A PSO-Based Hyperparameter Optimized Deep Neural Network for Accurate Heart Disease Classification

Ashitha Baby[1], Dr. V. Sujatha[2], Dr. Iwin C. Varkey[3]

Abstract

Cardiovascular diseases rank among the foremost causes of death worldwide, underscoring the necessity for precise predictions through effective predictive models. This study introduces an Optimized Deep Neural Network (ODNN) aimed at forecasting heart disease by utilizing electronic health record (EHR) databases sourced from the UCI machine learning repository. The methodology employed in this research is delineated as follows: initially, the research process encompasses data preprocessing, feature extraction, feature optimization, and ultimately, heart disease classification via deep learning algorithms. To begin with, missing data were addressed using K-Nearest Neighbor (KNN), while dimensionality reduction was achieved through Singular Value Decomposition (SVD) and Z-score normalization techniques. Subsequently, Principal Component Analysis (PCA) was applied to extract significant features from the dataset, which were then optimized using Particle Swarm Optimization (PSO). Furthermore, to mitigate overfitting concerns, the oversampling technique was adapted to one that is more suitable for educational contexts. Ultimately, the PRelu activation function and the move entropy loss function were employed to detect coronary heart diseases, facilitated by the optimized function. The experiments utilized the Cleveland and ORDBA datasets. The findings reveal that the proposed model attained classification accuracies of 99.14% and 98.45% for these datasets. These outcomes suggest that the proposed framework is capable of effectively and reliably predicting heart disease, thereby serving as a valuable tool for diagnosing the early stages of heart conditions through medical decision support systems.

[1]Research Scholar, Department of Computer Science,CMS College of Science and Commerce, Coimbatore, Tamil Nadu, India, Email: ashithaanki@gmail.com
[2]Associate Professor, Department of Computer Science, CMS College of Science and Commerce, Coimbatore, Tamil Nadu, India Email: sujathapadmakumar4@gmail.com
[3]Associate Professor, Department of Biotechnology, CMS College of Science and Commerce, Coimbatore, Tamil Nadu, India, Email: iwin.c.varkey@gmail.com

Keywords: Heart Disease Prediction; Deep Neural Network; Particle Swarm Optimization; Electronic Health Records; Feature Selection; Cardiovascular Disease Classification.

Introduction

Cardiovascular diseases rank among the foremost causes of mortality globally and present a considerable challenge to the current medical system1. Over the past few decades, there has been a notable increase in the burden of heart disease, particularly during the COVID-19 pandemic, which complicates the diagnosis and treatment of CVDs2. Heart disease typically refers to specific issues related to blood flow through arteries and blood vessels, thereby impacting the heart's proper functioning. Among the various types of cardiovascular diseases, rheumatic heart disease stands out as a significant contributor to mortality and disability3. Additionally, other cardiovascular conditions such as cerebrovascular disease, hypertension-related heart disease, cardiomyopathy, congenital heart disease, valvular heart disease, heart inflammation, aortic aneurysm, peripheral arterial disease, thromboembolism, and venous thrombosis are also common in the context of CVDs4.
The emergence of heart disease is frequently influenced by a combination of co-morbidities and health factors. Conditions like diabetes, cancer, and periodontitis can elevate the risk of heart-related issues5. Consequently, predicting cardiovascular disease based on the presence of co-morbidities presents a challenge. Therefore, there is a pressing need to develop advanced models for predicting and classifying heart diseases to enhance the quality of care. Intelligent heart disease prediction models have gained significance in contemporary medicine for various reasons6. Recently, electronic health record (EHR) datasets have proven valuable in constructing intelligent models that forecast heart disease due to their comprehensive patient data content. This encompasses information related to diagnoses and treatments, among other aspects7. Thus, EHR data sources can play a crucial role in advancing healthcare analytics8.

Conventionally, the identification of heart diseases involves a combination of traditional manual diagnosis and automatic diagnostic procedures. Nevertheless, conventional manual diagnosis methods tend to be cumbersome and inefficient since the analysis of the information related to the patient's condition is challenging. To enhance the accuracy and efficiency of diagnostics in medicine, new technologies have become increasingly popular and are widely used in research in healthcare. These technologies include big data analytics, artificial intelligence (AI), machine learning (ML), and deep learning (DL) [9-11]. Nonetheless, despite all the benefits that the above technologies provide for the examination of medical data, the application of such technologies in analyzing EHR datasets poses some challenges. These challenges include overfitting, high-dimensional space, and sparse data.

