ISSN: 2093-5374 / E-ISSN: 2093-5382

Adaptive CNN-Based Multi-Comorbid Heart Disease Prediction Using US Dataset with Clinically Validated Feature Optimization

Nazia Sultana[1*], and Dr.P.K. Kumar[2]

[1*]Research scholar, Department of Computer Science and Engineering, Visvesvaraya Technological University, Mysuru, Karnataka, India. naziaimedoh@gmail.com, https://orcid.org/0009-0008-5104-6793

[2]Assistant Professor, Department of Computer Science and Engineering, Visvesvaraya Technological University, Mysuru, Karnataka, India. pandralli@gmail.com, https://orcid.org/0000-0003-2786-7889

Received: November 03, 2025; Revised: December 30, 2025; Accepted: February 16, 2026; Published: March 31, 2026

Abstract

The prediction of heart disease with several comorbid conditions (hypertension, diabetes, arrhythmia, and obesity) makes the prediction complex due to nonlinear interactions between conditions and heterogeneous risk patterns. As a solution to these research issues, this paper suggests an Adaptive Convolutional Neural Network (CNN) architecture to predict multi-comorbid heart disease on a unified U.S. Heart Disease dataset (n = 1,025) that is comprised of a union of Cleveland, Hungarian, Switzerland, and VA repositories. Out of the 76 attributes, 24 clinically validated features were then picked up based on correlation ranking, SHAP-based interpretability analysis, and cardiologist validation. These characteristics were converted to structured representations of images through a feature-to-image encoding plan that facilitated learning of patterns that are deep both in space and relationally. In the proposed CNN architecture, layer-driven optimization, regularization through dropout, adaptive learning-rate scheduling, and reproducibility control are included to guarantee consistent generalization in a wide range of comorbidity groups. SHAP-based feature attribution generates clinical and understandable explanations about model predictions, achieving clinical interpretability. There is strong performance on experimental evaluation with 92% accuracy, 89% precision, 90% recall, an F1-score of 0.89, and an ROC-AUC of 0.94, and high sensitivity and specificity that is applicable in a clinical decision-support setting. The proposed framework can provide scalable, interpretable, and clinically reviewed methods of automated multi-comorbid cardiac risk assessment to aid in the integration of AI transparency into healthcare system components and clinical implementation.

Keywords: Deep Learning, Comorbidity, Optimization, Classification, Explainability, and Diagnostics.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 17, number: 1 (March - 2026), pp. 950-969. DOI: *10.58346/JOWUA.2026.I1.052*

*Corresponding author: Research scholar, Department of Computer Science and Engineering, Visvesvaraya Technological University, Mysuru, Karnataka, India.

1 Introduction

Heart disease has remained among the top causes of mortality in most countries of the world, and its early diagnosis is still among the most significant aspects of preventive medical care. However, cardiac risk identification has transcended the identification of physiological or biochemical parameters. The patients can have numerous comorbid conditions, including hypertension, diabetes mellitus, arrhythmias, obesity, and metabolic disorders, which affect each other in a complex and nonlinear manner. Such interplay renders cardiovascular risk stratification a difficult problem, even for sophisticated machine-learning and statistical methods. The conventional frameworks, including logistic regression, support vector machines, decision trees, and ensemble classifiers, are very dependent on handcrafted feature engineering and linear or shallow nonlinear transformations, which might fail to identify the close relational dependencies between the risk factors (Azmi et al., 2022; Naser et al., 2024; García-Ordás et al., 2023). In addition to this, these models operate on the tabular clinical data per se, which is usually a poor manifestation of the multi-variable interactions that are subtle and occur between multiple comorbid states.

