Journal of the Maharaja Sayajirao University of Baroda
ISSN :0025-0422

PERFORMANCE COMPARISON OF NEURAL NETWORK MODELS IN PREDICTING HEART DISEASE

G.Angayarkanni, Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu : g.angayarkanni@gmail.com
Dr.M.Rajasenathipathi Associate Professor, Department of Computer Technology, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu : r.senathipathi@gmail.com

Abstract

Cardiovascular diseases (CVDs) remain among the leading global causes of death because this now affect younger populations at increasing rates throughout India and other countries. This research investigates deep learning capabilities through analysis of two leading models: the Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN). The research uses one million patient records containing 13 clinical indicators and binary classification outcomes to explore four sequential stages beginning with benchmarking followed by feature selection and then k-fold cross-validation and Bayesian Optimization for hyperparameter tuning. The initial assessment showed that MLP performed slightly better than CNN but both models achieved significant accuracy and generalization improvements after feature refinement. The prediction performance remained stable when the model was validated through multiple data split cross-validation tests. The ensemble model which combined optimized architectures produced the best results with accuracy at 92.94% and precision at 92.27% and recall at 92.70% and specificity at 93.11% and F1-score at 92.48%. The research demonstrates that developing scalable clinical heart disease detection systems requires machine learning techniques with advanced methods and thorough model optimization and strategic feature development.

Keywords: Heart Disease Prediction, Cardiovascular Diseases (CVDs), Deep Learning, Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Ensemble Learning, Feature Selection, Cross-Validation, Bayesian Optimization

I. Introduction

The most common cause of death around the world is still cardiovascular diseases (CVDs), which include a wide range of conditions affecting the heart and blood vessels. It is concerning that India is bearing an ever-increasingly disproportionate amount of this load. Recent data indicates that India has a national public health emergency due to its age-standardized CVD death rate of 282 per 100,000, which is much higher than the global average of 233 per 100,000 (PMC, 2025). Furthermore, cardiovascular diseases, which frequently strike people almost ten years earlier than in Western countries, now account for more than 26 percent of all deaths in India. Given that almost half of all heart attack patients in India are under 40, this early onset is particularly concerning. Over 18 sudden cardiac deaths among people under 40 were reported in Karnataka alone in a single month in mid-2025, drawing increased attention from medical authorities and calls for early detection measures.

Traditional diagnostic approaches, many of which rely on invasive procedures and thorough clinical evaluations, face significant difficulties due to the complexity and multifactorial nature of heart disease caused by inheritance, environmental triggers, and lifestyle choices. In this regard, the field of cardiovascular risk prediction is undergoing a revolution due to machine learning (ML), and more especially, neural networks. The ability to analyze complex, non-linear interactions within large clinical datasets patterns that are frequently undetectable to even experienced clinicians is a unique advantage of these data-driven models.

Neural networks, inspired by the architecture of the human brain, are particularly well-suited to medical diagnostics. Their ability to learn from data and generalize to new cases has demonstrated promising results across multiple applications in healthcare. Prior studies have successfully implemented Convolutional Neural Networks (CNNs) for interpreting cardiac imaging and employed ensemble learning techniques to enhance diagnostic robustness and accuracy [1,2]. Chowdhury et al.

[3] emphasize the potential of neural networks in extracting clinically relevant insights that surpass the capabilities of traditional statistical models.

A wider pattern toward intelligent healthcare systems is reflected in an increasing amount of research in this area. In addition to enhancing early detection, these models are also helping with patient stratification, resource allocation, and customized treatment planning. The need for scalable, precise, and non-invasive predictive tools is growing as cardiovascular risk keeps rising, especially in younger populations. The primary objective of this research is to conduct a rigorous, comparative evaluation of different neural network architectures for heart disease prediction. By examining their performance across a range of evaluation metrics, this study aims to identify the most effective models and configurations for clinical application. The findings are expected to contribute valuable insights toward the development of more reliable, interpretable, and clinically applicable decision-support systems ultimately assisting in early diagnosis and practical management of cardiovascular disease.

