

Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis

Sunanda Budihal^{1, ✉}, Sheetalrani R. Kawale², Nitin Agarwal³, Bhagirathi Halalli¹

1. Computer Science, Government First Grade College, Bagalkot, IND

2. Computer Science, Karnataka State Women's University, Bijapur, IND

3. Consulting Physician and Cardiologist, Ayush Multi Specialty Hospital and Research Center Pvt. Ltd., Bijapur, IND

Received: March 17, 2026 | Review began: March 26, 2026 | Review ended: April 14, 2026 | Published: May 08, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Cardiovascular (cardiac) diseases are among the leading causes of death worldwide and require effective and dependable computational methods for early diagnosis and risk assessment. The proposed study introduces a Feature-Fusion Artificial Neural Network (FF-ANN) to predict cardiovascular diseases by directly dividing the numerical and categorical clinical variables into separate learning branches and fusing them via a gated fusion process. Real-time clinical data were obtained from approximately 2,000 patient records at Ayush Multispecialty Hospital and Research Centre, Vijayapura, Karnataka, India. A leakage-free stratified validation protocol was used to evaluate the model's performance and compare it with a CatBoost baseline and a standard ANN. The proposed FF-ANN achieved a test accuracy of 96.75%, a Matthews correlation coefficient of 0.931, and an area under the receiver operating characteristic curve of 0.996, outperforming the standard ANN. CatBoost also produced competitive results, further demonstrating the strong predictive value of categorical diagnostic variables. The findings suggest that the feature-group-aware gated fusion approach (FF-ANN) is a promising framework for analyzing heterogeneous cardiovascular clinical data to support cardiovascular decisions.

Categories: AI/ML-based decision support systems, Machine Learning (ML), Deep Learning

Keywords: prediction of cardiovascular diseases, learning on the basis of features, feature fusion, catboost, clinical decision support system, deep learning

Introduction

Cardiovascular diseases (CVDs) are among the leading causes of death worldwide and continue to have a significant impact on healthcare systems. The World Health Organization estimates that CVDs cause nearly 20 million deaths every year [1], with a significant number of these deaths believed to be preventable through proper diagnosis, risk categorization, and early medical treatment. Timely diagnosis of high-risk patients is therefore crucial in minimizing disease progression, the occurrence of severe heart attacks, and long-term patient outcomes. The extensive use of electronic health records and regular improvements in diagnostic processes have provided significant potential for computational solutions to support decision-making during the early stages of cardiovascular disease [2,3]. The application of machine learning in CVD prediction has been growing due to its ability to capture nonlinear relationships among structured clinical variables [2,4]. Publicly available benchmark datasets, especially the heart disease dataset proposed by Janosi et al. [5] in the UCI Machine Learning Repository and its extended versions available in Kaggle repositories [6], have been significant in speeding up predictive model development. Studies on early cardiovascular prediction in Bangladesh have used machine learning algorithms such as logistic regression, decision trees, and support

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

vector machines, and have shown that automated cardiovascular risk prediction is feasible and of clinical interest [7]. Despite the interpretable limits of decision made using these methods, they tended to require manual feature engineering and were low in their ability to model higher-order interactions between features.

This was later improved by the introduction of ensemble learning methods and the use of random forests, gradient boosting, and hybrid boosting-based approaches [8]. These methods have consistently shown strong performance on tabular cardiovascular data due to their ability to capture nonlinear feature associations. A hybrid random forest-linear model achieved 88.7% classification accuracy [9], while Biswas et al. [10] demonstrated that feature selection combined with random forests achieved 94.51% prediction accuracy in the early stages of heart disease detection. CatBoost has recently gained attention because of its intrinsic support for categorical variables and its consistent performance in structured medical prediction tasks [11]. The use of deep learning techniques has also been investigated due to their capability to automatically learn feature representations from clinical data. Multilayer perceptron architectures and related feedforward neural networks have demonstrated encouraging outcomes in cardiovascular diagnosis. Deep neural networks for coronary heart disease prediction achieved 83.7% accuracy and an area under the receiver operating characteristic curve (AUC) of 0.89 [12]. Subsequently, hybrid deep-learning approaches have also achieved more than 90% accuracy using optimized feature selection and regularization schemes [13,14]. Nevertheless, comparative studies have shown that traditional feedforward neural networks can usually match, or even be slightly worse than, the state-of-the-art ensemble algorithms when used on heterogeneous tabular clinical data [15-17]. The heterogeneous nature of the cardiovascular clinical datasets is one of the leading causes of such limitation. These datasets are usually a combination of continuous physiological data, demographics, binary risk factors, and categorical diagnostic data. Such groups of features vary significantly in scale, sparsity, and predictive power. Standard artificial neural networks can tend to directly concatenate all variables, which can destabilize representation learning since feature groups known to be clinically different are coerced into a common learning pathway [10,17]. Despite the success of feature-conscious learning systems in the multimodal learning context, little has been done to apply them to structured clinical tabular data. The gated fusion algorithms, which were initially developed to combine multimodal information integration, is an effective way of dynamically controlling feature contributions [18].

Nevertheless, previous studies on cardiovascular prediction have not clearly separated numerical and categorical clinical characteristics prior to neural fusion. Moreover, many studies in the literature are based on benchmark datasets and simplified validation methods, potentially lowering the clinical realism and external generalizability [19,20]. To overcome these drawbacks, our proposed work introduces a Feature-Fusion Artificial Neural Network (FF-ANN) that explicitly isolates the numerical and categorical clinical variables into separate learning streams and synthesizes them via an adaptive gated fusion process. In contrast to benchmark studies, the proposed framework is tested on the basis of actual clinical data collected from Ayush Multispecialty Hospital and Research Centre (AMSHRC), Vijayapura, Karnataka, India, in real time. In addition, CatBoost is used as a comparative baseline along with feature-group analysis under a leakage-free stratified evaluation protocol. The main contributions of this paper are threefold: First, numerical-categorical separation of features to predict CVDs; second, gated fusion to control features-group contribution; and third, comparative validation of feature-group analysis with CatBoost to enhance the predictive strength of real-time clinical data and decrease the phenomenon of disease misclassification. The experimented results show that the proposed method is robust for the early detection of heart disease. Hence, the proposed fusion-based deep learning model can assist experts in clinical decision-making, improve patient outcomes, and reduce mortality.

Materials And Methods

Problem formulation

Assume that the clinical data is represented as

$$\mathcal{D} = (\mathbf{x}_i, y_i)_{i=1}^N$$

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. *Cureus J Comput Sci* 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

N denotes the overall size of the patient records, $\mathbf{x}_i \in \mathbf{R}^d$ is the feature vector of the i -th patient in a d -dimensional real-valued feature space, and y_i is the binary disease label, where 0 represents no CVD and 1 represents CVD. In this case, $\mathbf{R}^d$ denotes the space of all d -dimensional real-valued vectors, where d is the sum of the number of clinical input features. In order to explicitly model heterogeneity of features, every feature vector is decomposed as

$$\mathbf{x}_i = (\mathbf{x}_i^{(n)}, \mathbf{x}_i^{(c)})$$

where $\mathbf{x}_i^{(n)}$ represents numerical clinical measurements and $\mathbf{x}_i^{(c)}$ represents categorical clinical indicators.

To estimate the learning objectives:

$$\hat{y}_i = f(\mathbf{x}_i)$$

where $\hat{y}_i \in (0, 1)$ represents the probability of CVD, which is estimated. This problem formulation is implemented in the following methodology.

Methodology

The proposed FF-ANN method works in the following steps, as shown in Figure 1. The dataset comprised about 2,000 anonymized patient records from the AMSHRC dataset. The records included physiological numerical measurements and categorical diagnostic indicators for the classification of CVD. The proposed methodology follows preprocessing, exploratory data analysis, feature fusion, ANN and CatBoost learning, stratified cross-validation, evaluation metrics, and final disease prediction.

