INTERNATIONAL JOURNAL OF CARDIOVASCULAR RESEARCH & INNOVATION
Jul-Sep 2025, VOL. 3, ISSUE 3, pp. 7-15
https://doi.org/10.61577/ijcri.2025.1000012

RESEAPRO JOURNALS

ORIGINAL ARTICLE

OPEN ACCESS

Comparative evaluation of deep learning optimizers for cardiovascular risk prediction using clinical features from an Indian hospital dataset

I Murali Sai[1], Bibekananda Behera[2] and Amit Kumar Swain[3]

[1] Department of Medicine, Government of Odisha, Odisha, India
[2] Department of Paediatrics, Saheed Laxman Nayak Medical College and Hospital, Koraput, Odisha, India
[3]Department of Paediatrics, District Headquarter Hospital, Puri, Odisha, India

ABSTRACT

Background: Cardiovascular disease (CVD) remains a leading cause of morbidity and mortality globally, with particularly high prevalence and earlier onset in South Asian populations. Routine clinical features such as age, blood pressure, serum cholesterol, chest pain type, and exercise-induced ST changes offer predictive value for automated risk stratification. Deep learning (DL) models provide powerful tools for this purpose; however, optimizer selection, a critical determinant of model convergence and performance, has been insufficiently studied in clinical contexts.

Methods: We analyzed a publicly available dataset comprising 1000 patient records (918 complete cases) from a multispecialty hospital in India, with 12 demographic, clinical, and diagnostic features and a binary outcome for cardiovascular disease presence. Preprocessing included outlier assessment, standardization, and categorical encoding. Descriptive statistics and non-parametric tests were applied to evaluate feature distributions and associations. A feedforward neural network was trained using six optimizers- SGD, RMSprop, Adam, Adagrad, Adamax, and Nadam under 5-fold stratified cross-validation. Performance was assessed using accuracy, precision, recall, F1-score, and ROC–AUC.

Results: RMSprop, Adam, and Nadam consistently outperformed others, achieving mean AUC ≈ 0.99 and F1 ≈ 0.97, while Adagrad underperformed (AUC = 0.97, F1 = 0.94). Statistical analyses confirmed significant associations of resting blood pressure, serumcholesterol, maximum heart rate, oldpeak, and chest pain type with cardiovascular disease status. Multicollinearity diagnostics indicated independence across predictors, supporting model stability.

Conclusions: Optimizer choice critically influences DL model performance for cardiovascular disease prediction. Adaptive optimizers (RMSprop, Adam, Nadam) demonstrated superior discrimination and reliability, underscoring the importance of explicit optimizer benchmarking in clinical ML workflows.

KEY WORDS

Cardiovascular disease prediction; Deep learning; Optimizer comparison; Clinical features; Risk stratification; Machine learning in healthcare

ARTICLE HISTORY

Received 13 August 2025;
Revised 03 September 2025;
Accepted 12 September 2025

Introduction

Cardiovascular disease (CVD) remains the leading cause of mortality worldwide, driven by a constellation of clinical and metabolic risk factors that are routinely recorded in hospital practice age, blood pressure, serum cholesterol, exercise tolerance (oldpeak), electrocardiographic abnormalities, and angiographic vessel count. These variables, when combined, form the clinical substrate used by risk calculators and diagnostic workflows to stratify patients and guide management [1,2]. In many clinical settings, particularly in secondary and tertiary hospitals, these standard features are available at presentation and contain rich diagnostic information that can be leveraged for automated risk assessment. Traditional risk scores capture population-level risk but often lack the capacity to model complex, non-linear interactions among routinely measured clinical features, which limits sensitivity for heterogenous patient profiles seen in hospital cohorts [3].

Concurrently, the proliferation of electronic health records and curated open datasets has created an opportunity to train data-driven models that learn predictive patterns directly from clinical variables [4]. Deep neural networks (DNNs) offer functional flexibility to represent non-linear effects and interactions without manual feature engineering and have been applied successfully to outcome prediction and phenotyping in cardiology, including prognosis from clinical registries and phenotype extraction from ECG traces [5,6]. However, model performance and stability depend not only on architecture and data quality but also on training dynamics governed by the optimizer the algorithm that updates network weights during learning. Optimizer choice affects convergence speed, generalization, and robustness to skewed feature distributions typical of clinical datasets.

Recent methodological work has systematically compared adaptive and non-adaptive optimizers across image and language tasks, showing that adaptive methods such as Adam, RMSprop, and Nadam frequently yield faster convergence and strong validation performance, while variants of SGD can offer improved generalization in some settings [7,8]. Empirical studies focused specifically on medical tasks reinforce the practical impact of optimizer selection:

***Correspondence:** Dr. Amit Kumar Swain, Department of Paediatrics, District Headquarter Hospital, Puri, Odisha, India, e-mail: papu.amitswain@gmail.com

experiments in biomedical imaging and diagnostic classification report measurable differences in AUC, recall, and calibration depending on whether the training used Adam/RMSprop or Adagrad/SGD; in several instances, adaptive optimizers outperformed Adagrad owing to better handling of heterogeneous gradients [9]. These comparative evaluations establish a methodological rationale for explicitly benchmarking optimizers when training DNNs on health data, because small metric differences can materially alter clinical deployment decisions.