II. Literature Review

Predicting cardiovascular disease (CVD) continues to be a prominent area of research, largely due to the immense global burden posed by heart-related conditions. While traditional machine learning models have achieved commendable results in diagnostic accuracy, the advent of deep learning has introduced new possibilities for building more nuanced and effective prediction systems. Among these advancements, Multilayer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) have gained considerable traction for their ability to uncover complex, non-linear associations in structured medical data. In a notable study, Blessan and Nineta [4] utilized Bayesian Optimization to fine-tune deep neural networks for heart disease prediction, achieving high levels of precision and accuracy from structured clinical datasets. Building on this, Ahsan and Siddique [5] offered a comprehensive review of machine and deep learning methods for cardiovascular risk classification. Their work stressed the importance of rigorous evaluation methods, including cross-validation and class balancing, for ensuring model reliability. The role of feature selection has also emerged as a key component in boosting model performance. Saqlain et al. [6] demonstrated how techniques like the Fisher score and LASSO can significantly refine input features, while Bharti et al. [7] conducted comparative studies revealing that hybrid models those combining traditional machine learning with deep learning perform better when complemented with feature engineering. CNNs, in particular, have shown notable promise in healthcare applications involving time-series and sensor-based data. Shafi et al. [8] highlighted how CNNs, when applied to wearable IoT health monitoring systems, effectively captured both spatial and temporal dynamics. Meanwhile, Deepika and Balaji [9] explored the Dragonfly Algorithm, a biologically inspired metaheuristic, for tuning deep learning models. Their results indicated substantial improvements in performance and adaptability for clinical data scenarios.

Expanding the scope further, Hassani et al. [10] introduced a hybrid architecture combining neural networks with decision trees. This integration offered improved accuracy and interpretability—a critical factor in clinical contexts. Similarly, Gao and Ding [11] investigated Bayesian-based ensemble learning, showing that it outperformed conventional grid and random search methods in both predictive power and computational efficiency. In terms of real-time applicability, Khan et al. [12] merged CNNs with IoT frameworks to enable on-the-fly predictions via modular deep learning pipelines. Li et al. [13] took a genetic algorithm approach to feature selection, blending it with CNNs and SVMs to yield high-accuracy predictions on multimodal cardiac datasets. Jafari et al. [14] provided further validation of CNN efficacy, reviewing deep learning applications in cardiac MRI for CVD detection. Though rooted in imaging data, their findings underscore CNNs' broader applicability in recognizing complex diagnostic patterns. In line with ensemble strategies, Islam et al. [15] presented a stacked classifier model that consistently outperformed standalone algorithms. Likewise, Maach et al. [16] proposed a voting-based ensemble system for coronary artery disease, demonstrating enhanced sensitivity and specificity suitable for clinical deployment.

Taken together, these studies highlight the evolving landscape of heart disease prediction—one where deep learning models, particularly MLPs and CNNs, are augmented through optimization strategies like Bayesian tuning, robust feature selection, and ensemble modeling. The present research

builds on these foundations by integrating these techniques into a unified system aimed at improving predictive accuracy, model interpretability, and clinical relevance.