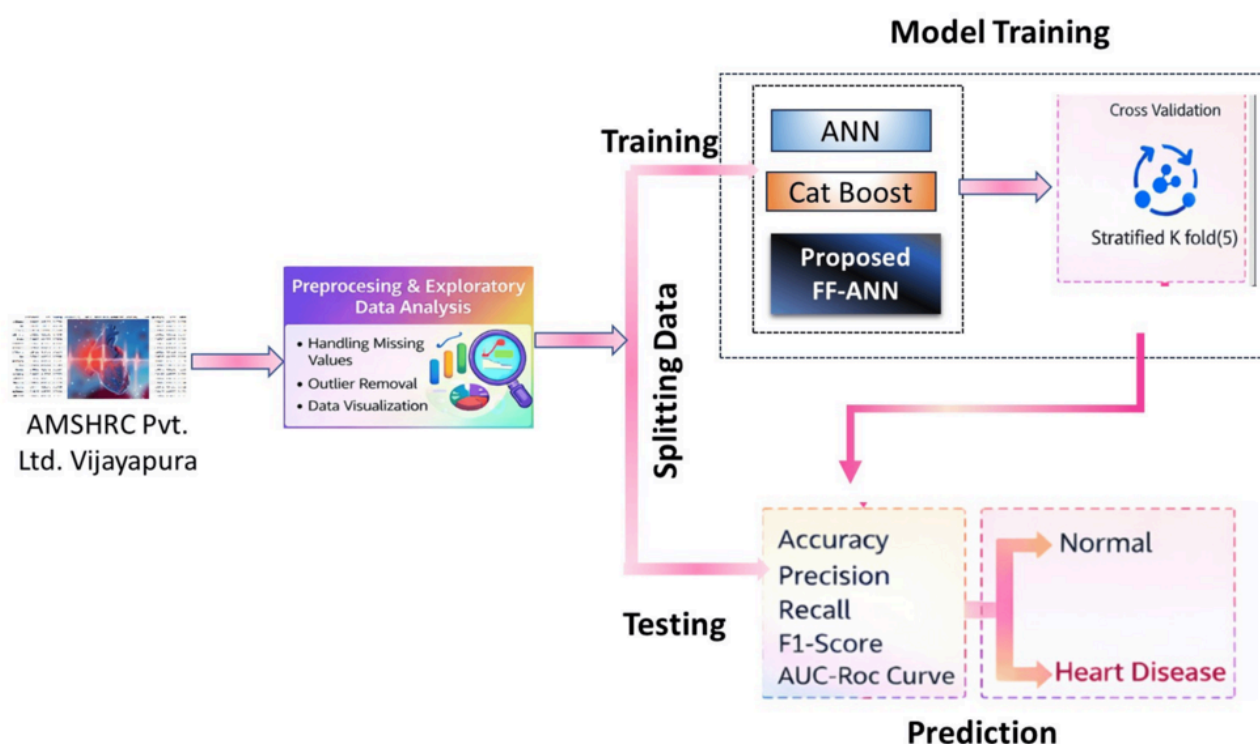


FIGURE 1: Proposed methodology

AMSHRC, Ayush Multispecialty Hospital and Research Centre; ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network; AUC-ROC, Area Under the Receiver Operating Characteristic

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

Dataset

The data employed in this research consists of real-time clinical cardiovascular records from AMSHRC. The dataset includes about 2,000 patient records collected during routine clinical assessment and diagnostic evaluation. Each record represents a single patient and comprises demographic data, physiological data, comorbidities, lifestyle factors, and diagnostic test results related to CVD diagnosis. The target variable is discrete and indicates the presence or absence of CVD based on clinical diagnosis. The analysis of patient records was performed anonymously, and no personally identifiable data were stored. The research was based on retrospective clinical records conducted for computational research under institutional data handling approval.

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

Feature Group	Feature	Description	Type
Numerical Features	Age	Patient age measured in years (20–95 years).	Continuous
	BP_systolic	Systolic blood pressure is a reflection of the cardiac contraction of the arteries (normal: 90-140 mmHg).	Continuous
	BP_diastolic	Diastolic blood pressure is a measure of arterial pressure at a particular time at the time of cardiac relaxation.	Continuous
	PR	Pulse rate measured in beats per minute (normal range: 60–100 bpm).	Continuous
	SpO ₂	Peripheral oxygen saturation percentage (normal range: 95–100%).	Continuous
	Weight	Body weight measured in kilograms.	Continuous
Categorical Features	Sex	Biological sex of the patient.	1 = Male, 0 = Female
	CP	Chest pain severity level.	0 = No pain, 1 = Mild pain, 2 = Moderate pain, 3 = Severe pain
	Chol	Serum cholesterol status.	1 = Elevated cholesterol, 0 = Normal
	Diabetics	Diabetes status.	1 = Diabetic, 0 = Non-diabetic
	Hypertension	Hypertension status.	1 = Hypertensive, 0 = Non-hypertensive
	Habits	Lifestyle-related behavioral risk factors.	0 = Healthy lifestyle, 1 = Smoker/Tobacco/Alcohol/Unhealthy habits
	ECG	Resting electrocardiography result.	0 = Not performed, 1 = Normal, 2 = Abnormal
	ECHO	Echocardiography result.	0 = Not performed, 1 = Normal, 2 = Abnormal

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

	TMT	Treadmill test outcome.	0 = Not performed, 1 = Negative, 2 = Positive
	CAG	Coronary angiography finding.	0 = Not performed, 1 = Normal, 2 = Single-vessel disease, 3 = Double-vessel disease, 4 = Triple-vessel disease
Target Variable	Target	Presence or absence of cardiovascular disease.	1 = Disease present, 0 = No disease

TABLE 1: Cardiovascular disease prediction using clinical feature description and grouping

To facilitate feature-group-conscious learning, the variables were divided into numerical and categorical feature groups based on clinical characteristics and data representation needs, as shown in Table 7. The numerical features are age, systolic blood pressure, diastolic blood pressure, pulse rate, oxygen saturation, and body mass. These variables are continuous physiological measurements and provide information on cardiovascular condition. Age ranges between 20 and 95 years, and blood pressure indicators represent arterial pressure during systolic and diastolic heart contractions. Pulse rate is recorded in beats per minute, oxygen saturation represents peripheral oxygen concentration, and body weight is measured in kilograms. Calcified characteristics are sex, type of chest pain, the presence of cholesterol, diabetes, high blood pressure, lifestyle, electrocardiogram (ECG), echocardiography (ECHO), treadmill test (TMT), and coronary angiography (CAG). These variables capture clinical conditions, diagnosis results, and risk variables. Chest pain has four levels of intensity, ranging from no pain to severe pain. A binary clinical indicator is presented as cholesterol, diabetes, and hypertension. Behavioral risk factors encompassed in lifestyle habits include smoking, tobacco use, alcohol use, and other factors that are not healthy. The diagnostic variables are ECG, ECHO, and TMT results, recorded as normal, abnormal, or unavailable. CAG records the severity of vessel involvement such as normal coronary to single, double, or triple vessels. Preliminary analysis of data revealed an intermediate level of class imbalance, whereby the cases of CVD were lower than those of non-disease cases. The ultimate dataset distribution was 779 disease-positive samples and 1,221 normal samples that equate to around 39% and 61% positive and negative samples, respectively. This distribution of classes encouraged the application of the class-weighted learning and threshold optimization when training the model. The non-homogenous composition of the dataset itself inspired the suggested feature-group-sensitive learning approach, according to which numerical and nominal variables are handled with the help of separate neural streams and then merged via gates. The dataset properties and the range of the numerical features are the same as were described in our previous study [21] as the same clinical dataset was used in the current research.

Case	Training (80%)	Testing (20%)
Heart Disease	623	156
Normal	977	244
Total	1,600	400
Total Samples	2,000	

TABLE 2: Total, training, and testing samples

Table 2 illustrates that the total sample size, population of the patient with heart diseases, training cases, and test cases.