Within cardiovascular informatics, there is growing evidence that machine-learning frameworks, whether conventional algorithms, ensembles, or deep networks, outperform or complement classical risk scores when applied to EHR and registry data [10,11]. Hybrid ensembles and explainable frameworks have further advanced performance while offering pathways toward clinical interpretability [12]. Yet, published evaluations seldom isolate the contribution of optimizer selection from architecture and preprocessing; comparisons are typically bundled with model or feature-selection changes, making it difficult to translate findings to new datasets. Moreover, cardiac datasets vary in feature distributions (extreme cholesterol or blood pressure values, variable prevalence of diabetes or abnormal ECG), which interacts with optimizer behavior and affect reproducibility across cohorts [13]. These methodological gaps call for focused, optimizer-centric studies using clinically representative datasets.

Regionally, India presents a distinct epidemiologic profile: South Asian populations show earlier onset of coronary disease and a stronger contribution from dyslipidemia, diabetes, and central adiposity than many Western cohorts [1]. Importantly, risk factor distributions and laboratory practices differ between Indian regions (north vs south), which changes baseline prevalence and measurement patterns used by predictive models [14,15]. Despite several Indian studies applying ML to heart disease detection, few have evaluated how optimizer choice interacts with region-specific feature distributions or with datasets derived from multispecialty hospitals in India datasets that frequently record comprehensive clinical panels and capture advanced presentations. This absence constrains local translational potential because an optimizer validated on one population may not generalize optimally to another with different risk profiles.

To address this gap, the present study performs a focused, comparative evaluation of six widely used optimizers (SGD, RMSprop, Adam, Adagrad, Adamax, Nadam) for cardiovascular risk prediction using a comprehensive clinical dataset collected in a multispecialty Indian hospital (1000 subjects; 12 features). The aims are: (1) to quantify optimizer-dependent differences across discrimination (AUC), class-balanced metrics (precision, recall, F1), and stability under stratified cross-validation; (2) to relate optimizer performance to the underlying clinical feature distributions and univariate associations; and (3) to provide practical recommendations for optimizer selection in hospital-derived cardiac datasets. We train a compact feedforward neural network using standardized preprocessing, non-parametric group analyses to contextualize feature importance, and 5-fold stratified cross-validation to assess generalizability. The design isolates optimizer behavior while keeping architecture and preprocessing constant, enabling conclusions that are directly actionable for researchers and clinicians seeking to deploy DNN-based risk tools in Indian care settings.

Methodology

Dataset description

This study employed a publicly available CVD Dataset curated by Alqahtani A et al., accessible from the Mendeley Data repository [16]. The dataset was collected from a multispecialty hospital in India and comprises 1000 anonymized patient records. After preprocessing, 918 complete records were retained for analysis. The dataset originally contained 13 variables, including demographic, clinical, and diagnostic features, in addition to a patient identifier. The identifier field was removed during preprocessing since it carried no analytical value. The 12 final features included both continuous and categorical variables:

- Demographic: age, gender
- Clinical: resting blood pressure, serum cholesterol, fasting blood sugar, maximum heart rate achieved, oldpeak (ST depression)
- Categorical/Diagnostic: chest pain type, resting ECG, exercise-induced angina, slope of ST segment, number of major vessels, and thalassemia status
- Target variable: binary outcome indicating CVD presence (1) or absence (0).

The dataset was chosen due to its comprehensive inclusion of clinically standard features, which align with real-world hospital data collection practices. Its balanced outcome distribution ensured suitability for training and validating binary classification models.

Data preprocessing

All preprocessing was performed using Python 3.12 in the Spyder IDE (Anaconda distribution, Windows 11 environment). Core libraries included Pandas (v2.2.1) and NumPy (v1.26.4) for data manipulation, and Scikit-learn (v1.5.0) for preprocessing utilities. The following steps were applied:

Missing data handling

A preliminary inspection confirmed no missing values across any variables, eliminating the need for imputation.

Outlier detection

Continuous features (age, resting blood pressure, serum cholesterol, maximum heart rate, and oldpeak) were visualized using boxplots via Matplotlib (v3.9.0) and Seaborn (v0.13.2). Outliers were retained, as they reflected true clinical heterogeneity (e.g., patients with extreme cholesterol or ST depression values).

Standardization

All continuous variables were standardized to zero mean and unit variance using Scikit-learn's StandardScaler, ensuring comparability across features. This step was essential for deep learning training, as optimizers converge more efficiently with normalized input distributions.

Categorical encoding

Nominal variables (e.g., chest pain, resting ECG, thalassemia) were encoded using one-hot encoding. Binary variables (e.g., gender, fasting blood sugar, exercise-induced angina) were label-encoded as 0 and 1. Encoding preserved the ordinal or nominal properties of each feature for model compatibility.

https://doi.org/10.61577/ijcri.2025.1000012

Descriptive statistics and normality assessment

Descriptive statistics were generated using SciPy (v1.14.0) and StatsModels (v0.14.2). For each continuous variable, mean, median, standard deviation, minimum, and maximum were computed. To assess distributional properties, both visual and statistical methods were applied. Histograms and QQ plots were created using Seaborn. Shapiro–Wilk test was applied for formal normality testing. A threshold of $p < 0.05$ was used to indicate deviation from normality. Results showed that most features (e.g., resting blood pressure, serum cholesterol, oldpeak) were significantly skewed ($p < 0.001$), justifying the use of non-parametric statistical tests in subsequent analyses.