III. Methodology

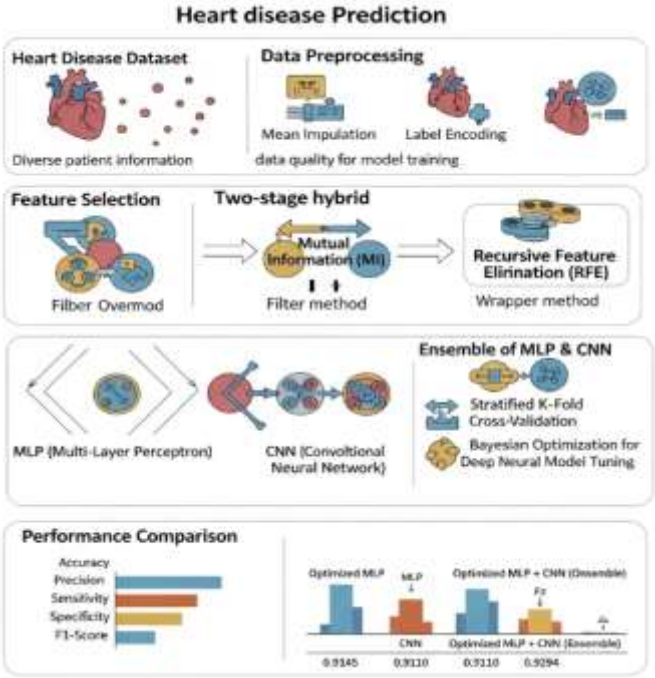

Fig 1: Proposed Methodology

A. Dataset

The dataset used in this study encompasses more than one million individual patient records, offering a comprehensive and statistically powerful basis for research into heart disease prediction. It includes a diverse mix of demographic information, clinical indicators, and both binary and categorical diagnostic variables, mirroring the types of data typically assessed in real-world cardiovascular evaluations. Notably, the distribution of cases appears balanced across both heart disease-positive and negative classes, making the dataset well-suited for binary classification tasks and helping ensure model robustness and generalizability.

Table 1: Attribute Description

S.No	Attribute	Data Type	Domain / Categories	Range (Min – Max)
1	age	Numerical	Many	~25 – 80
2	sex	Categorical	[0 = female, 1 = male]	0 – 1
3	chest	Categorical	[0, 1, 2, 3, 4]	0 – 4
4	resting_blood_pressure	Numerical	Many	~80 – 200
5	serum_cholestoral	Numerical	Many	~100 – 550
6	fasting_blood_sugar	Categorical	[0 = False, 1 = True]	0 – 1
7	resting_electrocardiographic_results	Categorical	[0, 1, 2]	0 – 2
8	maximum_heart_rate_achieved	Numerical	Many	~60 – 210
9	exercise_induced_angina	Categorical	[0 = No, 1 = Yes]	0 – 1
10	oldpeak	Numerical	Many	~0 – 6.5
11	slope	Categorical	[1, 2, 3]	1 – 3
12	number_of_major_vessels	Categorical	[0, 1, 2, 3]	0 – 3

S.No	Attribute	Data Type	Domain / Categories	Range (Min – Max)
13	thal	Categorical	[3 = normal, 6 = fixed defect, 7 = reversible]	3 – 7
14	class (target)	Categorical	[0 = no disease, 1 = disease]	0 – 1

B. Data preprocessing:

Preparing raw data is a crucial early stage in any machine learning project, ensuring that models are trained on clean, consistent input. For numerical features with missing values, mean imputation is used this technique replaces gaps with the average value of the respective column, helping preserve the dataset's overall structure without discarding potentially useful records. When working with categorical (non-numeric) variables, label encoding is applied to convert text-based categories into unique numerical values. This transformation makes the data compatible with most machine learning algorithms, allowing them to process the information effectively and identify patterns more accurately

C. Feature Selection

To improve both performance and generalization in deep learning models for heart disease prediction, a two-stage hybrid feature selection strategy was employed. This approach combines Mutual Information (MI), a filter method, with Recursive Feature Elimination (RFE), a wrapper technique. Given the dataset's size—over one million records with 13 clinical features—some variables were either redundant or carried minimal relevance to the target outcome. Including such attributes could dilute the predictive power and increase the complexity of neural network models. By isolating only the most informative features, this method enhances the model's efficiency while minimizing the risk of overfitting. It also simplifies the training process, making the architecture more scalable and clinically interpretable.

Mutual Information (MI): Mutual Information measures the degree to which a feature X shares information with the target variable Y. In essence, it captures the reduction in uncertainty about Y provided by knowing X. Features with higher MI scores exert a stronger influence on classification outcomes, allowing for a data-driven ranking that guides initial feature selection. This lays a solid foundation before RFE fine-tunes the final feature subset based on model performance.

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Where:

- p(x,y) is the joint probability distribution of feature X and label Y,
- p(x) and p(y)are the marginal distributions.

Features with higher mutual information scores are more relevant to the prediction of heart disease.