Exploratory data analysis

Exploratory data analysis was also used to discuss the structure, distribution, and relationships in the cardiovascular data before model development. The data consisted of about 2,000 patients records containing numerical values (e.g., age, blood pressure, pulse rate, oxygen saturation, and weight) and nominal attributes (e.g., the type of chest pain, comorbidities, and test outcomes of diagnostic tests such as ECG, ECHO, TMT, and CAG). The dependent variable was nominal in nature in that it is the presence or absence of CVD. The preliminary examination showed that there was a moderate class imbalance whereby there were more disease-positive cases and thus the adoption of class-weighted learning and powerful evaluation measures, including AUC and Matthews correlation coefficient (MCC) [10,12]. The numerical features showed varying skewness and outliers, as indicated in Figure 2, while categorical variables showed distribution across disease classes, as presented in Figure 3. The feature-target analysis also revealed that the results of diagnostic tests were more likely to discriminate than the demographic variables, and the correlation trends are presented in Figure 4.

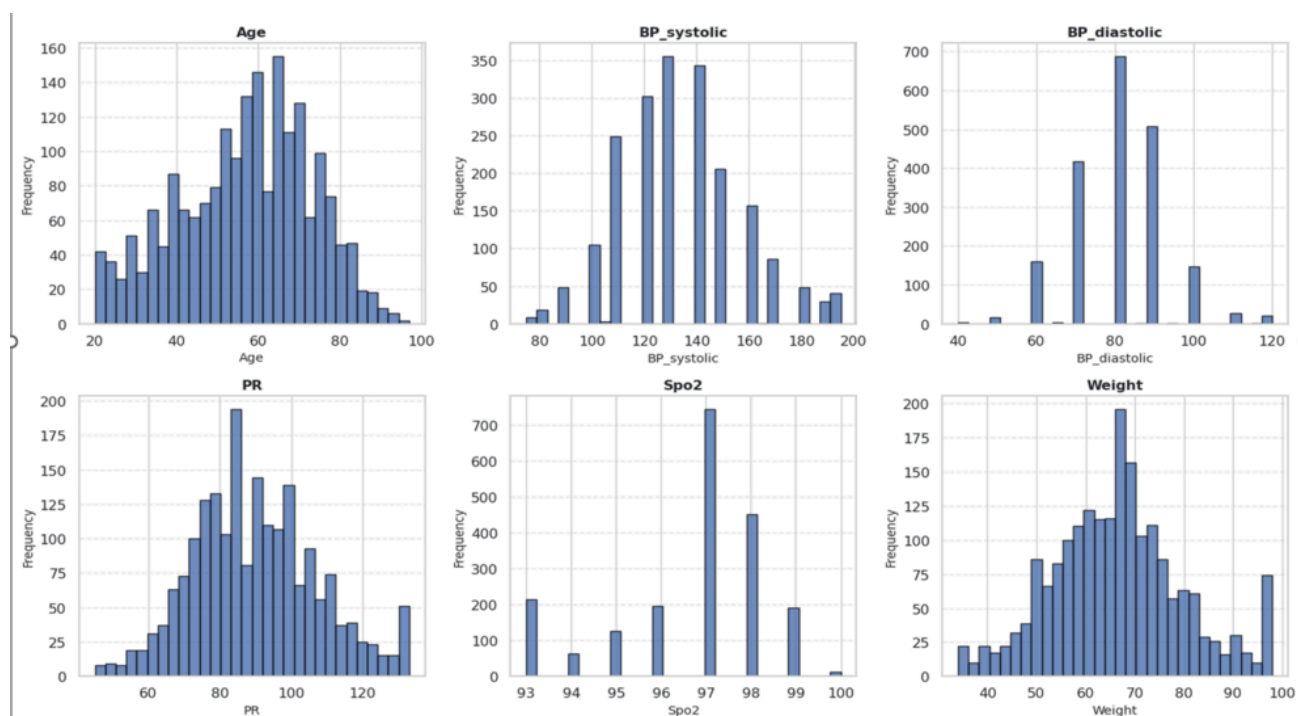


FIGURE 2: Distributions of selected numerical features using histograms

Figure 3 shows the distributions of the chosen numerical characteristics with the help of the histograms. The histogram demonstrates the distribution of the major continuous clinical variables. The age, pulse rate, and weight have a rough normal distribution and moderate variance, whereas diastolic blood pressure has slight skewness but within the expected ranges. Oxygen saturation (SpO_2) is tightly clustered around normal values with little variance. Cholesterol on the other hand does not follow a normal distribution as it has pronounced outliers, implying that it is more varied than other variables.

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. *Cureus J Comput Sci* 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

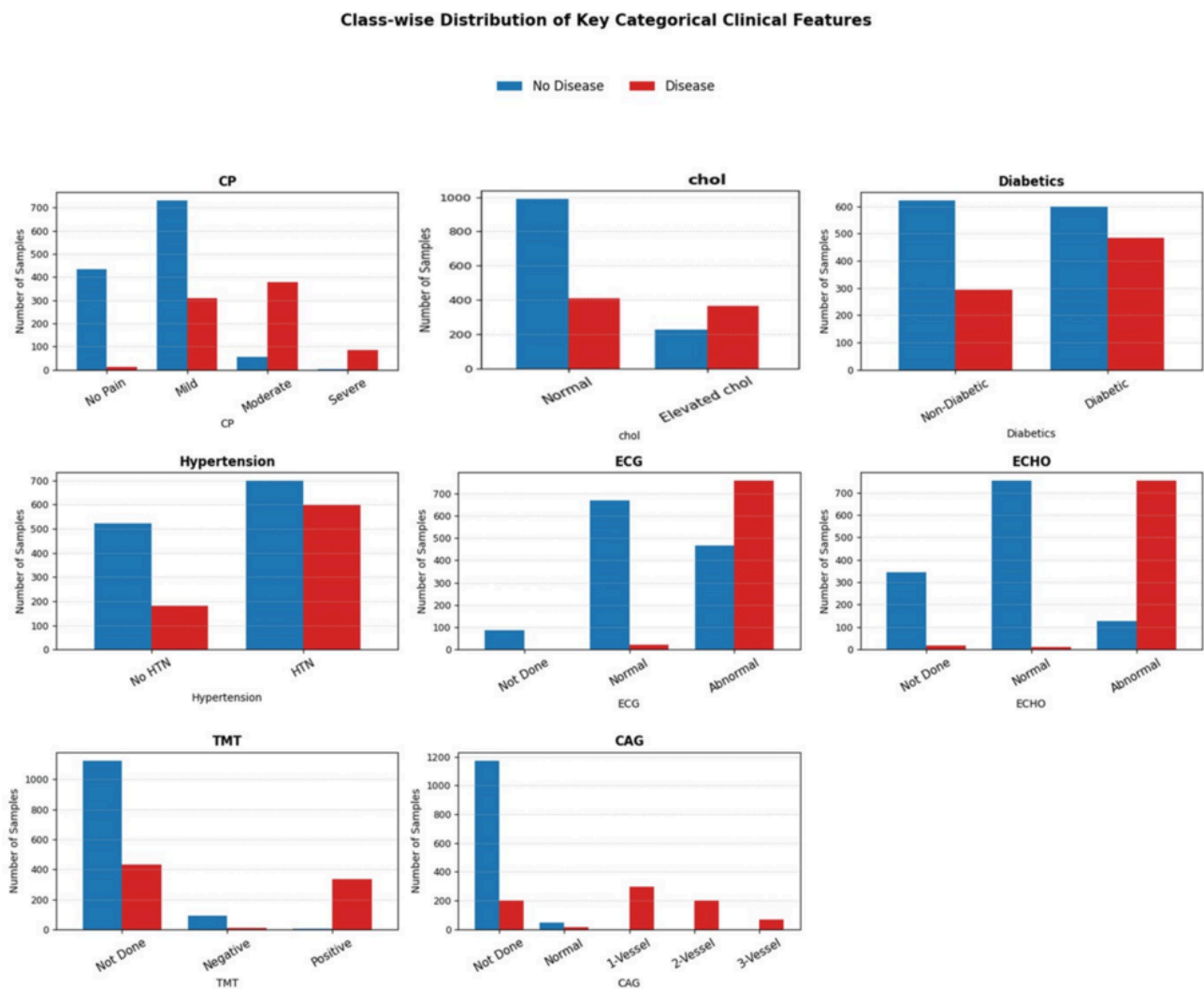


FIGURE 3: Class-wise distribution of key categorical distribution

Figure 4 indicates the proportion of the main categorical clinical characteristics between diseased and healthy patients. The diagnostic variables including ECG, ECHO, TMT, and CAG show clear separation, with the higher category values being more often linked to the disease and thus are strong predictors. There are also differences in types of chest pains in the two groups, with diabetes and hypertension showing significant overlap and a weaker association.