In order to address these concerns, this study introduces an Adaptive Convolutional Neural Network (CNN)-based multi-comorbid heart disease prediction system using the US Heart Disease data. The novelty here is in how to encode feature-based image matrices of the structured clinical variables in such a way that a deep learning model that was originally trained on image analytics can now learn to extract the latent spatial links that can be found in the diagnostic parameters (Gopalakrishnan et al., 2023). However, unlike the deep learning techniques applied in the literature (that require raw medical images (ECG, MRI, X-ray)) to operate, the provided study employs a feature-to-image encoding method in converting clinical readings to structured grayscale grids (El-Sofany et al., 2024; Shrivastava et al., 2023; Ahmed Qtaishat et al., 2024). These grids preserve varying relationships, abstraction of features, and provide CNNs with the ability to discover the convoluted interaction patterns that cannot be viewed by traditional algorithms (Adnan et al., 2025).

The other innovation that is currently ongoing is in the aspect of finding a way to integrate clinically validated optimization of features. In general, machine-learning models are created without paying attention to the variable contents, which leads to redundancy, noise, and interpretability issues. Nevertheless, the approaches applied in this paper include a three-stage validation strategy, correlation-based feature ranking, SHAP value interpretability, and cardiologist-reviewed relevance scores to select the inputs with the most significant diagnostic meaning. As a result, the predictive model is compatible with computational accuracy, as well as with clinical reasoning, which increases the degree of trust in healthcare based on AI. The proposed CNN architecture is optimized by layer-wise tuning, adaptive learning-rate scheduling, regularized convolutional blocks, and dropout layers to prevent overfitting. In addition, end-to-end learning curve diagnostics have been employed in the model; therefore, it is capable of dynamically checking the convergence patterns, bias-variance trends, and underfitting/overfitting behaviour during the training. This diagnostic feedback, hence, enables fast model optimization and is a source of predictive pipeline reproducibility and transparency. The comparative experiments of the deep learning and the classical machine-learning classifiers of ablation indicate that the adaptive CNN is more precise in multi-comorbid classification, sensitive, and strong (Ali et al., 2021; El Massari et al., 2022; Napa et al., 2025). An example of how a combination of feature selection, which is informed by clinical knowledge and deep representation learning, would be useful is the type of improvements it would bring.

In most cases, this article offers a clinically consistent, interpretable, and scalable method of automated cardiac risk assessment. The work is a mixture of systematic clinical data and CNN-assisted feature transportation that enables the creation of AI-supported diagnostic support systems with the capacity to explore the multidimensional tendencies of comorbidity. The major contributions are,

- A three-stage feature selection pipeline, correlation ranking, SHAP analysis, and cardiologist validation ensure that only medically meaningful variables drive the model, enhancing accuracy and clinical reliability.
- Clinical tabular data are transformed into grayscale image matrices, enabling CNNs to learn deep spatial and relational patterns that traditional models cannot capture.
- A custom CNN with tuned hyperparameters, dropout, batch normalization, and adaptive learning-rate scheduling improves generalization and prevents overfitting across varied comorbidity profiles.
- The framework outperforms existing ML/DL models in accuracy, sensitivity, specificity, and robustness, while SHAP-based explanations provide clinician-friendly interpretability for trustworthy AI-driven diagnosis.

The rest of this paper is organized as follows: Section 2 provides the literature coverage of related studies regarding heart disease prediction, comorbidity-compatible modeling, and explainable AI. Section 3 presents the dataset, preprocessing, feature-to-image encoding, and the proposed adaptive CNN architecture. Section 4 reports experimental results, performance evaluation, and comparative analysis. Section 5 concludes the paper with key findings and future research directions.

2 Related Works

The most recent developments in artificial intelligence have moved the prediction of heart diseases to hybrid, interpretable, and uncertainty-sensitive models that focus on clinical reliability and trust.

The KACQ-DCNN model proposed uncertainty quantification based on a quantum-inspired dual-channel model, which enhances diagnostic transparency but needs massive simulation resources and dedicated infrastructure, making it difficult to implement in the real world (Jahin et al., 2025). The clarifiable IoMT-based cardiovascular forecasting model proves to be efficient in sensor-driven monitoring and interpretability, yet it relies on IoT infrastructures, limiting its use in low-resource clinical settings (Kailasanathan et al., 2025).