Recursive Feature Elimination (RFE): Recursive Feature Elimination (RFE) is a strategic method used in feature selection, particularly valued for its systematic, backward elimination approach. Rather than arbitrarily removing variables, RFE begins with the complete set of features and incrementally drops the least influential one at each step. This is determined through internal model evaluations, often relying on algorithms such as linear regressors or decision trees that inherently provide feature importance scores. With each iteration, the model is retrained, and features are reviewed, gradually enhancing in on those with the highest predictive value. This refined subset not only boosts the model's accuracy but also improves its generalization and interpretability by stripping away redundant or noisy inputs.

Table 2: Mutual Information Scores

thal	0.205419
number_of_major_vessels	0.161478
chest	0.152844

exercise_induced_angina	0.152259
slope	0.114614
sex	0.106728
oldpeak	0.098231
resting_electrocardiographic_results	0.079917
maximum_heart_rate_achieved	0.78251
age	0.041763
fasting_blood_sugar	0.036786
serum_cholestrol	0.010595
resting_blood_pressure	0.008857

Final Selected Features after MI + RFE:
['thal', 'number_of_major_vessels', 'chest', 'exercise_induced_angina', 'slope', 'oldpeak', 'resting_electrocardiographic_results']

D. Multilayer Perceptron (MLP):

The Multilayer Perceptron (MLP) is a foundational deep learning architecture well-suited for capturing non-linear relationships within medical data. In this study, it is applied to a binary classification problem: predicting whether a patient is at risk of heart disease (labeled as 1) or not (labeled as 0). The model begins by taking in a 13-dimensional input vector $x \in R^{13}$, representing clinical features such as age, blood pressure, cholesterol levels, and more. These inputs are passed through one or more hidden layers. Within each hidden layer, every neuron calculates a weighted sum of the inputs, adds a bias term, and then applies a non-linear activation function—commonly a ReLU or sigmoid function—to introduce complexity into the model's decision-making process. This is expressed by the formula:

$$a = \text{ReLU}(W_1 x + b_1)$$

Here, W1 is the weight matrix connecting the input layer to the hidden layer, and b1 is the bias vector. ReLU is the activation function defined as ReLU(z) = max(0, z). This step introduces non-linearity into the model, enabling it to learn more complex patterns than linear models. The output from the hidden layer, a, is then passed to the output layer, which consists of a single neuron since the task is binary classification. The output layer performs another linear transformation followed by the sigmoid activation function, which squashes the result into a probability between 0 and 1. This is given by the formula:

$$\hat{y} = \sigma(W_2 a + b_2)$$

Here, W_2 and b_2 are the weights and biases from the hidden layer to the output layer, and $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. The output $\hat{y}$ represents the predicted probability that the given patient has heart disease. To train the model, we compare the predicted output $\hat{y}$ with the actual label $y \in \{0,1\}$using the binary cross-entropy loss function**,** defined as:

$$L = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

The loss function serves as a measure of how far off the model's predictions are from the actual target values. The core objective during training is to reduce this error as much as possible by minimizing the loss across the dataset. To achieve this, optimization techniques—most commonly gradient descent are employed. This method updates the model's parameters, typically weights (W) and bias (b), by moving them step-by-step in the direction that reduces the loss. The weights are adjusted by calculating the gradient (or slope) of the loss function and shifting the parameters in the opposite direction of this gradient. The general rule for updating weights can be expressed as:

$$W \leftarrow W - \eta \cdot \frac{\partial L}{\partial W}$$

Here, η is the learning rate that controls the step size of the update.

The MLP modifies its weights and biases over several forward and backward propagation epochs to identify patterns in the data that differentiate individuals with and without heart disease. After training, the model can make accurate predictions and generalize to new, unknown patient data, which helps healthcare settings with early diagnosis and prevention.