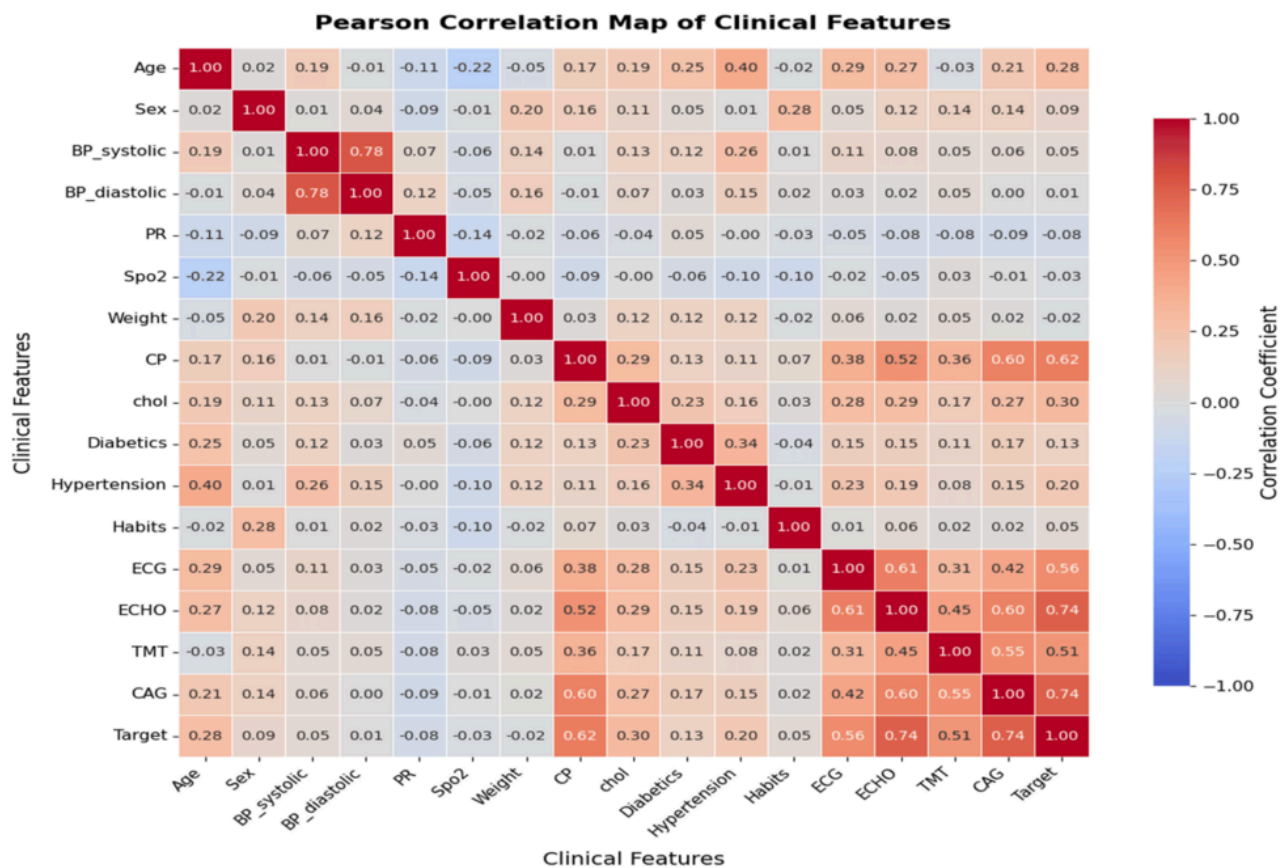


FIGURE 4: Correlation heatmap

The Pearson correlation heatmap, as shown in Figure 4, displays interrelations of clinical characteristics and how they are correlated with CVD. The diagnostic variables, namely, ECHO, CAG, and TMT, are positively correlated with the target at high levels, and as such, they are more predictive. There are moderate relationships between risk factors such as hypertension, diabetes, and cholesterol. On the other hand, demographic and physiological factors, including age, blood pressure, pulse rate, and SpO₂, show lower correlation with the target, which indicates a low individual contribution. The positive relationship between systolic and diastolic blood pressure is an indication of the consistency of measurements, and the low values of multicollinearity of most features confirm the information diversity. On the whole, the findings indicate that diagnostic features have a higher discriminative capacity than simple clinical measurements, which contributes to the employment of feature-group-sensitive modeling techniques.

The exploratory data analysis demonstrates high levels of heterogeneity of features, average levels of class imbalance, and diverse predictive relevance of features. These results were the influence that directly led to the use of feature-group-aware preprocessing, neural architecture asymmetry, and gated fusion in the designed structure.

Data preprocessing

The dataset with CVDs has heterogeneous clinical attributes of different types of data, scale, and quality. To be able to guarantee model training and evaluate objectively, preprocessing pipeline was conducted in a systematic form before model development. First, non-informative attributes were eliminated and included serial numbers, unnamed indexing columns, and empty columns. Columns with invalid characters only were dropped, and duplicates of the patients were removed to avoid biased learning. All numerical characteristics were directly translated to numerical values in order to

provide consistency in the computations. Outliers in continuous clinical measurements were addressed using a clipping strategy, which was based on interquartile range (IQR) [4]. Values of each of the numerical features x were restricted to the interval.

$$\text{IQR} = Q_3 - Q_1, x \in [Q_1 - 1.5 \cdot \text{IQR}, Q_3 + 1.5 \cdot \text{IQR}]$$

Q_1 and Q_3 are the first and third quartiles, respectively. All samples are retained in this method, but the effect of the extreme values is minimized, which is desirable in the case of clinical data. In order to avoid leakage of information, preprocessing was done on each of the training folds. To impute numerical features, the median was used, and to standardize it, the training-set statistics were used:

$$\tilde{x}^{(n)} = \frac{x^{(n)} - \mu}{\sigma}$$

where μ and σ were calculated only using the training data.

The most common category was imputed as categorical features and coded through one-hot encoding:

$$x^{(c)} \rightarrow z^{(c)} \in [0.1]^{d'_c}$$

where d'_c represents the dimensionality of the encoded categorical feature space. The final processed and clean dataset was the same one used in every experiment, ensuring fair comparability and reproducibility of the results.

Machine learning and deep learning models

The baseline ANN and CatBoost models were added to allow the comparative assessment of the proposed FF-ANN. The pipeline will start by preprocessing and analyzing exploratory data, such as addressing missing values, removing outliers, and visualizing data. The trained features are then applied to model training, in which the proposed feature-fusion mechanism is applied in the FF-ANN. The performance of the model is measured with stratified 5-fold cross-validation, and the predictions are measured using various evaluation measures. The resultant final outputs are the predicted probabilities in binary class labels, which are normal or heart disease.

Standard ANN

A baseline model adopted to predict CVD is the standard ANN. All features were combined into a single feature vector and analyzed in a homogeneous manner without explicitly separating feature groups. The network has two fully connected hidden layers with 64 and 32 neurons, respectively. After each hidden layer, batch normalization, ReLU activation, and dropout regularization are applied with a dropout rate of 0.20.

The hidden layer of ANN transformations is defined as

$$\mathbf{h} = \text{Dropout}(\text{ReLU}(\text{BN}(\mathbf{W}\mathbf{x} + \mathbf{b}_h)))$$

where h denotes the hidden-layer output representation, $\mathbf{W}$ denotes the trainable weight matrix, b_h denotes hidden-layer bias, $\text{BN}(\cdot)$ denotes batch normalization, ReLU denotes rectified linear activation, and Dropout denotes regularization through random neuron deactivation during training.

ReLU activation is defined as

$$\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x})$$

where x denotes the pre-activation input to a neuron.