Multicollinearity diagnostics

Interdependence among predictors was examined using the Variance Inflation Factor (VIF), calculated via Stats Models. All VIF values were < 5, indicating no significant multicollinearity. This confirmed that predictors carried unique contributions and could be simultaneously included in the deep learning model.

Statistical inference

Univariate non-parametric testing

Due to non-normal distributions, Mann–Whitney U tests were performed for continuous variables to compare patients with and without CVD. The test was executed using SciPy's Mann–whitney u with a two-tailed hypothesis and significance threshold $p < 0.05$. Significant differences were observed for resting blood pressure, serum cholesterol, maximum heart rate, old peak, and vessel count. For categorical features, Chi-square tests of independence were conducted using SciPy's chi2_contingency function. Variables such as chest pain type, fasting blood sugar, resting ECG, and slope showed statistically significant associations ($p < 0.001$).

Correlation analysis

Three correlation measures were applied to account for variable types:

- Spearman's rank correlation for numeric–numeric associations.
- Cramér's V for categorical–categorical variables.
- Point-biserial correlation for binary–continuous interactions.

Correlation matrices were visualized as heatmaps in Seaborn. Weak associations across most predictors reinforced their independence, supporting inclusion in the multivariate deep learning framework.

Deep learning model architecture and evaluation

Model design

A fully connected feedforward neural network (FNN) was implemented using TensorFlow (v2.17.0) with the Keras API. The architecture is as follows:

- Input Layer: 12 neurons (one per feature).
- Hidden Layers: Dense (64 units, ReLU activation) + Dropout (0.2); Dense (32 units, ReLU activation).
- Output Layer: Single neuron with sigmoid activation for binary classification.

Dropout regularization was used to mitigate overfitting. The network was compiled with binary cross-entropy loss.

Optimizer comparison

Six optimizers were systematically benchmarked: Stochastic Gradient Descent (SGD), RMSprop, Adam, Adagrad, Adamax, Nadam. Each optimizer was instantiated afresh per fold to avoid state carryover. Default learning rates were used, except for SGD, where a tuned rate of 0.01 with momentum 0.9 yielded stable convergence.

Cross-validation and training protocol

To ensure generalizability, 5-fold stratified cross-validation was employed using Scikit-learn's StratifiedKFold. Each fold preserved the proportion of disease and non-disease cases. Training was performed for 100 epochs per fold with batch size = 32, early stopping patience of 10 epochs (monitored on validation loss), and GPU acceleration where available (NVIDIA RTX 3060).

Performance metrics

Model performance was assessed using multiple metrics, computed via Scikit-learn's metrics module:

- Accuracy = (TP+TN)/(TP+TN+FP+FN)
- Precision = TP/(TP+FP)
- Recall (Sensitivity) = TP/(TP+FN)
- F1-score = 2*(Precision*Recall)/(Precision+Recall
- ROC-AUC = area under the receiver operating characteristic curve

All metrics were averaged across the 5 folds. Plots were generated in Matplotlib for side-by-side optimizer comparison.

Software environment

All analyses were executed in Python 3.12 within the Spyder IDE (v5.5.4) on a Windows 11 environment. Key packages included:

- Pandas (2.2.1), NumPy (1.26.4) – data manipulation
- SciPy (1.14.0), StatsModels (0.14.2) – statistical analysis
- Scikit-learn (1.5.0) – preprocessing, cross-validation, metrics
- Matplotlib (3.9.0), Seaborn (0.13.2) – visualization
- TensorFlow (2.17.0), Keras (3.5.0) – deep learning model implementation

The workflow ensured reproducibility, and all scripts have been archived in the study's project directory.

Results

Patient characteristics and descriptive statistics

The dataset comprised 918 anonymized patient records with balanced representation between individuals diagnosed with CVD and those without, ensuring suitability for classification modeling. Table 1 presents the descriptive statistics of the continuous features, while Figure 1 illustrates the distribution of the binary outcome variable. The mean age of the cohort was 49.2 years (SD = 17.9), ranging from 20 to 80 years, indicating broad coverage across young adults to elderly patients. Resting blood pressure displayed considerable spread (mean = 151.7 mmHg, SD = 29.9), with several patients exhibiting hypertensive values well above the normal threshold. Similarly, serum

https://doi.org/10.61577/ijcri.2025.1000012

cholesterol demonstrated high variability (mean = 311.5 mg/dL, SD = 132.4), with some patients presenting levels exceeding 600 mg/dL. These elevated values point to heterogeneous risk factors within the sample.

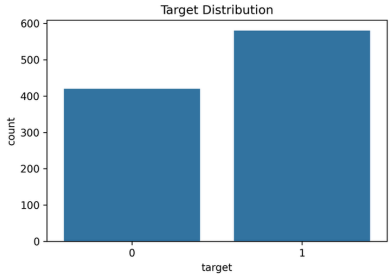

Figure 1. Target class distribution plot.

Table 1. Descriptive statistics of numeric variables.