In this regard, this research provides a proposed optimal hyperparameter tuning method for a DNN used in classifying heart diseases. This proposed strategy incorporates sophisticated data pre-processing methods, feature selection techniques, and feature optimization processes. To ensure data quality and analytical reliability, the electronic health record (EHR) dataset will be initially preprocessed by performing missing value imputation, normalization of feature scales, and the application of dimensionality reduction techniques. After that, feature selection will be conducted by employing PCA. Feature optimization will follow where PSO will be employed to optimize the set of selected features. Lastly, a deep neural network classifier will be developed, where optimal hyper parameters and optimal activation functions will be sought out for better classification of heart diseases.

The remainder of this paper is structured to provide a clear progression of the study. It begins with a review of existing machine learning and deep learning approaches for heart disease classification, highlighting their relevance to the dataset and model design adopted in this work. This is followed by a detailed explanation of the proposed methodology, including the architecture of the optimized deep neural network and the strategies employed for hyperparameter tuning. The subsequent part presents the experimental findings and performance evaluation of the model using benchmark datasets and multiple evaluation metrics, along with a comparison against current state-of-the-art methods. The paper concludes with a summary of the key outcomes and outlines potential directions for future research.

Related Work

Cardiac prognosis has been extensively studied using computational intelligence strategies to aid in early analysis and treatment choice formulation. With the increasing availability of digital health statistics (EHRs) and medical datasets, the introduction of tools and deep knowledge of models are ending up as powerful tools for identifying cardiovascular disease patterns and chance elements[12,13]. Several research studies have investigated specific types of strategies to improve the accuracy and reliability of coronary heart disease prediction systems. Earlier studies predominantly examined conventional machine learning methods, such as logistic regression, decision tree–based models, support vector machine (SVM) algorithms, and k-nearest neighbor (KNN) classifiers. These algorithms included age, LDL cholesterol degree, blood pressure, and electrocardiogram results to anticipate the chance of heart disease. According to research on health-related factors of people suffering from a particular problem, SVM and random forest techniques generally provide

better classification accuracies than the traditional statistical method[14,15]. Ensemble mastering strategies have likewise advocated using couples of classification errors to increase prediction accuracy and reduce classification errors[16].
Several research projects have utilized public databases, including the Cleveland Coronary Heart Disease database and the UCI Machine Learning Repository database, to evaluate the prediction models of coronary heart disease. Random wood area, gradient boosting, and other group detection approaches are expected to be accurate due to their effectiveness in handling nonlinearity and complicated scientific data[17,18]. However, these approaches can be computationally costly and performance-intensive for high-dimensional data.
For addressing these constraints, task selection and dimensionality reduction approaches have been extensively studied for the prediction of coronary heart disease. The principal component analysis (PCA), reciprocal records, and correlation selection algorithms have been adopted to perform dimensional reduction on the data set for improving classification accuracy [19, 20]. Through discarding irrelevant attribute characteristics, these approaches enhance the overall efficiency of the algorithmic model.
Metaheuristic optimization algorithms have also been explored to select the optimal subset of attributes in scientific datasets. Metaheuristic algorithms that comprise GA, PSO, and ACO have been employed due to their robustness in solving complicated optimization problems [21,22]. Among those techniques, PSO has gained good-sized interest due to its rapid convergence speed and green search mechanism. Several studies have shown that PSO-based absolute feature selection improves type accuracy with the help of selecting the most informative features from a highly dimensional data set[23]. In current years, techniques for healthcare information evaluation were widely explored for deep gain knowledge due to their ability to test complex representations from large data sets. Deep, Convolutional and Recurrent Neural Networks have proven promising results in predicting coronary heart disease[24]. These models would allow automatic study of hierarchical symptom representations and capture of nonlinear dependencies between clinical attributes.
CNNs proved particularly efficient in medical data analysis and disease classifications. CNNs learned relevant skills using convolutional and pooling layers and then classified data using the connected layers only. It was claimed by several researchers that CNN-based models had high efficiency in classification of cardiovascular disorders, with clearly observable efficiencies exceeding 98% in benchmarks[25].
Deep neural networks have also found their application in predicting cardiac disorders using dependent medical data. Such models contained several hidden layers, which facilitated prediction efficiency more than shallow machine learning models[26]. Still, deep learning models require fine-tuning of the model's hyperparameters, such as rate, range, activation function, and batch size, for optimal efficiency. To resolve the issues, the current study suggested the use of hybrid approaches combining feature selection, optimization algorithm applications, and deep understanding of classification models.
These models combine the advantages of feature optimization strategies with the powerful expertise functionality of deep neural networks[27]. Reduces the dimensionality of the dataset, improves ranking accuracy, and increases version normalization. Many limitations remain despite full-size advances in machine learning, cardioprognosis techniques, and gaining deep knowledge. Many current fashions still suffer from hyperdimensionality of clinical datasets; inefficient trait choice and suboptimal hyperparameter tuning can affect reliability[28]. Therefore, an optimized framework integrating powerful feature detection with hyperparameter tuning of deep neural networks is needed to improve the overall performance of cardiovascular disease types.