The last output layer approximates the presence of CVD as the result of a sigmoid activation:

$$\hat{y} = \sigma(\mathbf{w}^\top \mathbf{h}_2 + b)$$

where $\hat{y}$ denotes the probability of the CVD being predicted, $\sigma(\cdot)$ denotes sigmoid activation, w denotes the weight vector in the output layer, $\mathbf{w}^\top$ denotes the transpose of the weight vector, and b identifies the output bias.

How to cite this article:

The AdamW optimizer was used with a learning rate of 3×10^{-4} and a decay weight of 1×10^{-4} , a batch size of 16, and 100 epochs as the maximum number of training steps. Binary cross-entropy with label smoothing was used as the objective function, and early stopping was used to avoid overfitting and enhance generalization performance. This is a conventional neural network design with a feedforward architecture that has traditionally been used in cardiovascular prediction studies. The loss function was binary cross-entropy, and label smoothing was set to 0.05, with a batch size of 16 and early stopping with a patience of 10 epochs and a maximum of 100 epochs. This default architecture is based on a conventional feedforward neural network design typically employed in studies of cardiovascular prediction. The architecture of the standard ANN is shown in Figure 5.

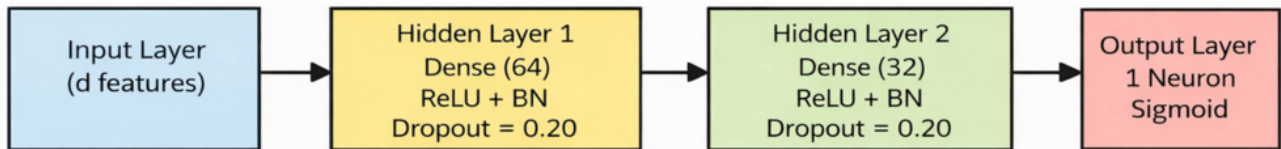


FIGURE 5: Standard artificial neural network architecture

ReLU, Rectified Linear Unit; BN, Batch Normalization

Proposed FF-ANN Method

The proposed FF-ANN clearly divides both numerical and categorical clinical features into two dedicated lines of learning to obtain feature-group-specific representations and then adaptively fuses them. The adaptive gating mechanism is based on the overall principle of gated multimodal units introduced in [18], modified in this case for clinical tabular data. The numerical branch comprises a dense layer consisting of 64 neurons with a dropout rate of 0.10.

$$\mathbf{x} = [\mathbf{x}^{(n)}, \mathbf{x}^{(c)}]$$

where $\mathbf{x}$ represents the entire pre-processed patient feature set, $\mathbf{x}^{(n)}$ represents the numerical feature subset, and $\mathbf{x}^{(c)}$ represents the categorical feature subset. The original independent generation of the branch-specific hidden representations was based on dense layers dedicated to every group of features. Each of the branches was then regulated with sigmoid-based gating:

$$\tilde{\mathbf{h}}^{(k)} = \mathbf{h}^{(k)} \odot \sigma(\mathbf{W}g^{(k)}\mathbf{h}^{(k)} + \mathbf{b}g^{(k)})$$

where $k \in (n, c)$ represents branch index, with n denoting the numerical branch and c denoting the categorical branch $\mathbf{h}^{(k)}$ masked by branch-specific hidden representation, gated by gate branch representation; $\mathbf{W}g^{(k)}$ denotes gating weight matrix, $\mathbf{b}g^{(k)}$ represents gating bias, $\sigma(\cdot)$ is the gating function, and $\odot$ represents element-wise multiplication.

The concatenated gated outputs were used to obtain the fused feature representation:

$$\mathbf{z} = [\tilde{\mathbf{h}}^{(n)}, \tilde{\mathbf{h}}^{(c)}]$$

where $\mathbf{z}$ indicates the combined hidden representation.

The dense fusion layer uses sigmoid activation to obtain the final disease probability:

$$\hat{y} = \sigma(\mathbf{w}^T \mathbf{z} + \mathbf{bo})$$

Here, output-layer weights are represented by $\mathbf{w}$ and output bias by $\mathbf{bo}$.

The hyperparameters of AdamW training were kept the same as those of the baseline ANN. Minimization of weighted binary cross-entropy was used, and the final classification threshold was chosen among candidate thresholds from 0.1 to 0.9 by maximizing the validation F1-score.

The overall FF-ANN algorithm is described in Algorithm 1, and the architecture is presented in Figure 6.

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

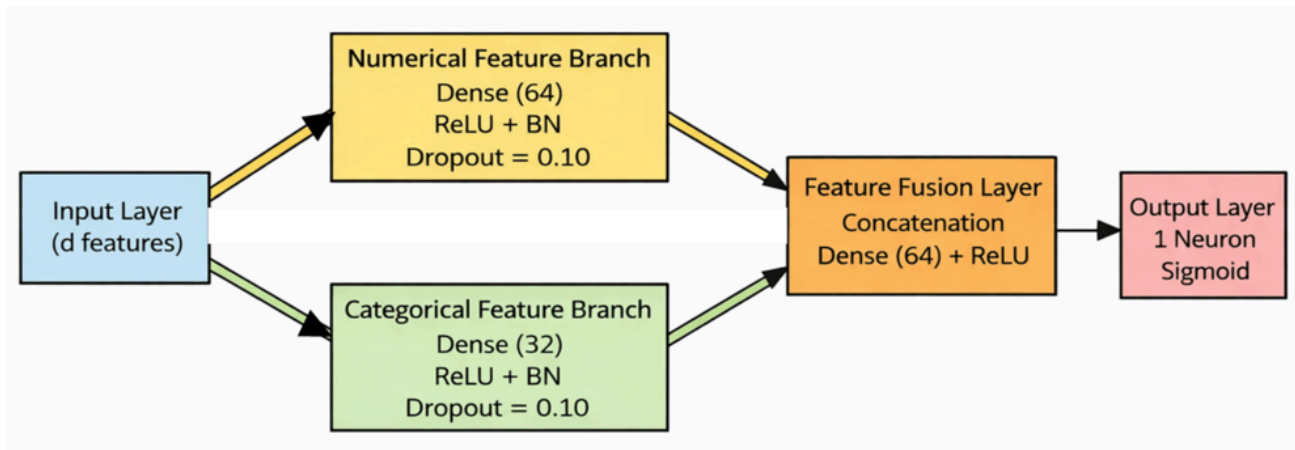


FIGURE 6: Architecture of the proposed FF-ANN

FF-ANN, Feature-Fusion Artificial Neural Network; ReLU, Rectified Linear Unit; BN, Batch Normalization

Algorithm for the proposed FF-ANN:

Input: Clinical dataset

$$\mathcal{D} = (\mathbf{x}_i, y_i)_{i=1}^N$$

Output: Probability of cardiovascular diseases prediction.

$$\hat{y}_i \in (0, 1)$$

Step 1. Load the AMSHRC dataset.

Step 2. The dataset is processed under a leakage-free protocol, which involves missing values being filled in on a case-by-case basis, and scaling and encoding is performed.

Step 3. The records of patients are categorized into groups of numerical and categorical features:

$$\mathbf{x}_i = (\mathbf{x}_i^{(n)}, \mathbf{x}_i^{(c)})$$

Step 4. Latent representations are learned by distinct neural branches in both groups of features:

$$\mathbf{h}^{(n)} = f_n(\mathbf{x}^{(n)}), \mathbf{h}^{(c)} = f_c(\mathbf{x}^{(c)})$$

Step 5. The sigmoid-based gating is employed to normalize the contribution before the fusion:

$$\mathbf{z} = \mathbf{h}^{(n)} \odot \mathbf{g}^{(n)}, \mathbf{h}^{(c)} \odot \mathbf{g}^{(c)}$$

Step 6. To evaluate the dense probability, the fused representation is accelerated via dense layers :

$$\hat{y}_i = \sigma(\mathbf{w}^T \mathbf{z} + \mathbf{b})$$

Step 7. Optimization of model parameters is done by mini-batch learning, which is regularized and stops early.

Step 8. Selection of Thresholds: Select the best threshold.

Step 9. Assign final class labels.