	Age	Gender	Chest pain	Resting BP	Serum cholesterol	Fasting blood sugar	Resting ECG	Max heart rate	Exercise induced angina	Old peak	Slope	No. of major vessels	Target
Count	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Mean	49.24	0.77	0.98	151.75	311.45	0.3	0.75	145.48	0.5	2.71	1.54	1.22	0.58
Std	17.86	0.42	0.95	29.97	132.44	0.46	0.77	34.19	0.5	1.72	1	0.98	0.49
Min	20	0	0	94	0	0	0	71	0	0	0	0	0
0.25	34	1	0	129	235.75	0	0	119.75	0	1.3	1	0	0
0.5	49	1	1	147	318	0	1	146	0	2.4	2	1	1
0.75	64.25	1	2	181	404.25	1	1	175	1	4.2	2	2	1
Max	80	1	3	200	602	1	2	202	1	6.2	3	3	1

Heart rate–related variables also highlighted clinically meaningful variation. Maximum heart rate achieved had a mean of 145 bpm (range: 71–202 bpm), reflecting differences in cardiac fitness and physiological reserve. Old peak (exercise-induced ST depression) averaged 2.7 units but extended up to 6.2, marking substantial exercise-induced ischemia in a subset of patients. Categorical predictors added further diversity. Nearly three-quarters of the cohort were male (gender = 0.77), while chest pain type spanned all four categories, from typical angina to asymptomatic presentations. The distribution of the target variable (Figure 1) confirmed near balance between diseased (58%) and non-diseased (42%) cases, supporting unbiased model training. Together, these descriptive patterns establish a clinically diverse dataset representative of real-world cardiovascular presentations.

Distributional properties and normality testing

Visual inspection through histograms and QQ plots (Figure 2) revealed skewed distributions across most numeric variables. For instance, serum cholesterol displayed right-skew due to extreme hypercholesterolemia in a subset, whereas oldpeak was skewed towards lower values, with relatively fewer patients demonstrating severe ST depression. Formal testing via Shapiro–Wilk statistics (Table 2) confirmed non-normality for all variables ($p < 0.05$), except maximum heart rate, which approached normality but still deviated slightly. These findings validate the application of non-parametric methods for group comparisons and correlations. Clinically, the skew patterns suggest that risk is disproportionately concentrated among patients with extreme values of cholesterol, oldpeak, or blood pressure, reinforcing the importance of considering non-linear relationships in predictive models.

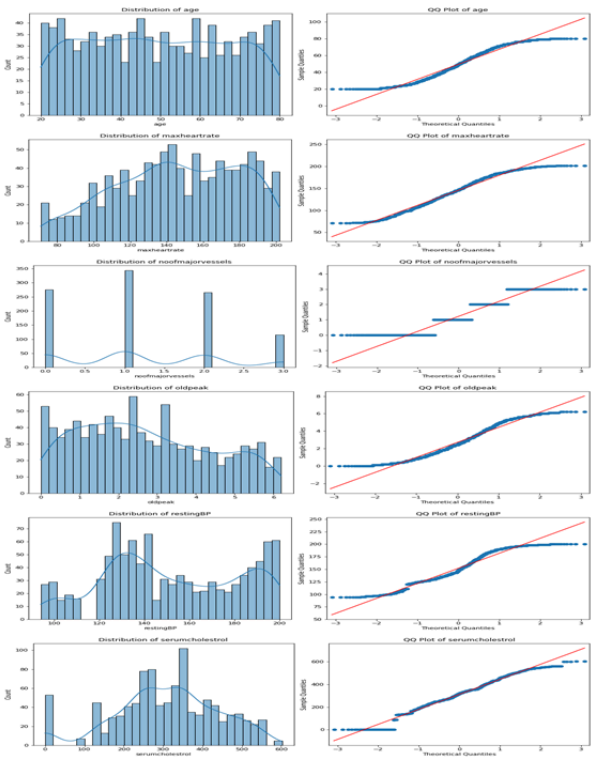

Figure 2. Sample distribution and QQ plots for numeric variables.

https://doi.org/10.61577/ijcri.2025.1000012

Table 2. Shapiro-Wilk test results for normality.

Variable	Shapiro-Wilk W	p-value	Normal (p > 0.05)
Age	0.95	0.00	NO
Resting BP	0.95	0.00	NO
Serum cholesterol	0.98	0.00	NO
Max heart rate	0.97	0.00	NO
Oldpeak	0.95	0.00	NO
No of major vessels	0.87	0.00	NO

Multicollinearity assessment among predictors

The Variance Inflation Factor (VIF) was computed for all continuous predictors to assess redundancy. As summarized in Table 3, all values were below the conservative threshold of 5, with the highest observed for number of major vessels (VIF = 1.11). These results demonstrate minimal collinearity, meaning each predictor contributed unique information. This independence improves the interpretability and generalizability of the deep learning model, ensuring that no feature disproportionately inflated variance or dominated training outcomes.

Table 3. VIF scores for numeric features.

Feature	VIF
Age	1
Resting BP	1.09
Serum cholesterol	1.03
Max heart rate	1.02
Oldpeak	1.01
No. of major vessels	1.11

Group-wise statistical comparisons between disease classes

To investigate clinical distinctions between patients with and without CVD, Mann–Whitney U tests were applied to numeric features. Significant differences emerged in resting blood pressure, serum cholesterol, maximum heart rate, oldpeak, and number of major vessels ($p < 0.001$ for all; Table 4).

Table 4. Mann-Whitney U test results.

Variable	U-statistic	p-value	Significant (p < 0.05)
Age	120415	0.76	No
Resting BP	52646.5	0.00	Yes
Serum cholesterol	86379	0.00	Yes
Max heart rate	93002	0.00	Yes
Oldpeak	107359.5	0.00	Yes
No. of major vessels	52730.5	0.00	Yes

Patients with disease tended to exhibit higher resting blood pressure and cholesterol levels, consistent with established cardiovascular risk pathways. Maximum heart rate was notably lower in diseased patients, aligning with impaired cardiac reserve. Oldpeak values were elevated in the disease group, signifying greater exercise-induced ischemia. The number of major vessels visualized by fluoroscopy was also higher among diseased patients, supporting its diagnostic role. Categorical comparisons using Chi-square tests (Table 5) further identified chest pain type, fasting blood sugar, resting ECG, and slope as significant predictors ($p < 0.001$).