MATERIALS AND METHODS:

Proposed Methodology and Model

In the present study, a novel framework termed "Optimized Hyperparameters of a Deep Neural Network for Heart Disease Classification using Particle Swarm Optimization (PSO)" is proposed. The method aims to enhance classification performance by integrating data preprocessing, feature extraction, feature optimization, and deep learning–based classification within a unified pipeline. The overall architecture of the proposed framework is illustrated in Fig. 1.

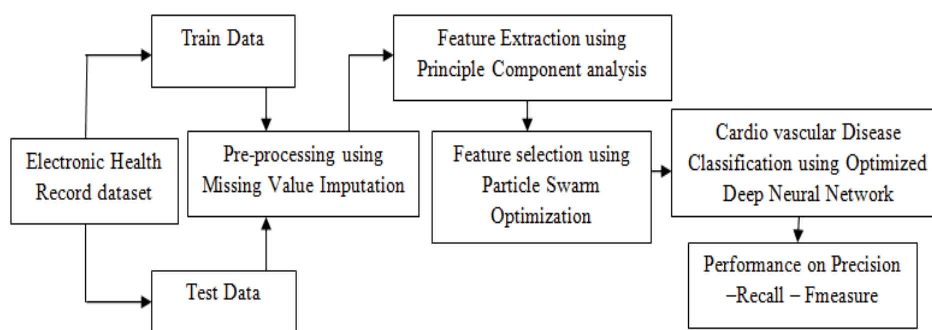

Figure 1. Architecture of Current methodology

Initially, electronic health record (EHR) data are processed using several preprocessing strategies to exceptionally improve the records and reduce dimensionality. The preprocessing stage includes missing value extraction, dimensionality reduction, and fact normalization. Subsequently, principal component analysis (PCA) is applied to extract features associated with rampant disease. These capabilities are further optimized using Particle Swarm Optimization (PSO) to identify the most applicable feature subset for the type. Finally, the optimized functions are fed to an optimized deep neural network (ODNN) classifier with tuned hyperparameters to classify cardiovascular disease conditions.

Data Preprocessing

The medical databases usually consist of many variables, noise, and missing values, which have an adverse effect on the performance of machine learning algorithms. As such, data preprocessing is used to transform the unstructured raw data to a structured form that can be used for classification purposes.

Missing Value Imputation using K-Nearest Neighbor (KNN)

There are many instances of missing values in medical databases due to the incompleteness of clinical reports. Therefore, the KNN imputation algorithm is applied for estimation of missing values according to the similarity among records. The computation of the centroid is done by taking into account the average and standard deviation of surrounding values. Similarity of records is calculated by applying the Euclidean distance. The nearest neighbors of the record are located; then, missing values are replaced with values of those records[29].

Dimensionality Reduction using Singular Value Decomposition (SVD)

The high-dimensional data sets may consist of redundant features and noise, which cause complexities during computations and affect classification accuracy. In order to solve this problem, the Singular Value Decomposition (SVD) method is applied to convert the high-dimensional data set into a lower-dimensional feature space.

SVD converts the data set matrix into three sub-matrices, which represent the patterns in the data set. This process eliminates noise from the data set and helps prevent overfitting. The matrix form of the EHR data set is:

$E_m = Di \sum_{i=0}^{n} V$ ……….Eq.1

Where V is represented as vector form of the dataset
U is represented as the Domain Parameter
Matrix operation of the data minimizes the dimensions of the dataset on considering only domain adapted attributes[30].

Data Normalization using Z-Score Normalization

Feature scaling is performed after reducing the number of dimensions through Z-score normalization. The idea behind Z-score normalization is that the data features are scaled based on their means and standard deviations. It helps ensure that all features have uniformity, which makes them contribute equally when used in training the model.

The normalized value is calculated as:

$$X'i,n = \frac{Xi,n - \mu}{\sigma_i} \text{....} Eq.2$$

Where μ, treated the mean and $\sigma_i as$ standard deviation value of i attribute respectively[19].

Feature Extraction using Principal Component Analysis (PCA)

Following normalization, principal component analysis (PCA) was applied to identify the most informative features related to the disease. This approach reduces dimensionality by converting correlated variables into a smaller set of independent variables, referred to as principal components.

This is accomplished by calculating the covariance matrix for the data set and deriving its eigenvectors and eigenvalues. Eigenvectors refer to the direction of the features, while eigenvalues indicate their significance. The extracted disease feature vector is defined as:

$\alpha_j = \sum q_j^T x$ where j=1,2,…,m ….Eq.3
Eigen Value of the α_j contains the principle disease attribute sets is {f1, f2, f3….Fn}[20].