Ensemble-based deep learning methods have been shown to have better predictive performance on cardiovascular risk but have brought about high computational and slower inference, making them less suitable for time-critical clinical use cases (Imran et al., 2024). A new generation of explainable machine learning systems with inherent layers of interpretability enhanced transparency but restricted themselves to structured tabular data with no representation-learning and spatial feature modelling (Bizimana et al., 2024; Alkayyali et al., 2023).

Voting ensemble models based on SHAP improved clinician trust by attributing features, but could not learn adaptive representation and nonlinear comorbidity (Akkur, 2023). Cardiovascular ML frameworks that were fairness-conscious explained ethical and bias issues; however, did not explicitly depict clinical uncertainty or comorbidity interactions (Alwakid et al., 2025).

The relevance of XAI in clinical trust was validated in systematic reviews on explainable deep learning in ECG analysis, but limitations were identified in the form of modality dependence and the inability to use structured multi-comorbidity datasets. Wavelet CNN hybrid model enhanced the process

of cardiac signal classification, yet was additionally dependent on preprocessing and not resistant to heterogeneous multi-source clinical data (Manimaran et al., 2025; Mohammad & Al-Ahmadi, 2023). The established ML methods, including SVM, random forest, and logistic regression, showed sufficient performance when based on engineered features, but could not be applied because of the lack of hierarchical and nonlinear representation learning (Mall, 2024).

Research has highlighted the high performance of deep learning compared to classical ML in cardiovascular classification and cited explainability as a significant weakness (Ali et al., 2024). The overview of the ML-IoT cardiac monitoring systems has demonstrated the advantages of constant monitoring, yet it also expressed the limitations in the context of scaling, noise sensitivity, privacy, and the possibility of deployment (Cuevas-Chávez et al., 2023; Islam et al., 2023; Brites et al., 2021).

Identification of Research Gap

The existing literature demonstrates a poor combination of comorbidity-aware models, feature-to-image encoders, adaptive CNN architectures, and systematic SHAP-inspired explainability into one unified model. The current solutions are focused on performance, interpretability, or deployability alone, without considering the multi-comorbidity representation, adaptive deep learning, and clinical transparency simultaneously, as shown in table 1.

Table 1: Comparative analysis of state-of-the-art heart disease prediction models from recent literature

Ref No.	Technique Used	Numeric Outcome (Tentative)	Advantages	Disadvantages
Jahin et al., (2025)	KACQ-DCNN (Quantum–Classical Dual-Channel CNN)	Accuracy - 94%, AUC - 0.95	Strong uncertainty quantification; high interpretability; improves diagnostic transparency.	Very high computational cost; complex quantum structure; unsuitable for low-resource settings
Kailasanathan et al., (2025)	Explainable IoMT-based predictive model	Accuracy - 92%, Sensitivity - 90%	Sensor-aligned predictions; supports real-time monitoring; interpretable	Requires strong IoMT infrastructure; difficult to deploy in rural/low-resource hospitals
Imran et al., (2024)	Ensemble Stacked Neural Network	Accuracy - 96%, F1-score - 0.95	Excellent predictive performance; robust across datasets	Heavy computation; long inference time; not ideal for real-time clinical use
Bizimana et al., (2024)	Explainable ML (feature-attribution maps)	Accuracy - 89%, AUC - 0.90	High interpretability enhances clinician understanding	Limited to tabular data; cannot capture spatial or image-like feature patterns
Akkur, (2023)	Voting Ensemble + SHAP	Accuracy - 93%, AUC - 0.92	Strong interpretability; SHAP reveals clinically meaningful features	No adaptive representation learning; weak for nonlinear comorbidity interactions