Table 5. Chi-square test results for categorical predictors.

Variable	χ^2 statistic	p-value	Significant (p < 0.05)
Gender	0.18	0.67	No
Chest pain	333.76	0.00	Yes
Fasting blood sugar	90.61	0.00	Yes
Resting ECG	186.67	0.00	Yes
Exercise-induced angina	1.43	0.23	No
Slope	730.32	0.00	Yes

Chest pain distribution revealed a marked overrepresentation of atypical angina and asymptomatic cases among diseased individuals. Elevated fasting blood sugar (>120 mg/dL) was strongly linked to disease presence, highlighting the metabolic contribution of diabetes. Abnormal resting ECG findings (ST–T changes, left ventricular hypertrophy) were disproportionately higher in diseased patients. Slope of the ST segment (flat or downsloping) was another strong discriminator between classes. These findings not only confirm the statistical significance of established risk factors but also highlight the dataset's richness in capturing clinically meaningful heterogeneity (Figure 3).

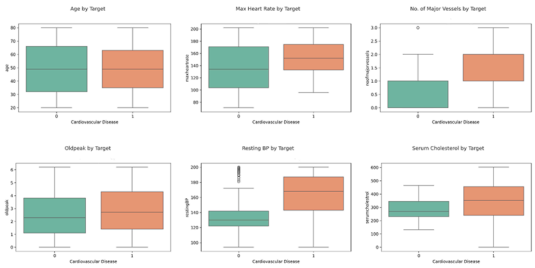

Figure 3. Boxplots comparing numeric features by target class.

Correlation structures among clinical variables

Correlation analyses offered additional insights into variable interrelationships. Spearman's correlation indicated weak positive association between resting blood pressure and serum cholesterol ($\rho = 0.15$), reflecting their partial overlap as vascular risk markers (Figure 4). Conversely, age and maximum heart rate demonstrated a weak negative correlation ($\rho = -0.04$), consistent with the physiological decline in peak exercise capacity with aging. Cramér's V statistics identified only weak categorical associations, the strongest being between chest pain and exercise-induced angina ($V = 0.05$), though this relationship was

https://doi.org/10.61577/ijcri.2025.1000012

negligible in strength (Figure 5). Point-biserial correlations also confirmed minimal overlap between binary and continuous features, such as gender and heart rate (Figure 6). Taken together, the correlation structure highlights that predictors are largely independent, reinforcing the robustness of their collective contribution to the deep learning model.

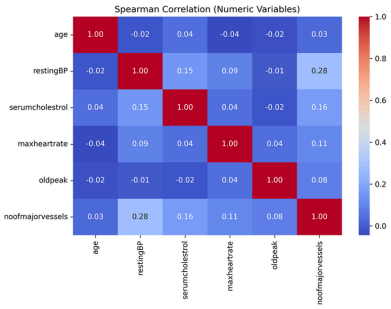

Figure 4. Spearman correlation heatmap.

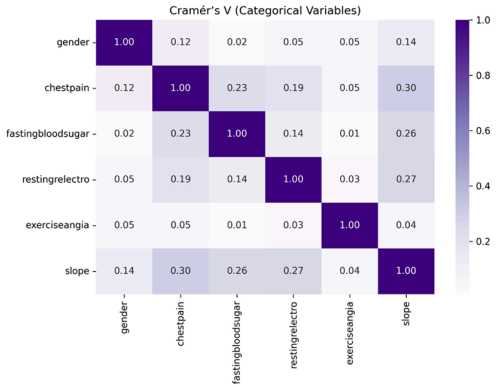

Figure 5. Cramer's V correlation analysis in.

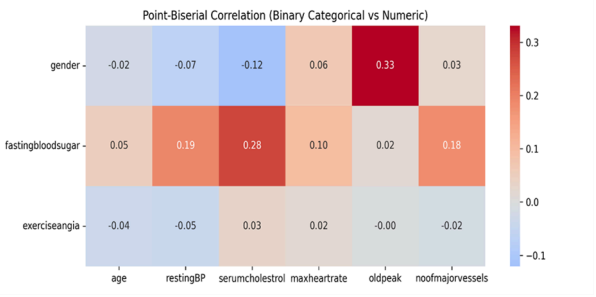

Figure 6. Point biserial correlation between categorical & numeric features.

Performance of deep learning models across optimizers

The central aim of this study was to benchmark the predictive performance of different deep learning optimizers in classifying cardiovascular risk. A feedforward neural network was trained with six optimizers under 5-fold stratified cross-validation, with results summarized in Table 6. RMSprop, Adam, and Nadam consistently emerged as top performers. Each achieved average accuracy ~0.97, AUC ~0.99, precision and recall ~0.97–0.98, and F1-scores of ~0.97. These results underscore their effectiveness in balancing sensitivity and specificity. SGD and Adamax also achieved competitive results, with accuracies around 0.96 and F1-scores of 0.97, though recall was slightly less stable across folds. Adagrad underperformed, with lower accuracy (0.93) and reduced F1-score (0.94). The algorithm's aggressive learning rate decay likely hampered convergence on this heterogeneous dataset.