$\mathcal{D}$ represents the dataset with N samples; $\mathbf{x}_i$ represents the input feature vector of the i -th patient, where $\mathbf{x}_i^{(n)}$ and $\mathbf{x}_i^{(c)}$ represent a numerical subset of features, respectively. $f_n(\cdot)$ and $f_c(\cdot)$ denote the corresponding neural transformation, $\mathbf{h}^{(n)}$ and $\mathbf{h}^{(c)}$ represent a hidden representation, $\mathbf{g}^{(n)}$ and $\mathbf{g}^{(c)}$ represent gating vectors, $\odot$ denotes element-wise multiplication,

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

and z represents the fused feature. w and b represent the weights and bias of the output layer, respectively, and $\sigma(\cdot)$ denotes the sigmoid activation, and $\hat{y}_i$ denotes the predicted probability.

CatBoost-Based Feature-Group Analysis

The use of CatBoost classifier [11] was based on its effectiveness when working with heterogeneous tabular data, as well as categorical variables. This model prediction is given as

$$\hat{y}_i = \sum_{m=1}^M f_m(\mathbf{x}_i)$$

where $f_m(\mathbf{x}_i)$ represents the decision tree in the m^{th} iteration, $\mathbf{x}_i$ represents the input feature vector, M represents the number of iterations of boosting, and $\hat{y}_i$ represents the predicted output for the i^{th} patient sample.

The training of this model was done with 600 trees, depth = 5, and learning rate = 0.05, where the objective function and evaluation metric were Logloss and AUC. Early stopping was used after 50 iterations, and balanced weights of classes were used because it improved the generalization. There were three feature settings considered, namely numerical, categorical, and a combination of all features. Stratified five-fold cross-validation and independent hold-out test set were used to determine performance.

Models performance evaluation

Stratified five-fold cross-validation followed by an 80:20 hold-out test split is used to test all the models. The performance is reported with the help of accuracy, precision, recall, F1-score, AUC, and MCC:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1-Score} = 2 \bullet \frac{\text{Precision} \bullet \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

where TP, TN, FP, and FN represent the True Positive, True Negative, False Positive, and False Negative values, respectively. These metrics are widely recommended for imbalanced clinical classification tasks because they jointly reflect discrimination, balance, and classification stability [10,12].

The hyperparameters were optimized in each fold by maximizing validation F1-score. In the case of neural networks, validation probabilities $\hat{y}_i \in [0, 1]$ were further optimized, as opposed to setting a default classification threshold of 0.5. All the model configurations were based on what was implemented so that they could be reproduced.

Results And Discussion

Experimental setup

The CVD prediction model was coded in Python using Google Colab with scikit-learn, PyTorch, and CatBoost. Anonymized clinical records of approximately 2,000 cases from AMSHRC were analyzed, including numerical (e.g., age, blood pressure, pulse rate, oxygen saturation, and weight) and categorical (e.g., chest pain type, diabetes, hypertension, ECG, ECHO, TMT, and CAG) features. All personally identifiable information was removed.

The 80:20 stratified train-test split was used to divide the dataset, and stratified five-fold cross-validation was used in the course of model development to maintain the class distribution. Numerical features were median-imputed and standardized, while categorical features were imputed using the most frequent category and one-hot encoded.

How to cite this article:

Three models were compared: ANN, FF-ANN, and CatBoost. ANN and FF-ANN were trained using AdamW optimizer (learning rate = 3×10^{-4} , weight decay 1×10^{-4} , batch size 16, maximum 100 epochs, and early stopping), with binary cross-entropy loss and validation-directed optimization of the F1-score threshold. Logloss and AUC were used to train CatBoost, with 600 trees, depth 5, and learning rate of 0.05 on balanced weights of the classes, with stopping early [11]. Accuracy, precision, recall, F1-score, AUC, and MCC were used to measure performance, which are suggested to be used when the clinical dataset is unbalanced.

Model performance

Stratified five-fold cross-validation and independent hold-out test data were used as the performance evaluation tools to determine the robustness of the classification and generalization of the proposed models. It was compared to the standard ANN, the proposed FF-ANN, and CatBoost on several performance metrics that included accuracy, precision, recall, F1-score, MCC, and AUC.

Performance of the Proposed FF-ANN

The proposed FF-ANN demonstrated the best-balanced performance among the analyzed neural models, as summarized in Table 3. Compared to the standard ANN, the FF-ANN improved test accuracy (95.50% to 96.75%), recall (0.9038 to 0.9359), F1-score (0.9400 to 0.9574), and MCC (0.9059 to 0.9318). These improvements reflect a reduction in FN from 15 to 10 cases, while FP remained unchanged at 3 cases, indicating increased sensitivity for disease-positive patients without reduced specificity. The proposed FF-ANN achieved high test accuracy, recall, F1-score, and MCC compared to both the standard ANN and CatBoost.

Model	Accuracy	Precision	Recall	F1-score	MCC	AUC
Standard ANN	95.5	0.9792	0.9038	0.94	0.9059	0.9929
CatBoost	95.75	0.9793	0.9103	0.9435	0.9111	0.9934
Proposed FF-ANN	96.75	0.9799	0.9359	0.9574	0.9318	0.996

TABLE 3: Comparative performance of standard ANN, feature-group ANN, and CatBoost models

ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network; MCC, Matthews Correlation Coefficient; AUC, Area Under the Receiver Operating Characteristic curve

Classification Report Comparison

Comparing with the standard ANN, in the Normal class, the precision, recall, and F1-score values were 0.9414, 0.9877, and 0.9640, respectively. In the Disease class, the precision, recall, and F1-score values were 0.9792, 0.9038, and 0.9400 respectively, which gave an overall accuracy of 95.50%. CatBoost model was also able to provide good classification performance. In the case of the Normal class, precision, recall, and F1-score were 0.9451, 0.9877, and 0.9659 respectively, whereas in the case of the Disease, the precision was 0.9793, the recall was 0.9103, and the F1-score was 0.9435, with an overall accuracy of 95.75%. The proposed method was shown to be more effective with respect to the two classes. In the case of the Normal class, the precision rose to 0.9602 and the recall stood at 0.9877, which generated an F1-score of 0.9737. In the Disease class, the recall increased to 0.9359, with the precision of 0.9799, which provided the F1-score of

How to cite this article:

0.9574 and the accuracy of 96.75%. The proposed method offered the most moderate classification performance, especially in terms of better disease recall as well as high level of normal case discrimination. Table 4 shows a comparison of classification reports for the standard ANN, CatBoost, and the proposed model.

Model	Class	Precision	Recall	F1-score	Support
Standard ANN	Normal(0)	0.9414	0.9877	0.9640	244
	Disease(1)	0.9792	0.9038	0.9400	156
	Accuracy	0.9550	0.9550	–	400
	Macro Avg	0.9603	0.9458	0.9520	400
	Weighted Avg	0.9561	0.9550	0.9546	400
FF- ANN	Normal(0)	0.9602	0.9877	0.973	244
	Disease(1)	0.9799	0.9359	0.9574	156
	Accuracy	0.9675	0.9675	–	400
	Macro Avg	0.9700	0.9618	0.9656	400
	Weighted Avg	0.9678	0.9675	0.9674	400
CatBoost	Normal(0)	0.9451	0.9877	0.9659	244
	Disease(1)	0.9793	0.9103	0.9435	156
	Accuracy	0.9575	0.9575	–	400
	Macro Avg	0.9622	0.9490	0.9547	400
	Weighted Avg	0.9584	0.9575	0.9572	400

TABLE 4: Comparison of classification reports for the standard ANN, feature-group ANN, and CatBoost models

ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network

Confusion Matrix Analysis

The test set computation of confusion matrices provides detailed information on model-specific classification behavior and error distribution. The confusion matrices for the standard ANN, CatBoost, and proposed FF-ANN are shown in Figure 7. In the case of the standard ANN, 241 TN and 141 TP were correctly identified, along with 3 FP and 15 FN. The comparatively higher number of FN indicates reduced sensitivity to disease-positive cases. The proposed FF-ANN showed improved classification performance, correctly identifying 241 TN and 146 TP, with 3 FP and 10 FN. This reduction in FN directly led to better recall and overall classification balance. The CatBoost model correctly identified 241 TN and 142 TP,

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

with 3 FP and 14 FN. Although CatBoost showed competitive overall performance, the FF-ANN achieved the lowest FN rate among all evaluated models. Overall, the FF-ANN performed best in terms of confusion matrix outcomes, showing the least disease misclassification without compromising high specificity, particularly in CVD screening.