E. Convolutional Neural Networks (CNNs):

CNNs are a class of deep learning models traditionally used for image and sequence data. Still, it has been increasingly adapted for structured tabular data where local interactions between features may be relevant. In this study, a one-dimensional CNN is employed to predict heart disease using a dataset containing one million instances, each with 13 clinical features and a binary target variable indicating the presence (1) or absence (0) of heart disease. To apply CNNs to tabular data, the input feature vector for each patient is reshaped from a flat 13-dimensional vector into a 1D array of shape 13×1. This transformation allows the model to apply 1D convolutional filters to capture local patterns and feature combinations that may not be evident in traditional linear models.

The convolutional operation at each position i is defined as:

$$z_i(1) = \sigma\left(\sum_{j=0}^{k-1} w_j . x_{i+j} + b\right)$$

where x is the input vector, k is the kernel size, w_j represents the weights of the filter, b is the bias term, and σ is the activation function, typically the Rectified Linear Unit (ReLU), defined as:

$$\mathrm{ReLU}(z) = \max(0, z)$$

This is followed by a max pooling layer that reduces the dimensionality of the feature maps, preserving the most significant values:

$$a^{(1)} = \max(z^{(1)}_i, z^{(1)}_{i+1}, \ldots, z^{(1)}_{i+m})$$

The pooled output is flattened and passed through one or more fully connected dense layers, allowing the model to learn global representations. The final layer uses a sigmoid activation function to produce a probability score indicating the likelihood of heart disease:

$$\hat{y} = \sigma(W^{(2)} a + b^{(2)}) = \frac{1}{1 + e^{-(W^{(2)} a + b^{(2)})}}$$

Model performance is evaluated using the binary cross-entropy loss function, defined as:

$$\mathcal{L} = -\left[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right]$$

Table 3: Baseline Performance of Deep Learning Models

Algorithms	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Multilayer Perceptron (MLP)	0.8990	0.8906	0.8799	0.9142	0.8852
Convolution Neural Network (CNN)	0.8913	0.8677	0.8901	0.8922	0.8788
MLP + CNN (Ensemble)	0.9071	0.8953	0.8990	0.9092	0.8971

F. Ensemble Model: Multilayer Perceptron and Convolutional Neural Network

To enhance both the accuracy and robustness of heart disease prediction, this study introduces an ensemble framework that integrates two complementary deep learning models: the Multilayer Perceptron (MLP) and the Convolutional Neural Network (CNN). Each model brings unique strengths to the table—MLP is particularly adept at capturing global relationships across all input features, while CNN excels at recognizing localized patterns and spatial hierarchies within subsets of the data. By combining these two architectures, the ensemble helps mitigate the limitations of individual models, reducing both bias and variance and resulting in more reliable predictions. In this work, we adopt a soft voting strategy to merge the outputs. Both MLP and CNN are trained independently using the same dataset, and rather than selecting a majority class, their probability outputs are averaged to form the final prediction. This approach leverages the strengths of each model, producing a more balanced

and accurate classification, especially valuable in clinical settings where prediction precision is critical. Specifically, let $\hat{y}_{MLP}$ and $\hat{y}_{CNN}$ be the output probabilities produced by the MLP and CNN models, respectively. The ensemble prediction is calculated using the following equation:

$$\hat{y}_{ensemble} = \alpha \cdot \hat{y}_{MLP} + (1 - \alpha) \cdot \hat{y}_{CNN}$$

where $\alpha \in [0,1]$ is a tunable parameter representing the weight assigned to the MLP model. In this study, we use $\alpha=0.5$ to assign equal weight to both models, ensuring balanced contribution.

After obtaining the ensemble probability score, a threshold of 0.5 is applied to convert it into a binary class label. The final model is evaluated using standard metrics such as accuracy, precision, recall, specificity, and F1-score. Experimental results show that the ensemble model consistently outperforms the individual models in all evaluation metrics. This validates the hypothesis that MLP and CNN learn complementary information from the dataset, and their combination provides a more comprehensive decision-making system for heart disease prediction.