Feature Selection using Particle Swarm Optimization

In order to enhance the classification performance, the Particle Swarm Optimization (PSO) algorithm is employed for choosing the most appropriate feature subset out of the feature sets that were generated using Principal Component Analysis (PCA). PSO is an optimization technique based on swarm intelligence where

birds' flocking is imitated. In PSO, every particle represents a possible solution, and particles' movements are guided according to pBest and gBest. Velocity is updated based on the following equation:

$V = v = v + w_1 * rand * (LBest – p) + w_2 * rand * (gBest – p)$... Eq.4
Where V is the fitness attribute selected for disease classification
P is the particle or suitable disease attribute
W1 and W2 is represented as attribute weight of the specified disease matrix

The fitness function evaluates the quality of selected features based on their contribution to heart disease classification accuracy.

Disease Classification using Optimized Deep Neural Network

The PSO-derived optimal feature set was then provided as input to the optimized deep neural network (ODNN) model. The architecture consisted of several hidden layers and a final output layer, facilitating the capture of intricate nonlinear associations among clinical parameters.

Hidden Layer

The hidden layers use convolution-like operations to generate feature maps from the customized feature matrix and function transformation. Each hidden layer applies a filter to capture significant disease styles.
An example feature map generation process is shown in Fig. 2, where the feature matrix is combined with kernel filters to produce a transformed feature representation.

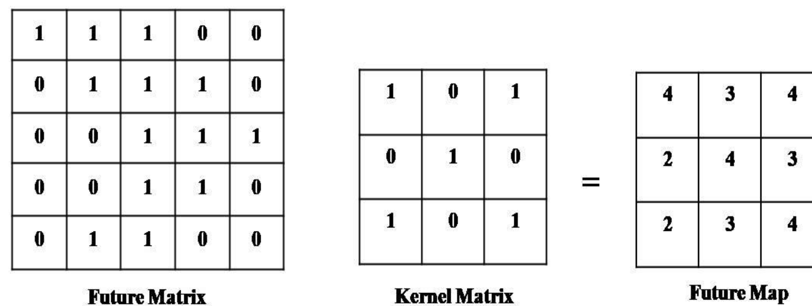

Figure 2. Feature map

To improve model convergence and reduce vanishing gradient problems, the **Rectified Linear Unit (ReLU)** activation function is applied:
$C_f = y(m^t f^t + c)$(Eq.5)
ReLU introduces nonlinearity and accelerates the learning process by retaining only positive activations.

Output Layer

The output layer plays types using a Softmax function, which converts feature distributions into feature distributions representing specific coronary heart disease indications, and also has a cross-entropy loss function to reduce ranking errors in the network. The connection between the hidden layer and the output layer is shown in Figure 3, in which the task maps are flattened and processed for the final class.

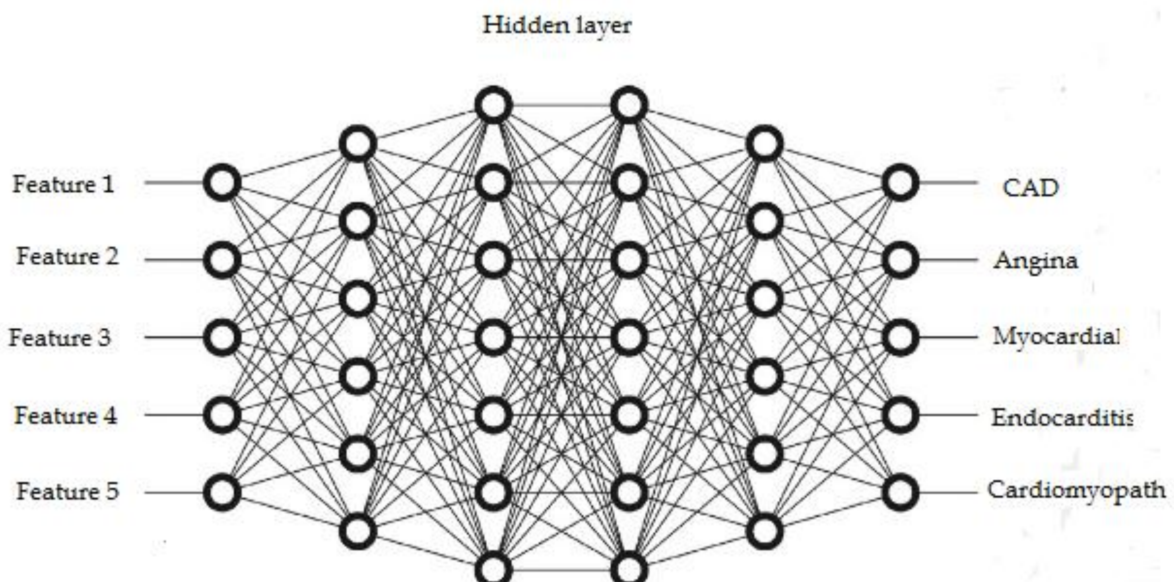

Figure 3. Connected layer of the Deep Neural Network

Hyperparameter Configuration

The performance of the proposed ODNN model depends on several hyperparameters, which are optimized during the training process. The hyperparameter configuration used in this study is shown in Table 1.