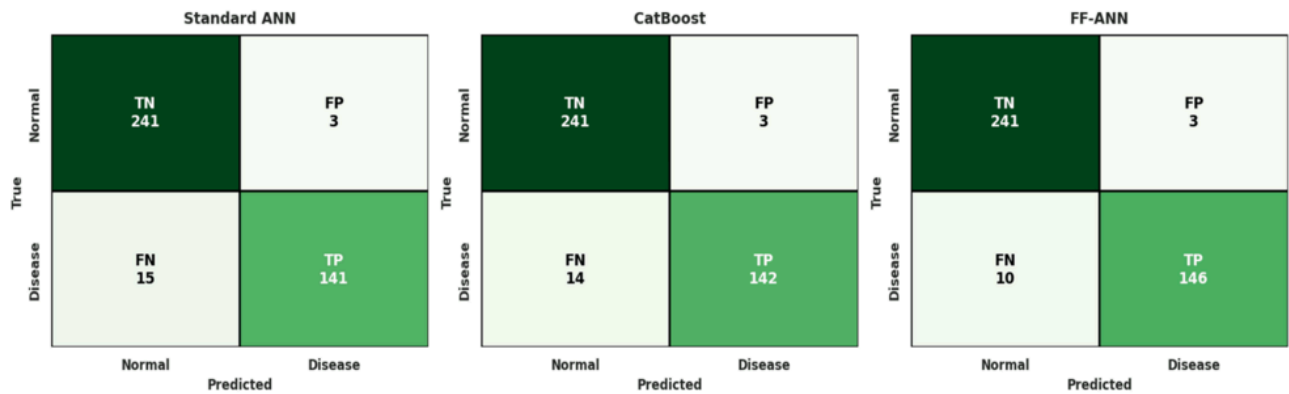


FIGURE 7: Confusion matrix of ANN, CatBoost, and FF-ANN models

ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network

ROC Curve Analysis

Figure 8 indicates high discrimination and convergence stability in the assessed neural models, as indicated by the ROC curves. Excellent classification performance is observed in the upper ROC plots, which have AUC of 0.992, 0.996, and 0.9934 corresponding to standard ANN, proposed FF-ANN, and CatBoost model, respectively, and this shows good separation between CVD and normal cases. The accuracy curves exhibiting a low accuracy are characterized by a fast rate of learning in the first training epochs, and thereafter a stabilized convergence rate. Standard ANN had an approximate test accuracy of 95.50%, and FF-ANN had an approximate test accuracy of 96.75%.

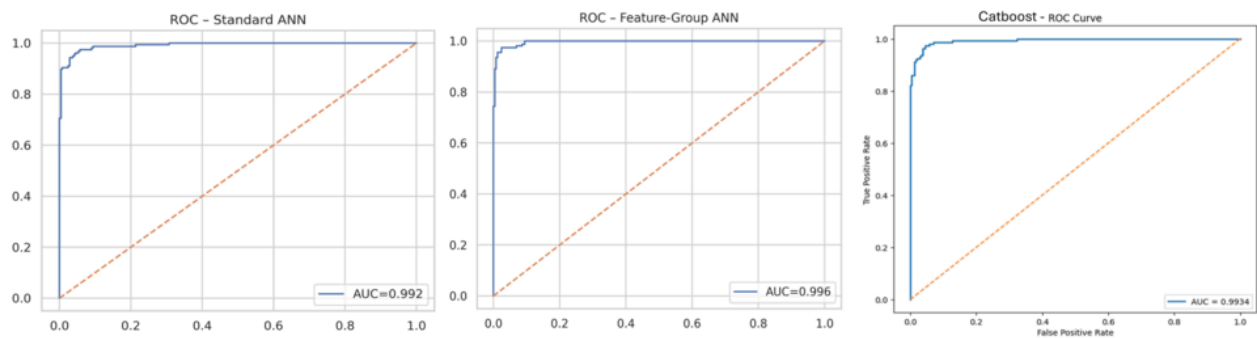


FIGURE 8: ROC curves of standard ANN, FF-ANN, and CatBoost models

ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network; AUC-ROC, Area Under the Receiver Operating Characteristic

Training and Validation Loss Analysis

The measurement of convergence behavior and generalization in the neural architectures was observed using the training and validation loss curves. Figure 9 shows the representative loss curves of the standard ANN, proposed FF-ANN, and CatBoost model. The standard ANN showed a steep decrease in loss during the initial epochs, followed by slow

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group–Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

stabilization after approximately 10 epochs. The difference between training and validation loss became small in the later epochs, indicating slight overfitting and the overall stable convergence. The proposed FF-ANN showed smooth optimization dynamics, where the training and validation loss decreased steadily and had a smaller generalization gap during training. This indicates greater regularization and improved feature learning stability compared to the standard ANN. The convergence of the CatBoost was also characterized by constant convergence with a little variation in validation loss. The close parallelism of the training and validation loss curves in any configuration may imply that regularization scheme was successful and overfitting was low.

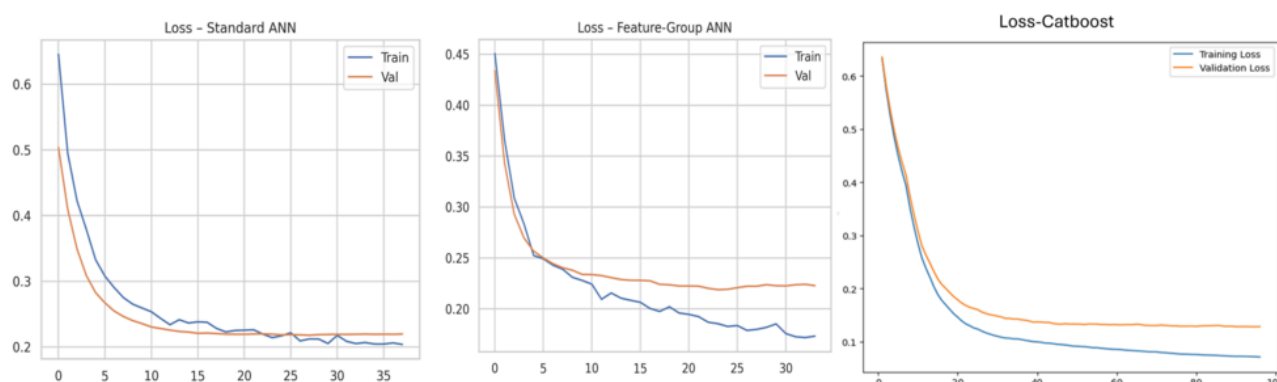


FIGURE 9: Training and validation loss curves of standard ANN, FF-ANN, and Catboost models

ANN, Artificial Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network

Overall, the findings show that, on the variable computations, the proposed FF-ANN is superior to the baseline models and provides more credible and clinically significant estimations to detect CVD.

Discussion

The findings of this study show that the applications of ANN and CatBoost demonstrated good performance in the prediction of CVD. However, the standard ANN processes features in a uniform way and does not consider heterogeneity, whereas CatBoost uses implicit processing of categorical variables [10]. This leaves a limitation in modeling heterogeneous clinical manifestations explicitly and in minimizing FN, which are of great importance in clinical diagnosis. To overcome this, the proposed FF-ANN splits both numerical and categorical features into separate branches and uses gated fusion based on a sigmoid function to control their contributions. This facilitates feature-group-based learning and better representation of organized clinical data.

Findings indicate that FF-ANN has higher accuracy, higher recall, higher F1-score, and higher MCC, and the number of false negatives (15 to 10 cases) is significantly lower, which suggests better sensitivity. The proposed model has a higher test-set AUC than CatBoost, indicating superior generalization. These results can be compared with previous research studies that have prove the usefulness of feature-conscious architectures in medical prediction tasks [4,17].

