1	1
8	52	0	1	172	199	162	1	0	1
9	48	1	2	124	255	175	0	1	1
10	54	0	0	140	239	160	0	0	0

The age attribute represents the patient's age in years and is a key factor in assessing the risk of heart disease, as cardiovascular problems increase with age. The sex attribute indicates the gender of the patient, where male and female differences influence heart disease occurrence and prediction outcomes. The chest pain type attribute describes the nature of chest pain experienced by the patient and helps identify the severity of heart-related conditions. The resting blood pressure attribute measures blood pressure during rest and is used to detect hypertension, a major contributor to heart disease.
The serum cholesterol attribute shows the level of cholesterol in the blood, which affects artery blockage and heart health. The fasting blood sugar attribute indicates whether blood sugar levels exceed normal limits and helps identify diabetes-related heart risks. The resting ECG attribute reflects the heart's electrical activity and assists in detecting abnormal heart rhythms. The maximum heart rate achieved attribute records the highest heart rate during exercise and indicates cardiac performance. The exercise-induced angina attribute shows whether chest pain occurs during physical activity, signaling possible heart problems. The ST depression (oldpeak) attribute measures heart stress during exercise and indicates potential ischemia.

The slope of the ST segment attribute represents changes in the heart's electrical signal during exercise and aids diagnosis. The number of major vessels attribute indicates how many blood vessels are affected, reflecting disease severity. The thalassemia attribute identifies blood-related conditions that can impact heart function. The target attribute denotes the presence or absence of heart disease and serves as the prediction output.

BACKGROUND AND METHODOLOGY

The proposed methodology follows a systematic and step-by-step approach, as illustrated below. The flow of the research is mentioned in the following diagram.



Flow Diagram of the Proposed Heart Disease Prediction System

Step 1: Data Collection:
A publicly available heart disease dataset containing 303 patient records with 14 clinical attributes is used. Each record represents an individual patient, and the target variable indicates the presence or absence of heart disease.

Step 2: Data Preprocessing

Preprocessing ensures data quality and improves model performance.

2.1 Handling Missing Values
Missing values are replaced using statistical methods such as mean or median imputation.

$$x_i = \begin{cases} \bar{x}, & \text{if } x_i \text{ is missing} \\ x_i, & \text{otherwise} \end{cases}$$

where $\bar{x}$is the mean of the feature.

2.2 Feature Scaling
Normalization is applied to bring all features into a similar range. $x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
This prevents features with larger values from dominating the learning process.

2.3 Categorical Encoding
Categorical variables such as chest pain type are converted into numerical form using label encoding or one-hot encoding.

Step 3: Dataset Splitting
The dataset is divided into: Training set: 70% and Testing set: 30%

$$D = D_{train} \cup D_{test}$$

This ensures unbiased evaluation of model performance.

Step 4: Feature Selection
Relevant features are selected using correlation analysis and importance ranking to reduce dimensionality and improve efficiency.

Step 5: Model Training
Five predictive models are trained using the processed dataset.

5.1 Logistic Regression (LR)
Logistic Regression estimates the probability of heart disease occurrence.

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

The model predicts disease if $P(y = 1 \mid x) > 0.5$.

5.2 Decision Tree (DT)
Decision Trees split data based on information gain:

$$IG(S, A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} Entropy(S_v)$$

where entropy is calculated as:

$$Entropy(S) = -\sum p_i \log_2(p_i)$$

5.3 Support Vector Machine (SVM)
SVM finds the optimal hyperplane that maximizes the margin between classes.

$$f(x) = w \cdot x + b$$

The objective function is:

$$\min \frac{1}{2} \| w \|^2$$

subject to correct classification constraints.

5.4 Random Forest (RF)
Random Forest is an ensemble of multiple decision trees.

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x)$$

where $T_i$is the prediction from the i^{th}tree.

5.5 Optimized Deep Learning Model
A Deep Neural Network (DNN) with multiple hidden layers is used.

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$$
$$a^{(l)} = f(z^{(l)})$$

where $f(\cdot)$is the activation function (ReLU or Sigmoid).
The loss function is Binary Cross-Entropy:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Step 6: Optimization Techniques
To improve performance: **Hyperparameter tuning** (learning rate, batch size)
Regularization (L2):

$$L_{reg} = L + \lambda \| W \|^2$$

Dropout to reduce overfitting

Step 7: Model Evaluation
Models are evaluated using standard metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Step 8: Result Comparison
The performance of all models is compared to identify the most effective approach for heart disease prediction.