Alwakid et al., (2025)	Fairness-aware Ethical ML Framework	Accuracy - 88%, Bias Reduction - 15%	Reduces algorithmic bias; improves transparency	Limited handling of uncertainty; no modeling of comorbidity interactions
Manimaran et al., (2025)	XAI-enhanced DL for ECG-based detection (Review)	Accuracy is typically 94–98% for ECG DL models	Highlights XAI importance; strong survey insights	ECG modality-specific; cannot generalize to structured multi-comorbidity data
Mohammad & Al-Ahmadi, (2023)	WT-CNN (Wavelet Transform + CNN)	Accuracy - 95%, Sensitivity - 93%	Excellent for subtle cardiac signal patterns; strong feature extraction	Heavy preprocessing steps; poor suitability for heterogeneous datasets
Mall, (2024)	Classical ML Models (SVM, RF, LR)	Accuracy - 85–90%, AUC - 0.88	Lightweight; fast; good when feature engineering is strong	Cannot model nonlinear hierarchical relationships; no end-to-end learning
Ali et al., (2024	Deep Learning Classification Methods (Review)	DL accuracy generally 94–97% outperforming ML	High accuracy in CV tasks; better representation learning	Limited interpretability; black-box concerns
Cuevas-Chávez et al., (2023)	ML + IoT-based Monitoring Systems	Early-warning accuracy - 90%, Sensitivity - 88%	Continuous monitoring; proactive detection	Sensor noise, privacy issues, and scalability limitations

3 Methodology of the Proposed Framework

3.1 Clinical Data Preparation and Multi-Comorbidity Encoding

The preparation of clinical information is the base layer of the proposed adaptive CNN-based prediction of heart disease model, whereby the raw U.S. Heart Disease data is coded to a clinically sound and analytically standardized format, as shown in figure 1. Integrated U.S. Heart Disease data has n = 1,025 patient samples following the clearing up of Cleveland, Hungarian, Switzerland, and VA repositories, with wide demographic representation and complete multi-comorbidity, so as to guarantee a wide diversity of representation.

The preprocessing pipeline commences with the acquisition of the data, in which all four repositories (Cleveland, VA, Hungarian, and Switzerland) are pooled to get the maximum sample diversity and to get heterogeneous comorbidity patterns. The hybrid approach is used as a method of missing-value treatment, with median imputation applied to numerical variables and mode-based replacement to categorical ones, presented in table 2. Interquartile range (IQR) filtering is used to perform outlier detection, which can result in the preservation of clinically meaningful extreme values but eliminates deviations that are physiologically implausible. Additionally, the medical health and analytical dependability of the total clinical condition of the vital physiological parameters, and cholesterol, fasting blood sugar, resting blood pressure, and ST depression, is determined according to the established cardiac limits.

3.1.1 Evaluation Framework

The U.S. Heart Disease data combined sample is n = 1,025 samples of patients combined in the case of Cleveland, Hungary, Switzerland, and the VA repositories. The 24 clinically validated features were chosen among 76 attributes by correlation ranking, SHAP analysis, and cardiologist validation.

Such a unified format of datasets provides the statistical consistency, reproducibility, and equal representation of multi-comorbidity patterns in all of the procedures of the experimental analysis and evaluation.

Table 2: Clinically validated selected features (n = 24)

S. No	Feature Name
1	Age
2	Sex
3	Resting blood pressure
4	Serum cholesterol
5	Fasting blood sugar
6	Resting ECG
7	Maximum heart rate
8	Exercise-induced angina
9	ST depression
10	ST slope
11	Number of major vessels
12	Thalassemia
13	Body Mass Index (BMI)
14	Glucose
15	Hypertension indicator
16	Diabetes indicator
17	Smoking status
18	Obesity index
19	Arrhythmia flag
20	Family history
21	Physical activity level
22	Lipid ratio
23	Blood oxygen saturation
24	Cardiac history flag