Table 1. Hyper Parameter to the Optimized Deep Neural Network

Hyper Parameter	Values
Feature Set Size	254
Learning Rate	0.06
Feature Dimensions	54
Epoch Value	50
Activation Function	ReLu
Optimizer	Gradient descent
Loss function	Cross entropy

Classification Algorithm

Algorithm : Heart Disease Classification
Input: ORDBA Dataset
Output: Categorized Classes of Heart Disease

1. Data Preprocessing
2. Missing Data Imputation with KNN
3. SVD for Dimension Reduction
4. Normalization of Data using Z-Score Normalization
5. Feature Extraction using PCA
6. PSO-based Optimal Feature Subset Selection
7. Optimized Deep Neural Network Training
8. Creation of Feature Maps in Hidden Layers
9. ReLU Activation Function &Softmax Classification
10. Heart Disease Class Output

Results And Discussion

The effectiveness of the proposed coronary heart disease classification framework was assessed using the ORDBA electronic health record (EHR) dataset obtained from the UCI Machine Learning Repository. This dataset comprises clinically relevant attributes associated with cardiovascular conditions and is frequently utilized as a benchmark for evaluating machine learning models in healthcare analytics research. The proposed model was applied using a Python programming environment with a TensorFlow backend to connect and train deep neural community architectures. Prior to the school version, the dataset was modified to be subjected to several preprocessing methods to better improve the records and reduce noise. Missing trait values were estimated using the K-nearest neighbor (KNN) approach, while singular value decomposition (SVD) is implemented to remove outlier traits and reduce dimensional redundancy and enhance the overall performance of the model by increasing the representation of underlying styles in scientific records.

To ensure reliable performance assessment, the dataset was divided into education, testing, and validation subgroups. In addition, k-fold go-validation was applied to reduce overfitting and increase version generalizability. The proposed optimized deep neural network (ODNN) version was converted into a trained experiment with batch lengths of 128 and 50 education epochs. The overall performance of the proposed model was compared with that of a baseline convolutional neural network (CNN) to demonstrate the effectiveness of the optimization strategy. Model evaluation was conducted using a confusion matrix derived from the validation dataset, which provides a detailed comparison between predicted and actual class labels. This enables the computation of key performance metrics. In this study, model performance was assessed using precision, recall, and F1-score.

Precision

Precision represents the fraction of correctly identified positive cases out of all instances predicted as positive, indicating the accuracy and reliability of the model's positive predictions.

$$\text{Precision} = \frac{\textbf{True positive}}{\boldsymbol{Truepositive + FalsePositive}} \text{...Eq.6}$$

TP denotes true positives, while FP refers to false positives.

An increase in precision indicates a reduction in the number of false positive predictions generated by the model. Figure 4 presents a comparative analysis of precision between the proposed ODNN model and the CNN

model. As illustrated in the figure, the ODNN model consistently achieves higher precision values than the CNN model.

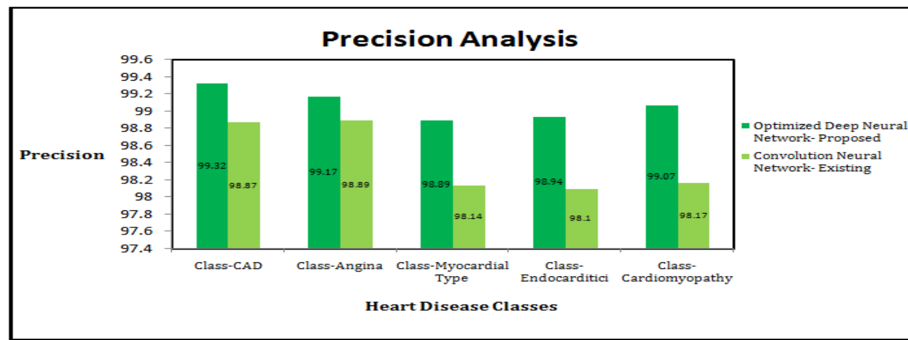

Figure 4. Performance comparison of the proposed ODNN and CNN models with respect to precision.