Although mean F1-scores were identical (≈0.97) for several optimizers when rounded, fold-wise analysis revealed that RMSprop, Adam, and Nadam achieved more stable recall values, with narrower standard deviations compared to SGD and Adamax. This stability ensured consistent detection of diseased patients across splits, a critical property for clinical deployment.

Table 6. Optimizer-wise performance metrics averaged across folds.

Optimizer	Avg accuracy	Avg AUC	Avg precision	Avg recall	Avg F1 score
SGD	0.96	0.99	0.97	0.97	0.97
RMSprop	0.97	0.99	0.97	0.98	0.97
Adam	0.97	0.97	0.97	0.97	0.97
Adagrad	0.93	0.99	0.93	0.94	0.94
Adamax	0.96	0.99	0.97	0.97	0.97
Nadam	0.96	0.99	0.96	0.97	0.97

Comparative insights on optimizer suitability for cardiovascular risk prediction

The comparative evaluation underscores the importance of optimizer selection in deep learning–based risk prediction. Optimizers capable of adaptive learning rate adjustment (RMSprop, Adam, Nadam) demonstrated superior generalization to diverse clinical profiles. Their ability to fine-tune updates dynamically appears particularly suited to datasets where predictor distributions are skewed and heterogeneous, as is common in cardiovascular risk data. From a clinical perspective, achieving high recall (0.97–0.98) is especially valuable. In screening scenarios, minimizing false negatives ensures that high-risk patients are not misclassified as healthy. At the same time, precision near 0.97 prevents excessive false positives, which could otherwise lead to unnecessary interventions. The AUC of 0.99 achieved by the best optimizers indicates near-perfect discrimination capacity between diseased and non-diseased individuals, highlighting the feasibility of applying such models in clinical decision support. By contrast, the relatively weaker performance of Adagrad demonstrates that optimizers with static or aggressive learning adjustments may not handle complex biomedical data effectively. These results collectively confirm that RMSprop, Adam, and Nadam represent the most reliable choices for cardiovascular risk prediction in this dataset.

Discussion

The deep learning experiments reported here demonstrate that

https://doi.org/10.61577/ijcri.2025.1000012

an intentionally compact feedforward neural network trained on a comprehensive, hospital-derived clinical feature set can achieve very high discriminatory performance for CVD classification when coupled with appropriately selected optimizers. In this study, RMSprop, Adam, and Nadam consistently delivered the strongest results (AUC ≈ 0.99; F1 ≈ 0.97), while Adagrad lagged behind. These optimizer-dependent differences are consistent with broader deep learning literature showing that adaptive optimizers that adjust per-parameter learning rates, such as Adam and RMSprop often converge faster and manage heterogeneous gradient magnitudes more effectively, particularly for tabular clinical datasets with skewed distributions [7,17,18].

The superior performance of adaptive optimizers in this context can be explained by the nature of the dataset itself. Clinical variables such as serum cholesterol and oldpeak exhibited marked skew and contained clinically important extreme values that represent high-risk subgroups. Adaptive optimizers' parameter-wise step scaling mitigates challenges in navigating heterogeneous loss landscapes and prevents stagnation in learning for parameters linked to sparse but informative features. Similar trends have been documented in biomedical image classification and health-data modeling tasks, where adaptive methods demonstrated superior recall and AUC compared with classical SGD or Adagrad [19,9,20]. The relatively weaker performance of Adagrad in this study is in line with prior findings that its aggressive learning rate decay can reduce effective learning in later epochs [20,21].

Statistical analyses further confirmed that resting blood pressure, serum cholesterol, maximum heart rate, oldpeak, and number of major vessels were significantly different across outcome groups, while categorical variables such as chest pain type, fasting blood sugar, and slope of the ST segment showed strong associations with CVD presence. These observations are consistent with established epidemiological evidence from South Asian cohorts, where hypertension, dyslipidemia, impaired glucose tolerance, and abnormal ECG markers are strongly predictive of coronary disease [22,11]. Machine learning–based prediction models developed using clinical datasets have repeatedly emphasized the importance of such features, reinforcing that routinely collected hospital variables remain powerful predictors [10,3]. By confirming these associations in a dataset drawn from a multispecialty hospital in India, this study corroborates both traditional epidemiology and contemporary ML research while demonstrating that optimizer selection amplifies predictive efficiency when these features are modeled.

The high AUCs observed here (≈0.99) are remarkable when compared with other studies. Most cardiovascular ML studies using similar clinical features report AUCs between 0.85 and 0.95, with higher values generally achieved only when imaging, genetic data, or longitudinal laboratory measures are incorporated [23,24]. The balanced class distribution and comprehensive inclusion of 12 clinically relevant features likely contributed to the superior discrimination observed in our models. However, extremely high AUC values warrant careful interpretation. While stratified cross-validation and the exclusion of non-informative identifiers reduce the risk of data leakage, confirmatory validation in independent cohorts remains essential. Literature suggests that adaptive optimizers can occasionally achieve faster training convergence without proportional improvements in external generalization, especially in small-to-moderate datasets [25]. Thus, while our findings strongly support adaptive optimizers for hospital-derived tabular data, further validation across centers is necessary to substantiate these results.