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group-Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. *Cureus J Comput Sci* 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

Authors	Dataset	Models Achieved High Accuracy	Accuracy Achieved
Nagavelli et al. [4]	PhysioNet	XGBoost Algorithm	95.9%
Mohan et al. [9]	UCI	Linear+ Random Forest	88.7%
Miao and Miao [12]	UCI	Enhanced DNN	83.67%
Rimal et al. [17]	Open-source dataset from Github	Random Forest	94%
Proposed FF-ANN	AMSHRC	FF-ANN	96.75%

TABLE 5: Comparison with existing work

AMSHRC, Ayush Multispecialty Hospital and Research Centre; DNN, Deep Neural Network; FF-ANN, Feature-Fusion Artificial Neural Network; UCI, University of California, Irvine Machine Learning Repository

Comparing with the existing models in Table 5, our proposed model achieved high accuracy. Hence, based on more influential features such as ECHO, ECG, TMT, and CAG, our model achieved a high AUC of 0.996. Nevertheless, along with these encouraging results, our study has a few limitations. To begin with, although the dataset includes about 2,000 patient records, larger multi-center cohorts can enhance model robustness and external validity. The proposed FF-ANN improves predictive performance but indicates the need for future inclusion of explainability methods such as SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME). Future work will focus on developing a hybrid model with nested cross-validation and the combination of other explainability techniques such as SHAP or LIME.

Overall, the findings show that the proposed FF-ANN enhances the balance of classification and minimizes misclassification during the prediction of CVDs.

Conclusions

The traditional machine learning and deep learning models are not able to process features uniformly in their predictions of CVD, and thus, the proposed FF-ANN explicitly separates numerical and categorical clinical features into separate learning branches and combines them via an adaptive gated fusion mechanism. An experimental assessment with a CVD dataset with about 2,000 patient records showed that the proposed FF-ANN is more effective in classification than a standard ANN. It was found that the proposed model demonstrated a test accuracy of 96.75% along with an MCC of 0.9318 and an AUC of 0.996, which is better than the baseline ANN, which showed a test accuracy of 95.50%, with an MCC of 0.9059 and an AUC of 0.992, indicating a better class-balanced prediction and less disease misclassification. Besides that, such results suggest that explicit proposed method provides better neural representation learning of heterogeneous clinical data and better CVD prediction performance in clinical decision support systems. Furthermore, this research is dedicated to creating a hybrid clinical decision support framework where machine learning and deep learning are combined with post hoc explainability algorithms like SHAP and LIME. The interpretation of such results is that architecture-sensitive feature-group learning would be an effective future direction in increasing the dependability of neural prediction of structured cardiovascular clinical data.

How to cite this article:

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sunanda Budihal, Sheetalrani R. Kawale , Nitin Agarwal, Bhagirathi Halalli

Acquisition, analysis, or interpretation of data: Sunanda Budihal, Sheetalrani R. Kawale , Bhagirathi Halalli

Drafting of the manuscript: Sunanda Budihal, Sheetalrani R. Kawale , Nitin Agarwal, Bhagirathi Halalli

Critical review of the manuscript for important intellectual content: Sunanda Budihal, Sheetalrani R. Kawale , Nitin Agarwal, Bhagirathi Halalli

Supervision: Sunanda Budihal, Sheetalrani R. Kawale

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Intellectual property info:** We have a plan to patent. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. World Health Organization. Cardiovascular diseases (CVDs). (2025). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29?utm_source=chatgpt.com.
2. Subramani S, Varshney N, Anand MV, et al.: Cardiovascular diseases prediction by machine learning incorporation with deep learning. *Frontiers in Medicine*. 2023, 10:1150933. [10.3389/fmed.2023.1150933](https://doi.org/10.3389/fmed.2023.1150933)
3. Zhou C, Dai P, Hou A, Zhang Z, Liu L, Li A, Wang F: A comprehensive review of deep learning-based models for heart disease prediction. *Artificial Intelligence Review*. 2024, 57:263. [10.1007/s10462-024-10899-9](https://doi.org/10.1007/s10462-024-10899-9)
4. Nagavelli U, Samanta S, Chakraborty P: Machine learning technology-based heart disease detection models. *Journal of Healthcare Engineering*. 2022, 2022:1-9. [10.1155/2022/7351061](https://doi.org/10.1155/2022/7351061)
5. Janosi A, Steinbrunn W, Pfisterer M, Detrano R. Heart Disease [Dataset]. UCI Machine Learning Repository. (1989). <https://doi.org/10.24432/C52P4X>.
6. Predicting Heart Disease Using Clinical Variables. Kaggle. <https://www.kaggle.com/datasets/thedevastator/predicting-heart-disease-risk-using-clinical-var>.
7. Hossain S, Hasan MK, Faruk MO, Aktar N, Hossain R, Hossain K: Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. *BMC Cardiovascular Disorders*. 2024, 24:214. [10.1186/s12872-024-03883-2](https://doi.org/10.1186/s12872-024-03883-2)
8. Chen T, Guestrin C: XGBoost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, 785-794. [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)
9. Mohan S, Thirumalai C, Srivastava G: Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*. 2019, 7:81542-81554. [10.1109/access.2019.2923707](https://doi.org/10.1109/access.2019.2923707)

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group–Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. *Cureus J Comput Sci* 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>

10. Biswas N, Ali MM, Rahaman MA, et al.: [Machine learning-based model to predict heart disease in early stage employing different feature selection techniques](#). BioMed Research International. 2023, 2023:1-13. [10.1155/2023/6864343](#)
11. Hamid M, Hajje F, Alluhaidan AS, bin Mannie NW: [Fine tuned CatBoost machine learning approach for early detection of cardiovascular disease through predictive modeling](#). Scientific Reports. 2025, 15:31199. [10.1038/s41598-025-13790-x](#)
12. Miao KH, Miao JH: [Coronary heart disease diagnosis using deep neural networks](#). International Journal of Advanced Computer Science and Applications. 2018, 9:1-8. [10.14569/ijacsa.2018.091001](#)
13. Dritsas E, Trigka M: [Application of deep learning for heart attack prediction with explainable artificial intelligence](#). Computers. 2024, 13:244. [10.3390/computers13100244](#)
14. Al Gharib S, Charafeddine J, Dornaika F, Haddad S: [Hybrid learning framework for explainable cardiovascular disease detection](#). IEEE Access. 2025, 13:130168-130184. [10.1109/access.2025.3591241](#)
15. Breiman L: [Random forests](#). Machine Learning. 2001, 45:5-32. [10.1023/a:1010933404324](#)
16. Wajgi R, Champaneria T, Wajgi D, Suryawanshi Y, Bhoyar D, Nilawar A: [Heart disease prediction using graph neural networks](#). International Journal of Intelligent Systems and Applications in Engineering. 2024, 12:280-287.
17. Rimal Y, Sharma N, Paudel S, Alsadoon A, Koirala MP, Gill S: [Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy \[RETRACTED\]](#). Scientific Reports. 2025, 15:13444. [10.1038/s41598-025-93675-1](#)
18. Arevalo J, Solorio T, Montes-y-Gómez M, González F: [Gated multimodal units for information fusion](#). arXiv. 2017, [10.48550/arXiv.1702.01992](#)
19. Rehman MU, Naseem S, Butt A, et al.: [Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment](#). Scientific Reports. 2025, 15:13361. [10.1038/s41598-025-96437-1](#)
20. El-Sofany H, Bouallegue B, El-Latif YMA: [A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method](#). Scientific Reports. 2024, 14:23277. [10.1038/s41598-024-74656-2](#)
21. Budhal S, Kawale S, Agarwal N: [Heart disease prediction using machine learning on a real-time clinical dataset with multi-diagnostic features](#). International Journal of Applied Mathematics. 2025, 38:2719-2742. [10.12732/ijam.v38i12s.1584](#)

How to cite this article:

Budihal S, Kawale S R, Agarwal N, et al. (May 08, 2026) Feature-Group–Aware Cardiovascular Disease Prediction Using Gated Neural Fusion and CatBoost-Based Clinical Feature Analysis. Cureus J Comput Sci 3 : es44389-026-00068-0. DOI <https://doi.org/10.7759/s44389-026-00068-0>