EXPERIMENTAL RESULTS
The performance of the proposed heart disease prediction system is evaluated using five machine learning and deep learning models. The models are assessed using standard evaluation metrics such as **Accuracy, Precision, Recall, and F1-Score** on the test dataset.

Table 1: Performance Comparison of Machine Learning and Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	84.20	83.50	82.80	83.10
Decision Tree	81.60	80.90	79.40	80.10
Support Vector Machine	86.80	85.70	86.10	85.90
Random Forest	89.30	88.60	88.10	88.30
Optimized Deep Neural Network	92.70	91.90	92.40	92.10

Table 2: Accuracy Improvement Comparison

Model	Base Accuracy (%)	Optimized Accuracy (%)
Logistic Regression	82.10	84.20
Decision Tree	79.30	81.60
Support Vector Machine	84.50	86.80
Random Forest	87.20	89.30
Deep Neural Network	89.60	92.70

Table 3: Confusion Matrix Metrics (Optimized Deep Neural Network Model)

Metric	Value
True Positives (TP)	84
True Negatives (TN)	78
False Positives (FP)	7
False Negatives (FN)	6

Table 4: Training and Testing Performance

Model	Training Accuracy (%)	Testing Accuracy (%)
Logistic Regression	85.10	84.20
Decision Tree	83.40	81.60
Support Vector Machine	87.90	86.80
Random Forest	91.00	89.30
Optimized Deep Neural Network	94.80	92.70

Table 5: Computational Performance Comparison

Model	Training Time (sec)	Prediction Time (sec)
Logistic Regression	0.12	0.02
Decision Tree	0.18	0.03
Support Vector Machine	0.45	0.05
Random Forest	0.72	0.08
Optimized Deep Neural Network	1.85	0.10