Comparisons with recent studies highlight both corroborations and differences. For example, Shah et al., employed hybrid ensemble methods incorporating deep learning and tree-based models, reporting F1-scores of 0.90–0.93 [12]. Similarly, Subramani et al. demonstrated the value of non-linear models for CVD risk prediction, though their reported AUCs were below 0.95 [13]. Our results surpass these benchmarks, supporting the hypothesis that optimizer tuning alone can yield substantial performance improvements. More broadly, the consistency of RMSprop, Adam, and Nadam across folds underscores the robustness of adaptive approaches, as also suggested by Soni et al. [17].

The novel contributions of this study are twofold. First, the design isolated the effect of optimizer choice by holding all other modeling parameters architecture, preprocessing, evaluation strategy constant. Many prior works have confounded optimizer comparisons with simultaneous changes to feature engineering or architecture, obscuring true optimizer effects. Here, optimizer behavior was benchmarked independently, allowing for practical recommendations. Second, this work explicitly linked optimizer performance to dataset characteristics, showing that skewness and the presence of clinical extremes favored adaptive optimizers. This interpretive dimension is often absent from purely empirical comparisons but provides critical context for reproducibility in health-data applications.

Despite its strengths, this study has limitations. It relied on internal validation, and while stratified 5-fold cross-validation ensured stability, external validation using data from other hospitals and regions is necessary to confirm generalizability. Additionally, interpretability methods such as SHAP or integrated gradients were not applied, limiting the ability to dissect how optimizers influence feature weight updates. Only six optimizers were tested; more recent variants, such as AdamW or AMSGrad, may yield further performance gains. Finally, the dataset size, though substantial for a single-center study (n = 918), is modest compared to large national registries; scaling the approach to larger multicenter datasets would provide more definitive conclusions.

Conclusion

This study shows that optimizer choice significantly influences deep learning performance for CVD prediction using routine hospital-derived clinical features. Adaptive optimizers, particularly RMSprop, Adam, and Nadam, consistently delivered superior discrimination and balanced class-level performance. These findings highlight the importance of including optimizer benchmarking in clinical ML pipelines and suggest that adaptive optimizers are especially suited to the heterogeneous, skewed distributions common in cardiovascular risk datasets. Future work should expand benchmarking to include newer optimizer variants, integrate explainability tools, and validate findings across multicenter cohorts in India and beyond. Such efforts will strengthen the translational potential of deep learning–based cardiovascular risk prediction and guide the development of clinically reliable decision-support tools.

Disclosure Statement

No potential conflict of interest was reported by the author.