Recall

Recall quantifies the model's ability to correctly detect all actual positive cases within the dataset, reflecting its sensitivity to identifying disease instances.

$$\text{Recall} = \frac{\text{True positive}}{True\ positive + False\ negative\ (FN)} \text{...Eq .7}$$

Where FN represents false negatives.

For recall in a clinical decision support system, it is essential because missing out on some of the patients who are actually having a disease can cause late diagnosis and treatment. Recall analysis of the suggested ODNN algorithm for the training and testing dataset is represented in Fig. 5 below. The figure indicates that the ODNN model achieves higher recall values than the CNN model for several cardiovascular conditions, including coronary artery disease (CAD), angina, myocardial infarction, endocarditis, and cardiomyopathy.

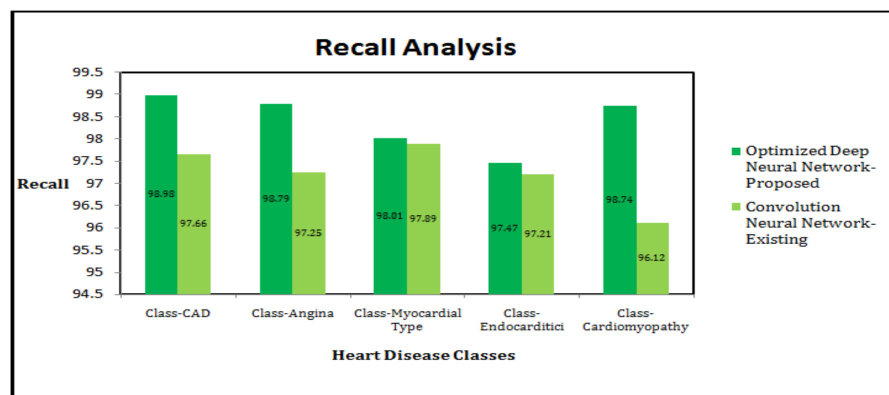

Figure 5. Recall performance comparison of the ODNN and CNN models.

F1-Score

The F1-score provides a comprehensive assessment of the classification model by integrating both precision and recall into a single performance metric.

F1=2×Precision×Recall/Precision + Recall...Eq .8

The following metric can be considered advantageous when dealing with data sets having unbalanced class distributions. The F1-scores for both the ODNN and CNN models are displayed in Figure 6. It is observed from Figure 6 that the ODNN model yields F1-scores that are always greater than those obtained from the CNN model.

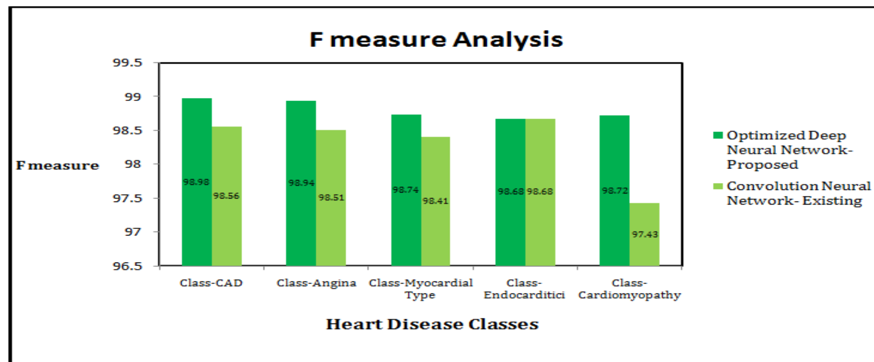

Figure 6. Comparison of ODNN and CNN models with respect to F1-score**.**

Comparative Performance Analysis

A comparative evaluation of the proposed model and existing approaches across various categories of heart diseases is summarized in Table 1. The results indicate that the proposed Optimized Deep Neural Network (ODNN) demonstrates superior performance compared to the conventional CNN, particularly in terms of classification accuracy and predictive capability.

Table 1: Comparative evaluation of performance between Optimized Deep Neural Network and Convolution Neural Network models

Disease Class	Method	Precision	Recall	F measure
Class-CAD	Optimized Deep Neural Network- Proposed	99.32	98.98	98.98
	Convolution Neural Network-Existing	98.87	97.66	98.56
Class-Angina	Optimized Deep Neural Network- Proposed	99.17	98.79	98.94
	Convolution Neural Network-Existing	98.89	97.25	98.51
Class-Myocardial Type	Optimized Deep Neural Network- Proposed	98.89	98.01	98.74
	Convolution Neural Network-Existing	98.14	97.89	98.41
Class- Endocarditic	Optimized Deep Neural Network- Proposed	98.94	97.47	98.68
	Convolution Neural Network-Existing	98.10	97.21	97.78
Class-Cardiomyopathy	Optimized Deep Neural Network- Proposed	99.07	98.74	98.72
	Convolution Neural Network-Existing	98.17	96.12	97.43

The findings clearly indicate that the suggested architecture is more efficient than other architectures using CNN, where it is able to perform with an accuracy exceeding 98.8% and the recall levels of which are almost 99%. The improved accuracy level for the proposed system can be explained by the fact that the system is able to incorporate the use of PSO technology in determining the right parameter settings and feature selection.