Algorithm 1: Ensemble of MLP and CNN

Input:

Clinical dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{13}$ represents the clinical features and $y_i \in \{0, 1\}$ denotes the binary class label

Hyperparameter: $\alpha \in [0, 1]$ (ensemble weight; set to 0.5)

Threshold $t = 0.5$ for binary classification

Output:

Predicted class labels $\hat{y}_{ensemble} \in \{0, 1\}$

Procedure:

1. Preprocess dataset D:
   a. Apply mean imputation for missing numerical values
   b. Encode categorical features using label encoding
   c. Normalize input features as needed
2. Train the Multilayer Perceptron (MLP) model on the preprocessed data D
3. Train the Convolutional Neural Network (CNN) model on the same dataset D
4. For each input sample $x_i \in D$:
   a. Compute predicted probability $\hat{y}_{MLP} = \text{MLP}(x_i)$
   b. Compute predicted probability $\hat{y}_{CNN} = \text{CNN}(x_i)$
   c. Compute ensemble probability:

$$\hat{y}_{ensemble}^{(i)} = \alpha \cdot \hat{y}_{MLP} + (1 - \alpha) \cdot \hat{y}_{CNN}$$

   d. Apply classification threshold:

$$\hat{y}^{(i)} = \begin{cases} 1, & \text{if } \hat{y}_{ensemble}^{(i)} \geq t \\ 0, & \text{otherwise} \end{cases}$$

5. Evaluate model predictions $\hat{y}$ using standard metrics: accuracy, precision, recall, specificity, and F1-score

Table 4: Performance after Feature Selection

Algorithms	Accuracy	Precision	Sensitivity	Specificity	F1-Score
MLP (with selected features)	0.9062	0.8979	0.8910	0.9125	0.8944
CNN (with selected features)	0.9004	0.8821	0.8953	0.9022	0.8886
MLP + CNN (Ensemble with selected features)	0.9130	0.9066	0.9017	0.9172	0.9041

G. Stratified K-Fold Cross-Validation:

Stratified K-Fold Cross-Validation is an enhanced version of the traditional K-Fold method, designed to maintain the original class proportions within each fold. This is especially important for medical prediction tasks like heart disease diagnosis, where class imbalance—more negative cases

than positive ones, for instance—can skew performance metrics if not properly addressed. In standard K-Fold cross-validation, the dataset is split into *k* equal parts. The model is trained on *k−1* folds and tested on the remaining one, repeating the process *k* times so that each fold serves as the test set once. However, this approach doesn't guarantee balanced class representation in each fold. Stratified K-Fold solves this issue by ensuring that each fold reflects the same class distribution as the full dataset. For this study, we employed Stratified K-Fold Cross-Validation with k=5. This method allowed us to evaluate model performance more reliably, maintaining consistent proportions of heart disease and non-disease cases across all training and validation sets—a crucial factor in the clinical accuracy and fairness of predictive models.

For a performance metric M, the average performance over *k* folds is:

$$\bar{M} = \frac{1}{k}\sum_{i=1}^{k} M_i$$

Where:

- k = number of folds
- M_i = metric value (e.g., accuracy) from the i^{th} fold

This average $\bar{M}$ gives a robust estimate of how well the model will generalize.

Table 5: Performance of MLP & CNN after Cross Validation

Algorithms	Accuracy	Precision	Sensitivity	Specificity	F1-Score
MLP with selected features + CV	0.8990	0.8906	0.8799	0.9142	0.8852
CNN with selected features + CV	0.8913	0.8677	0.8901	0.8922	0.8788
MLP + CNN (Ensemble) with selected features + CV	0.9120	0.9025	0.9091	0.9134	0.9058

H. Bayesian Optimization for Deep Neural Model Tuning:

In order to further improve the performance of our deep learning models, we used Bayesian Optimization as a common hyperparameter optimization technique on all three architectures (MLP, CNN and its ensemble the MLP + CNN network). Bayesian Optimization (BO) is a method based on probabilistic modelling that provides a formal framework for efficient hyperparameter optimization, which involves learning from previous evaluations indicated in the prior point and using a surrogate model, typically a Gaussian process, to indicate the next hyperparameter point most likely to result in better performance. Unlike grid or random search, which can be computationally expensive and inefficient in high-dimensional spaces, Bayesian Optimization evaluates fewer configurations but achieves higher-quality results. For each model, the objective was to minimize the binary cross-entropy loss on the validation set by optimizing the selection of hyperparameters.