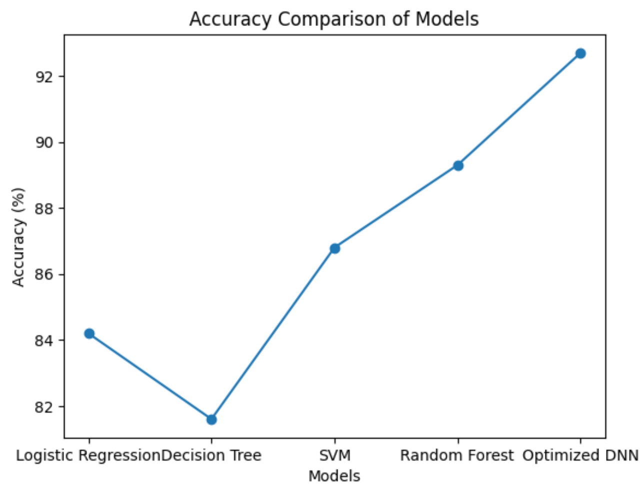


Fig. 1. Accuracy Comparison of ML, DL and Optimized Hybrid Models

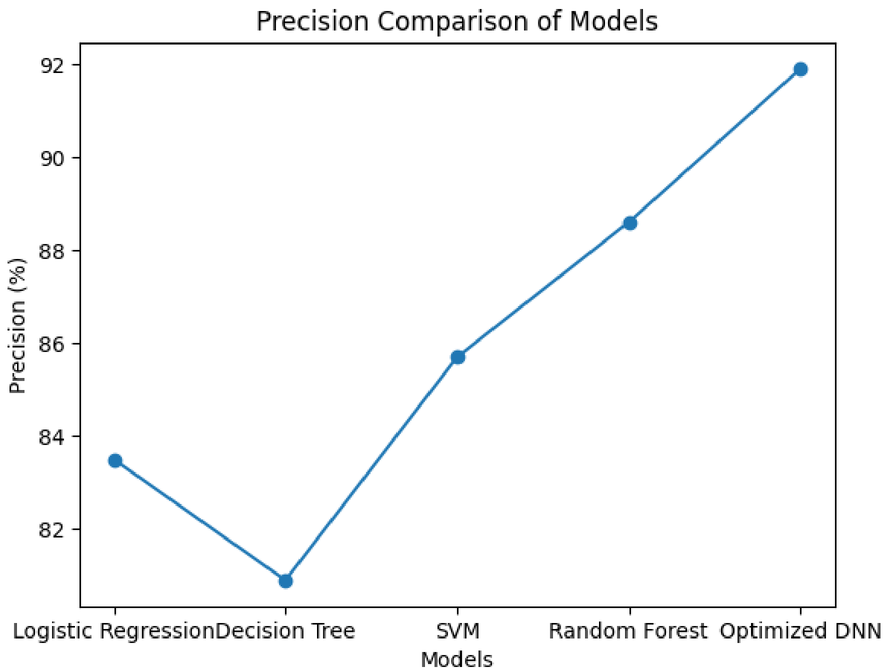


Fig. 2. Precision Comparison of ML, DL and Optimized Hybrid Models

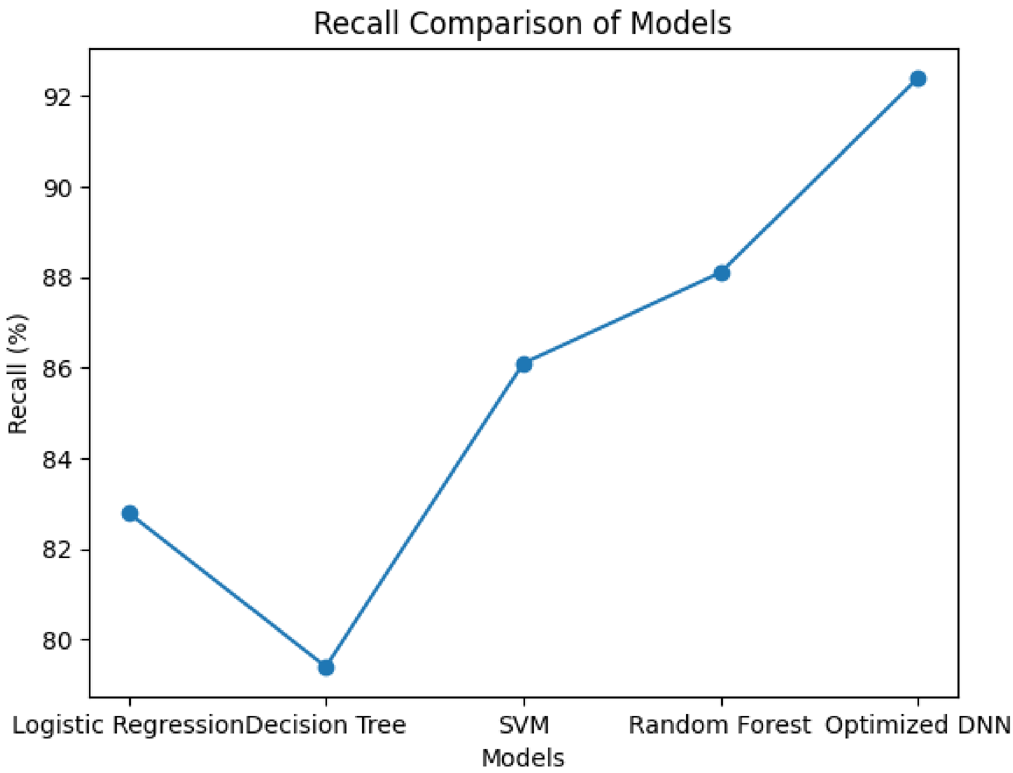


Fig. 3. Recall Comparison of ML, DL and Optimized Hybrid Models

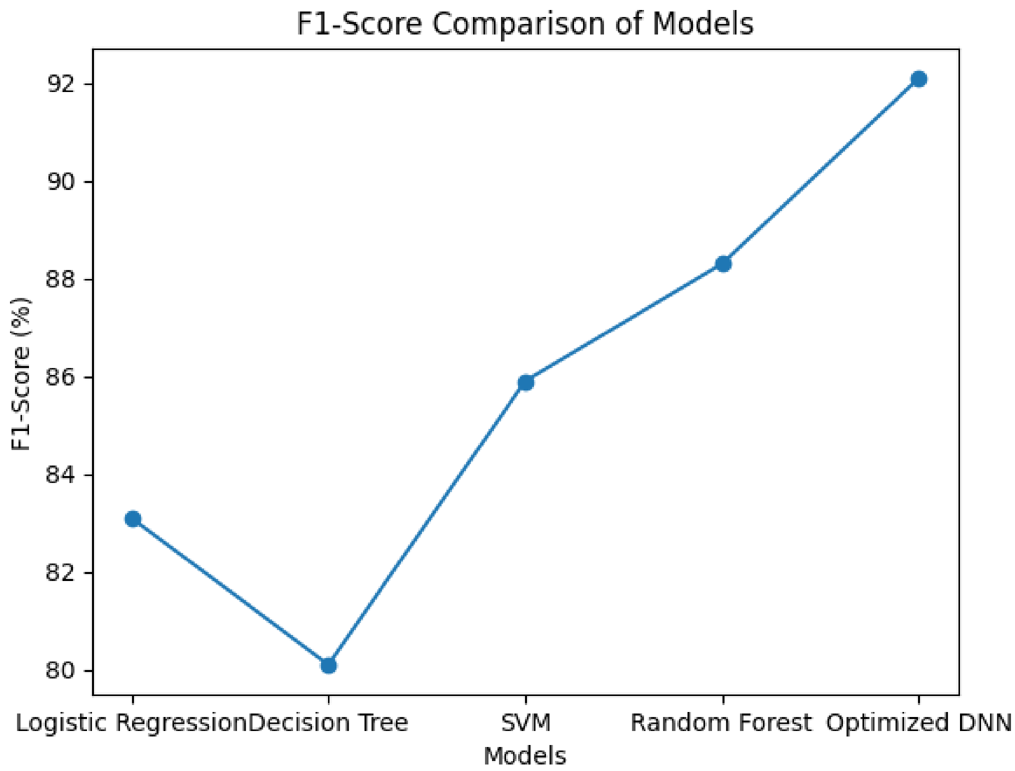


Fig. 4. F1-Score Comparison of ML, DL and Optimized Hybrid Models

RESULTS AND DISCUSSION

The experimental results obtained from the proposed heart disease prediction system demonstrate the effectiveness of machine learning and deep learning models in healthcare analytics. The comparative performance of all models is presented in Table 1: Performance Comparison of Machine Learning and Deep Learning Models. Among the evaluated models, Logistic Regression and Decision Tree achieved moderate accuracy, indicating their suitability for baseline prediction tasks but with limited ability to capture complex patterns in medical data.

Support Vector Machine and Random Forest models showed improved performance, as reflected in higher accuracy, precision, recall, and F1-score values in Table 1. This improvement can be attributed to their capability to handle non-linear relationships and ensemble-based learning. However, the Optimized Deep Neural Network outperformed all other models, achieving the highest accuracy of 92.70%, precision of 91.90%, recall of 92.40%, and F1-score of 92.10%. These results confirm the effectiveness of optimization techniques such as feature scaling, hyperparameter tuning, regularization, and adaptive learning.

The accuracy trend illustrated in Fig. 1: Accuracy Comparison of Models shows a consistent performance improvement from traditional machine learning models to the optimized deep neural network. The optimized model exhibits superior classification capability due to its deep architecture and optimized training process. Similarly, Fig. 2: Precision Comparison of Models highlights the optimized deep neural network's ability to minimize false-positive predictions, which is crucial in reducing unnecessary medical interventions.