Furthermore, hybrid frameworks that involve the combination of deep learning techniques and evolutionary algorithms have been found to improve the performance of predictive frameworks for diseases[15,31,32]. Moreover, deep learning-based algorithms have already been found to outperform conventional machine learning methods in diagnosing diseases like heart [diseases33,34]. The findings of experiments have shown that the proposed ODNN architecture is quite useful in diagnosing heart diseases.

Conclusion

In this research study, a unique ODNN architecture has been proposed for the CHD class through digital health records presented at the UCI Machine Learning Repository. The ODNN incorporated data preprocessing, feature selection, and hyperparameter optimization implemented through the Particle Swarm Optimization (PSO) algorithm. Experimental findings demonstrated that the ODNN significantly outperforms the CNN model concerning evaluation metrics, including accuracy, precision, recall, and F1 score. In addition, ODNN technique successfully improves the accuracy of the developed models by extracting the most important features for the prediction of coronary heart diseases, which enhances the model's ability to identify at-risk patients more effectively. The proposed framework shows strong potential to be used for the automatic detection and prediction of coronary heart disease from digitized medical records. The research direction ahead includes

evaluating the performance of the framework on bigger medical datasets and discovering superior combined deep learning models.

Informed Consent

Not applicable declarations.

Author Contributions

AB and VS conceptualized the study, developed the methodology, and supervised the project. AB conducted data analysis, implemented algorithms, and interpreted the results. ICV and AB assisted with data preparation, contributed to algorithm optimization, and wrote sections of the manuscript. All authors contributed to reviewing and editing the manuscript and approved the final version for submission.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Data Availability Statement

The data are available from the corresponding author on reasonable request.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript

References

1. Di Cesare M, Perel P, Taylor S, Kabudula C, Bixby H, Gaziano TA, McGhie DV, Mwangi J, Pervan B, Narula J, Pineiro D, Pinto FJ. The heart of the world. Glob Heart. 2024; 19(1): 11. doi: 10.5334/gh.1288.
2. Krishna BA, Metaxaki M, Sithole N, Landín P, Martín P, Salinas-Botrán A. Cardiovascular disease and COVID-19: a systematic review. Int J Cardiol Heart Vasc. 2024; 54: 101482. doi: 10.1016/j.ijcha.2024.101482.
3. Berhanu H, Tegene E, Sudhakar M, Gemechu TD, Mossie A. Rheumatic heart disease burden and determinants in cardiac patients: a follow-up care concern in Ethiopia. Int J Cardiol Cardiovasc Risk Prev. 2025; 24: 200352. doi: 10.1016/j.ijcrp.2024.200352.
4. Lopez EO, Ballard BD, Jan A. Cardiovascular disease. In: StatPearls [Internet]. StatPearls Publishing. 2023.
5. Serón C, Olivero P, Flores N, Cruzat B, Ahumada F, Gueyffier F, Marchant I. Diabetes, periodontitis, and cardiovascular disease: towards equity in diabetes care. Front Public Health. 2023; 11: 1270557. doi: 10.3389/fpubh.2023.1270557.
6. Khawla R, Talukder MA, Layek MA, et al. Enhanced cardiovascular disease risk prediction using explainable machine learning and data balancing. Discov Computing. 2026; 29: 100. doi: 10.1007/s10791-026-09973-3.
7. Oliver D, Arribas M, Perry BI, Whiting D, Blackman G, Krakowski K, et al. Using electronic health records to facilitate precision psychiatry. Biol Psychiatry. 2024; 96(7): 532–542. doi: 10.1016/j.biopsych.2024.02.1006.
8. Hani SB, Ahmad MM. Integrating large-scale data analytics for cardiovascular disease prediction: a scoping review. Healthc Inform Res. 2025; 31(4): 331–346. doi: 10.4258/hir.2025.31.4.331.
9. Chowdhury MA, Rizk R, Chiu C, Zhang JJ, Scholl JL, Bosch TJ, et al. The heart of transformation: exploring artificial intelligence in cardiovascular disease. Biomedicines. 2025; 13(2): 427. doi: 10.3390/biomedicines13020427.
10. Khera R, Oikonomou EK, Nadkarni GN, Morley JR, Wiens J, Butte AJ, Topol EJ. Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC state-of-the-art review. J Am Coll Cardiol. 2024; 84: 97–114. doi: 10.1016/j.jacc.2024.05.003.
11. [Singh M, Kumar A, Khanna NN, Laird JR, Nicolaides A, Faa G, Johri AM, Mantella LE, Fernandes JFE, Teji JS, et al. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. eClinicalMedicine. 2024; 73: 102660. doi: 10.1016/j.eclinm.2024.102660.
12. Liu T, Krentz A, Lu L, Curcin V. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. Eur Heart J Digit Health. 2025; 6(1): 7–22. doi: 10.1093/ehjdh/ztae080.
13. Nasarudin NA, Al Jasmi F, Sinnott RO, Zaki N, Al Ashwal H, Mohamed EA, Mohamad MS. A review of deep learning models and online healthcare databases for electronic health records and their use for health prediction. ArtifIntell Rev. 2024; 57(9): 249. doi: 10.1007/s10462-024-10876-2.
14. Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys. 2022; 105(1): 103825. doi: 10.1016/j.medengphy.2022.103825.
15. [Li Q, Campan A, Ren A, Eid WE. Automating and improving cardiovascular disease prediction using machine learning and EMR data features from a regional healthcare system. Int J Med Inform. 2022; 163: 104786. doi: 10.1016/j.ijmedinf.2022.104786.
16. [Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. Healthcare. 2023; 11(12): 1808. doi: 10.3390/healthcare11121808.

17. Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am J Cardiol. 1989; 64(5): 304–310. doi: 10.1016/0002-9149(89)90524-9.
18. Al-Alshaikh HA, P P, Poonia RC, Saudagar AKJ, Yadav M, AlSagri HS, AlSanad AA. Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. Sci Rep. 2024; 14(1): 7819. doi: 10.1038/s41598-024-58489-7.
19. Jolliffe I. Principal component analysis. Encyclopedia of Statistics in Behavioral Science. 2005.
20. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007; 23(19): 2507–2517. doi: 10.1093/bioinformatics/btm344.
21. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 International Conference on Neural Networks. IEEE. 1995; 4: 1942–1948.
22. Yang XS. Nature-inspired optimization algorithms. Academic Press. 2020.
23. Cao K, Liu C, Yang S, Zhang Y, Li L, Jung H, Zhang S. Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model. Sci Rep. 2025; 15(1): 12406. doi: 10.1038/s41598-025-96520-7.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553): 436–444. doi: 10.1038/nature14539.
25. Ameen A, Fattoh IE, Abd El-Hafeez T, Ahmed K. Advances in ECG and PCG-based cardiovascular disease classification: a review of deep learning and machine learning methods. J Big Data. 2024; 11(1): 159. doi: 10.1186/s40537-024-01011-7.
26. Filist S, Al-Kasasbeh RT, Shatalova O, Aikeyeva A, Korenevskiy N, Shaqadan A, et al. Developing neural network model for predicting cardiac and cardiovascular health using bioelectrical signal processing. Comput Methods Biomech Biomed Eng. 2022; 25(8): 908–921. doi: 10.1080/10255842.2021.1986486.
27. Kumar A, Dhanka S, Singh J, Ali Khan A, Maini S. Hybrid machine learning techniques based on genetic algorithm for heart disease detection. Innov Emerg Technol. 2024; 11: 2450008. doi: 10.1142/S2737599424500087.
28. Mim TJH, Hassan MM, Paul BK, Mollah MSH. Machine learning approaches for cardiovascular disease prediction: a comparative study. Biomed Mater Devices. 2025; 1: 1–15. doi: 10.1007/s44174-025-00564-2. Han J, Kamber M, Pei J. Data mining: concepts and techniques. Morgan Kaufmann Publishers, Waltham. 2012.
29. Golub GH, Van Loan CF. Matrix computations. JHU Press. 2013.
30. Zhang Z, Qiu Y, Yang X, Zhang M. Enhanced character-level deep convolutional neural networks for cardiovascular disease prediction. BMC Med Inform DecisMak. 2020; 20(Suppl 3): 123. doi: 10.1186/s12911-020-1118-z.
31. Aggarwal R, Singh G, Aggarwal E. A PSO-based hybrid cardiovascular disease prediction using artificial flora algorithm. In: Artificial Intelligence-enabled Blockchain Technology and Digital Twin for Smart Hospitals. 2024; pp. 231–243. doi: 10.1002/9781394287420.ch12.
32. Sadr H, Salari A, Ashoobi MT, Nazari M. Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models. Eur J Med Res. 2024; 29(1): 455. doi: 10.1186/s40001-024-02044-7.
33. Li F, Chen Y, Xu H. Coronary heart disease prediction based on hybrid deep learning. Rev Sci Instrum. 2024; 95(1). doi: 10.1063/5.0172368.