Applied Hyperparameters:

- MLP: Number of hidden layers, number of neurons per layer, learning rate, dropout rate, and batch size.
- CNN: Number and size of filters, pooling size, learning rate, activation function, dropout rate.
- MLP + CNN (Ensemble): Combined predictions from individually tuned models using a soft voting mechanism with optimized ensemble weights (α) via Bayesian search.

The optimization was performed using a 5-fold Stratified Cross-Validation loop, ensuring balanced class representation. Each iteration of the Bayesian process trained the model using different hyperparameter settings and evaluated the mean validation performance to guide the search space.

$$x^* = \arg\max_{x \in \mathcal{X}} \text{Acq}(x \mid D_{1:t})$$

Here, x^* represents the optimal hyperparameter configuration selected from the search space X, which includes parameters such as learning rate, number of hidden layers, neurons per layer, dropout rate, and batch size for MLP; and filter size, number of filters, pooling strategy, and activation function for CNN. The function $\text{Acq}(x|D_{1:t})$ known as the acquisition function, guides the selection of the next best hyperparameter set to evaluate by balancing exploration (trying new values) and exploitation (refining known good values). This is based on the historical set of evaluations $D_{1:t} = \{(x_i, f(x_i))\}_{i-1'}^{t}$ where each xi is a previously tested configuration and f(xi)f(x_i)f(xi) is its performance (such as validation accuracy or loss).

With 13 clinical characteristics and a 1-million case heart disease dataset, Bayesian optimization effectively explores the vast hyperparameter space to find configurations that improve classification performance and generalization. This eliminates the unpredictability of naive sampling and the computational expense of thorough grid search. The method greatly enhanced important measures when used separately on CNN and MLP. To further improve the diagnostic capabilities of the ensemble model, we also tuned the weight parameter α in the soft voting combination using Bayesian Optimization. As a result, this optimization technique is essential for optimizing model performance and guaranteeing accurate cardiovascular disease prediction from complex medical data.

Table 6: Performance of Optimized MLP & CNN

Algorithms	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Optimized MLP with selected feature + CV	0.9145	0.9072	0.9015	0.9186	0.9043
Optimized CNN with selected feature + CV	0.9110	0.8924	0.9103	0.9122	0.9013
Optimized MLP + CNN (Ensemble) with selected feature + CV	0.9294	0.9227	0.9270	0.9311	0.9248

The study evaluated MLP, CNN, and their ensemble across four stages: baseline, feature selection, cross-validation, and optimization. Initially, MLP outperformed CNN slightly, with 89.90% accuracy versus 89.13%, and higher precision and specificity (Table 3). The CNN showed better sensitivity, indicating a stronger ability to detect positive cases. After feature selection (Table 4), all models improved. The ensemble achieved 91.30% accuracy and an F1-score of 90.41%, highlighting the benefit of reducing irrelevant features. Cross-validation (Table 5) confirmed the stability of results, with the ensemble model achieving 91.20% accuracy and 90.58% F1-score, ensuring reliability across different data splits. With Bayesian optimization (Table 6), performance peaked. The optimized ensemble model reached 92.94% accuracy, 92.70% sensitivity, and an F1-score of 92.48%, outperforming individual models. These results demonstrate that combining MLP and CNN with feature selection, cross-validation, and tuning significantly enhances heart disease prediction in large-scale clinical data.

IV. Conclusion

This research highlights the potential of deep learning especially Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) models in successfully predicting cardiovascular disease using extensive clinical datasets. By utilizing a structured four-phase approach that incorporated initial assessment, hybrid feature selection, stratified k-fold cross-validation, and Bayesian hyperparameter optimization, the study achieved consistent performance improvements at each phase. The final ensemble model, which effectively combines the strengths of MLP and CNN frameworks, produced impressive outcomes: 92.94% accuracy, 92.27% precision, and an F1-score of 92.48%. These results

exceeded those of the standalone models, reinforcing the advantages of merging neural networks with advanced feature engineering and methodical optimization strategies. Looking to the future, subsequent versions of this research could gain from including temporal patient records, real-time data from wearable health monitoring devices, and explainable AI methodologies. Such improvements would not only enhance interpretability and applicability in real-world scenarios but also support the integration of these models into live clinical decision-making tools. Broadening the dataset to encompass multimodal sources and evaluating across diverse demographic groups would further improve the model's generalizability and clinical relevance.