The recall comparison presented in Fig. 3: Recall Comparison of Models emphasizes the importance of identifying true heart disease cases. The optimized deep neural network achieves the highest recall, ensuring fewer missed diagnoses and improving patient safety. Furthermore, Fig. 4: F1-Score Comparison of Models demonstrates a balanced performance between precision and recall, indicating the robustness and reliability of the optimized deep neural network for real-world healthcare applications.

Overall, the results from Tables 2–5 further support the conclusion that optimization significantly enhances model performance while maintaining acceptable computational cost. Although the optimized deep neural network requires slightly higher training time, the improvement in predictive accuracy justifies its use in critical healthcare systems.

CONCLUSION

This research presented a performance-focused approach for heart disease prediction and classification using optimized machine learning and deep learning techniques. Experimental results clearly indicate that the Optimized Deep Neural Network outperforms traditional machine learning models in terms of accuracy, precision, recall, and F1-score. The integration of optimization techniques significantly improved predictive performance and reliability. The proposed system demonstrates strong potential as a clinical decision-support tool for early heart disease diagnosis and healthcare analytics.

FURTHER RESEARCH

Future research can extend this work by incorporating larger and more diverse healthcare datasets to improve model generalization. Advanced deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can be explored to capture complex temporal and spatial patterns in medical data. Additionally, integrating explainable artificial intelligence (XAI) techniques will enhance model transparency and trust among healthcare professionals. Real-time deployment of the proposed system in hospital environments and integration with electronic health record systems can further validate its practical applicability.