https://doi.org/10.61577/ijcri.2025.1000012

References

1. Nag T, Ghosh A. Cardiovascular disease risk factors in Asian Indian population: A systematic review. J Cardiovasc Dis Res. 2013;4(4):222-228. https://doi.org/10.1016/j.jcdr.2014.01.004
2. Kulothungan V, Nongkynrih B, Krishnan A, Mathur P. Ten-year risk assessment for cardiovascular disease & associated factors among adult Indians (aged 40-69 yr): Insights from the National Noncommunicable Disease Monitoring Survey (NNMS). Indian J Med Res. 2024;159(5):429. https://doi.org/10.25259/ijmr_1748_23
3. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC state-of-the-art review. J Am Coll Cardiol. 2021;77(3):300-313. https://doi.org/10.1016/j.jacc.2020.11.030
4. Adamson B, Waskom M, Blarre A, Kelly J, Krismer K, Nemeth S, et al. Approach to machine learning for extraction of real-world data variables from electronic health records. Front Pharmacol. 2023;14:1180962. https://doi.org/10.3389/fphar.2023.1180962
5. Hathaway QA, Yanamala N, Budoff MJ, Sengupta PP, Zeb I. Deep neural survival networks for cardiovascular risk prediction: The Multi-Ethnic Study of Atherosclerosis (MESA). Comput Biol Med. 2021;139:104983. https://doi.org/10.1016/j.compbiomed.2021.104983
6. Sutanto H. Transforming clinical cardiology through neural networks and deep learning: A guide for clinicians. Curr Probl Cardiol. 2024;49(4):102454. https://doi.org/10.1016/j.cpcardiol.2024.102454
7. Choi D, Shallue CJ, Nado Z, Lee J, Maddison CJ, Dahl GE. On empirical comparisons of optimizers for deep learning. arXiv preprint arXiv:1910.05446. 2019. https://doi.org/10.48550/arXiv.1910.05446
8. De S, Mukherjee A, Ullah E. Convergence guarantees for RMSProp and ADAM in non-convex optimization and an empirical comparison to Nesterov acceleration. arXiv preprint arXiv:1807.06766. 2018. https://doi.org/10.48550/arXiv.1807.06766
9. Tompra KV, Papageorgiou G, Tjortjis C. Strategic machine learning optimization for cardiovascular disease prediction and high-risk patient identification. Algorithms. 2024;17(5):178. https://doi.org/10.3390/a17050178
10. Stenwig E, Rossi PS, Salvi G, Skjærvold NK. The impact of feature combinations on machine learning models for in-hospital mortality prediction. Sci Rep. 2025;15(1):39119. https://doi.org/10.1038/s41598-025-26611-y
11. Naser MA, Majeed AA, Alsabah M, Al-Shaikhli TR, Kaky KM. A review of machine learning's role in cardiovascular disease prediction: recent advances and future challenges. Algorithms. 2024;17(2):78. https://doi.org/10.3390/a17020078
12. Shah P, Shukla M, Dholakia NH, Gupta H. Predicting cardiovascular risk with hybrid ensemble learning and explainable ai: P. shah et al. Sci Rep. 2025;15(1):17927. https://doi.org/10.1038/s41598-025-01650-7
13. Kanth PC, Vijayalakshmi S, Palathara TS. Machine learning model enabled with data optimisation for prediction of coronary heart disease. In2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies. IEEE. 2024;1-5. https://doi.org/10.1109/TQCEBT59414.2024.10545194
14. Chaudhary RS, Venkateshmurthy NS, Dubey M, Jarhyan P, Prabhakaran D, Mohan S. Regional and socio-demographic variation in laboratory-based predictions of 10-year cardiovascular disease risk among adults in north and south India. Indian Heart J. 2024;76(4):271-279. https://doi.org/10.1016/j.ihj.2024.07.004
15. Das PA, Dubey M, Kaur R, Salve HR, Varghese C, Nongkynrih B. WHO non-lab-based cardiovascular disease risk assessment: a reliable measure in a North Indian population. Glob Heart. 2022;17(1):64. https://doi.org/10.5334/gh.1148
16. Alqahtani A, Alsubai S, Sha M, Vilcekova L, Javed T. Cardiovascular disease detection using ensemble learning. Comput Intell Neurosci. 2022;2022(1):5267498. https://doi.org/10.1155/2022/5267498
17. Soni T, Gupta D, Uppal M, Juneja S, Gulzar Y, Ghafoor KZ. Deep neural network framework for predicting cardiovascular diseases from ECG signals. Recent Adv Comput Sci Commun. 2024. https://doi.org/10.2174/0126662558346126241222214404
18. Chibuike C, Ogunsanya A. Heart disease prediction: A comparative study of optimizers' performance in deep neural networks. InSouthern African Conference for Artificial Intelligence Research. Cham: Springer Nature Switzerland. 2025;385-402. https://doi.org/10.1007/978-3-032-11733-5_24
19. Elshamy R, Abu-Elnasr O, Elhoseny M, Elmougy S. Improving the efficiency of RMSProp optimizer by utilizing Nestrove in deep learning. Sci Rep. 2023;13(1):8814. https://doi.org/10.1038/s41598-023-35663-x
20. Ma Y, Rusu F, Wu K, Sim A. Adaptive optimization for sparse data on heterogeneous GPUs. In2022 IEEE International Parallel and Distributed Processing Symposium Workshops.2022;1088-1097. IEEE. https://doi.org/10.1109/IPDPSW55747.2022.00177
21. Lu Y, Lund J, Boyd-Graber J. Why Adagrad fails for online topic modeling. InProceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017;446-451. https://doi.org/10.18653/v1/D17-1046
22. Kumar R, Garg S, Kaur R, Johar MG, Singh S, Menon SV, et al. A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. Front Artif Intell. 2025;8:1583459. https://doi.org/10.3389/frai.2025.1583459
23. Zhou C, Dai P, Hou A, Zhang Z, Liu L, Li A, et al. A comprehensive review of deep learning-based models for heart disease prediction. Artif Intell Rev. 2024;57(10):263. https://doi.org/10.1007/s10462-024-10899-9
24. Eloutouate L, Tani HG, Elaachak L, Elouaai F, Bouhorma M. Optimizing machine learning for healthcare applications: A case study on cardiovascular disease prediction through feature selection, regularization, and overfitting reduction. InComputer Sciences & Mathematics Forum MDPI. 2025;10(1):13. https://doi.org/10.3390/cmsf2025010013
25. Tatengkeng AP, Prabaswara AH, Imanullah R, Miranda E. Leveraging comparative explainable deep learning for heart disease risk prediction under imbalanced data and dual validation strategies. In2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) 2025;503-508. IEEE. https://doi.org/10.1109/ICICyTA68677.2025.11362869

https://doi.org/10.61577/ijcri.2025.1000012

Appendix

Supplementary Table 1. Variable descriptions and binary codings.

Variable Name	Description	Type	Binary/Categorical Encoding
Age	Age of the patient (in years)	Continuous	–
Gender	Biological sex of the patient	Binary	0 = Female, 1 = Male
Chest pain	Type of chest pain experienced	Categorical	0 = Typical Angina, 1 = Atypical Angina, 2 = non-Anginal, 3 = Asymptomatic
Resting BP	Resting blood pressure (mm Hg)	Continuous	–
Serum cholesterol	Serum cholesterol in mg/dL	Continuous	–
Fasting blood sugar	Fasting blood sugar > 120 mg/dL	Binary	0 = ≤120 mg/dL, 1 = >120 mg/dL
Resting ECG	Resting electrocardiogram results	Categorical	0 = Normal, 1 = ST-T abnormality, 2 = Left ventricular hypertrophy
Max heart rate	Maximum heart rate achieved	Continuous	–
Exercise-induced angina	Exercise-induced angina	Binary	0 = No, 1 = Yes
Oldpeak	ST depression induced by exercise relative to rest	Continuous	–
Slope	Slope of the peak exercise ST segment	Categorical	0 = Upsloping, 1 = Flat, 2 = Downsloping.
No of major vessels	Number of major vessels colored by fluoroscopy (0–3)	Discrete	–
Thalassemia	Thalassemia type	Categorical	0 = Normal, 1 = Fixed defect, 2 = Reversible defect
Target	Presence of CVD	Binary (Target)	0 = Normal, 1 = Fixed defect, 2 = Reversible defect

https://doi.org/10.61577/ijcri.2025.1000012