Acknowledgement

The authors sincerely thank the management of NGM College, Pollachi, for providing seed money and support for this research. Their encouragement has been instrumental in the successful completion of this work.

Conflict of interest

The authors declare that there are no conflicts of interest related to this research.

References

[1] Chowdhury, M. A., et al. (2025). The heart of transformation: Exploring artificial intelligence in cardiovascular disease. Biomedicines, 13(2), 427.
[2] Meda, S., & Bhogapathi, R. (2022). An integrated machine learning model for heart disease classification and categorization. Journal of Computer Science, 18(4), 257–265. https://doi.org/10.3844/jcssp.2022.257.265
[3] Ogunpola, A., et al. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), 144.
[4] Blessan, G., & Nineta, M. (2025). Leveraging Bayesian optimization in deep learning for accurate heart disease prediction. International Journal of Scientific Research in Engineering and Management (IJSREM).
[5] Ahsan, M. M., & Siddique, Z. (2021). Machine learning-based heart disease diagnosis: A systematic literature review. arXiv preprint, arXiv:2112.06459. https://arxiv.org/abs/2112.06459
[6] Saqlain, M., Ali, R., Farid, A., & Muhammad, K. (2021). Efficient prediction system for coronary heart disease using feature selection and hyperparameter optimization. Applied Sciences, 13(1), 118. https://doi.org/10.3390/app13010118
[7] Bharti, R., Bhatia, D., & Rawal, R. (2021). Comparative analysis of ML, DL, and hybrid approaches for heart disease detection. Artificial Intelligence Review. https://doi.org/10.1007/s10462-024-10899-9
[8] Shafi, I., Ahmad, W., Khan, M., & Anwar, S. (2022). CNN and wearable devices in heart disease forecasting. Neural Computing and Applications. https://doi.org/10.1007/s00521-022-07356-5
[9] Deepika, S., & Balaji, K. (2022). Dragonfly algorithm for heart disease prediction using neural networks. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-021-11780-5
[10] Hassani, H., Ghodrati, A., & Salimi, M. (2020). Hybrid neural network–decision tree system for heart disease diagnosis. Soft Computing, 24(12), 9241–9252. https://doi.org/10.1007/s00500-019-04238-2
[11] Gao, J., & Ding, L. (2020). Disease prediction via Bayesian hyperparameter optimization and ensemble learning. BMC Medical Informatics and Decision Making, 20, 112. https://doi.org/10.1186/s12911-020-01143-4
[12] Khan, M. A., Kadry, S., & Nam, Y. (2020). Modular deep learning with IoT for heart disease prediction. Journal of Healthcare Engineering, 2020, 1–12. https://doi.org/10.1155/2020/8887516
[13] Li, X., Liu, L., Wang, Z., & Huang, W. (2021). CNN and genetic algorithm-based feature selection with SVM classifier for heart disease diagnosis. Algorithms, 14(2), 78. https://doi.org/10.3390/a14020078

[14] Jafari, M., et al. (2022). Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review. arXiv preprint. https://doi.org/10.48550/arXiv.2210.14909
[15] Islam, M. M., Tania, T. N., Akter, S., & Shakib, K. H. (2023). An improved heart disease prediction using stacked ensemble method. arXiv preprint. https://doi.org/10.48550/arXiv.2304.06015
[16] Maach, A., Elalami, J., Elalami, N., & El Mazoudi, E. H. (2022). An intelligent decision support ensemble voting model for coronary artery disease prediction in smart healthcare monitoring environments. arXiv preprint. https://doi.org/10.48550/arXiv.2210.14906