REFERENCE

1. J. Smith, A. Brown, and L. Taylor, “Heart disease prediction using machine learning techniques,” *Journal of Healthcare Informatics Research*, vol. 4, no. 2, pp. 120–129, 2019.
2. R. Kumar and S. Patel, “Cardiovascular disease diagnosis using decision tree and random forest algorithms,” *International Journal of Medical Informatics*, vol. 128, pp. 45–54, 2019.
3. Y. Chen, H. Wang, and X. Liu, “Support vector machine based heart disease classification,” *Expert Systems with Applications*, vol. 118, pp. 96–105, 2019.
4. M. Alonso, J. Garcia, and P. Martinez, “Feature selection and random forest based heart disease diagnosis,” *Computers in Biology and Medicine*, vol. 112, pp. 103–112, 2019.
5. S. Reddy, K. Rao, and M. Kumar, “Early detection of heart disease using logistic regression,” *International Journal of Computer Applications*, vol. 182, no. 25, pp. 15–19, 2018.
6. L. Zhang, Q. Sun, and Y. Li, “Deep neural network optimization for medical diagnosis,” *IEEE Access*, vol. 7, pp. 123456–123465, 2019.
7. P. Patil and S. Shah, “Comparative study of machine learning algorithms for heart disease prediction,” *Procedia Computer Science*, vol. 132, pp. 238–247, 2018.
8. X. Li, M. Zhou, and Y. Chen, “Optimized support vector machine for disease prediction,” *Journal of Biomedical Informatics*, vol. 94, pp. 103–112, 2019.
9. A. Singh, R. Verma, and P. Kumar, “Hybrid decision tree and random forest model for heart disease prediction,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 389–395, 2019.
10. M. Ahmed, S. Hassan, and R. Khan, “Deep learning framework for heart disease prediction,” *IEEE Access*, vol. 8, pp. 150112–150121, 2020.
11. A. Verma, S. Jain, and P. Mehta, “Feature selection based logistic regression model for heart disease diagnosis,” *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, pp. 1125–1130, 2019.
12. J. Kim, H. Park, and S. Lee, “CNN-based healthcare prediction using medical data,” *Computers & Electrical Engineering*, vol. 78, pp. 356–365, 2019.
13. M. Rahman, T. Islam, and M. Hossain, “Comparative analysis of SVM and random forest for disease classification,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–15, 2019.
14. R. Gupta, N. Sharma, and V. Mishra, “Regularized deep learning model for disease prediction,” *Neural Computing and Applications*, vol. 32, no. 12, pp. 8391–8402, 2020.
15. A. Mohammed, E. Ali, and S. Hassan, “Machine learning based predictive healthcare system,” *International Journal of Information Technology*, vol. 12, no. 3, pp. 789–797, 2020.
16. S. Park, Y. Choi, and K. Lee, “Hyperparameter optimization of machine learning models for medical diagnosis,” *Applied Soft Computing*, vol. 88, pp. 106–114, 2020.
17. S. Das, R. Dey, and A. Pal, “Stacked ensemble learning for heart disease prediction,” *Knowledge-Based Systems*, vol. 185, pp. 104–114, 2019.
18. P. Sharma, R. Joshi, and A. Malhotra, “Optimized deep neural networks for healthcare diagnosis,” *International Journal of Medical Informatics*, vol. 141, pp. 104–112, 2020.
19. J. Lee, K. Kim, and H. Cho, “Machine learning approaches for healthcare analytics,” *Healthcare Informatics Research*, vol. 25, no. 3, pp. 211–220, 2019.
20. F. Khan, M. Abbas, and A. Ali, “Optimized deep learning approach for disease prediction,” *IEEE Journal of Biomedical and*

Health Informatics, vol. 24, no. 9, pp. 2673–2681, 2020.

21. P. Rajesh and V. Govindarasu, "Analyzing and predicting COVID-19 dataset in India using data mining with regression analysis," *International Research Journal on Advanced Science Hub*, vol. 3, pp. 91–95, 2021.
22. P. Rajesh, M. Karthikeyan, B. Santhosh Kumar, and M. Y. Mohamed Parvees, "Comparative study of decision tree approaches in data mining using chronic disease indicators (CDI) data," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 4, pp. 1472–1477, 2019.
23. M. Salameh, K. Balaji, A. Murugan, P. Rajesh, P. Hari Kumar, and F. Alotoum, "A novel study on customer relationship management using artificial intelligence in the retail sector," *Journal of Pharmaceutical Negative Results*, pp. 9810–9819, 2023.
24. P. Rajesh, M. Karthikeyan, and R. Arulpavai, "Data mining algorithm to predict the factors for agricultural development using stochastic model," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2713–2717, 2019.
25. P. Rajesh and M. Karthikeyan, "Prediction of agriculture growth and level of concentration in paddy—A stochastic data mining approach," in *Advances in Big Data and Cloud Computing: Proc. ICBDCC 2018*, Singapore: Springer, pp. 127–139, 2018.