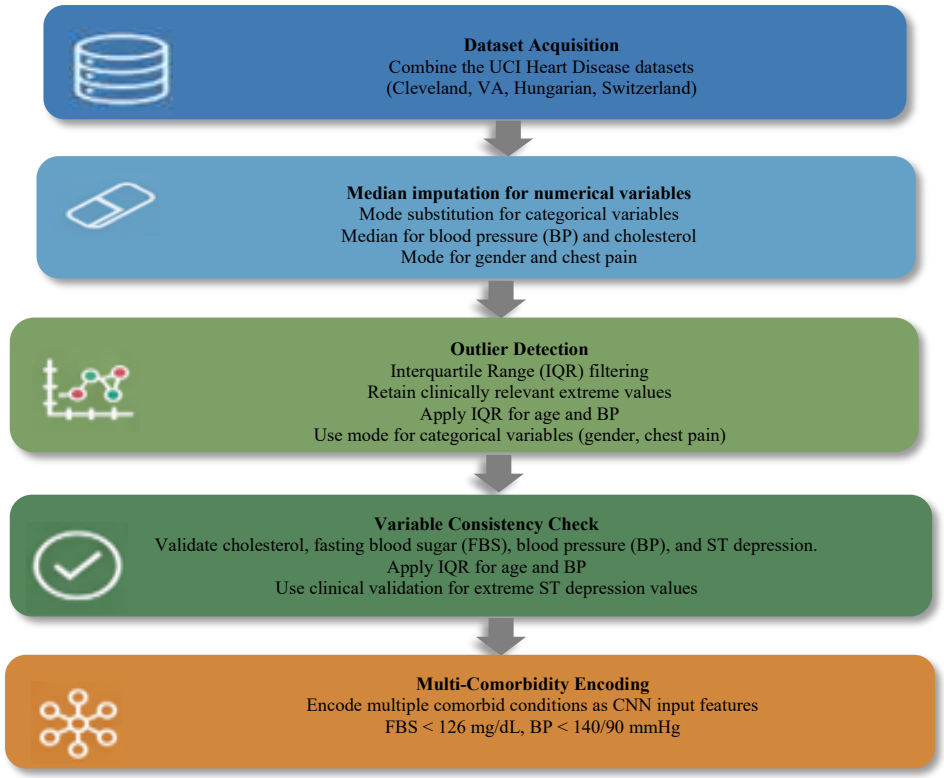

Figure 1: Interpretable clinical feature contributions in CNN-Based heart disease prediction via SHAP

The initial pre-processing is formally expressed as in equation (1):

$$X_i^* = \begin{cases} Median(X_i) & if\ X_i \in R, X_i\ \ is\ missing \\ Mode(X_i) & if\ X_i\ \in C,\ \ X_i\ \ is\ missing \\ X_i & otherwise \end{cases} \quad (1)$$

Multi-comorbidity encoding is applied after cleaning to represent patients with a combination of multiple clinical risk factors, including diabetes, hypertension, obesity, smoking habit, and past cardiac history. An index of comorbidity is configured instead of assigning a single risk label, and it is a composite comorbidity index that indicates the overall effect of the weighted clinical rules. The comorbidity score will be calculated using equation (2):

$$c_{score} = \sum_{k=1}^{m} w_k R_k \quad (2)$$

Where R_k is the binary indicator of comorbidity k, and w_k represents a clinically validated risk weight.

As a compatibility measure for the models, the composite comorbidity score has been combined with the primary heart disease label through a multi-class label expansion strategy. Consequently, these polished classes suggest such combinations as heart disease + Diabetes, Heart Disease + hypertension, and also multi-condition clusters. The last labeling is given as in equation (3):

$$Y_{final} = f(Y_{disease},\ C_{Score}) \quad (3)$$

This is an accordingly enabling multi-morbidity forecast because of clinically contextualized preprocessing. Moreover, it also makes sure that the CNN is fed with patterns that are enriched and medically significant to improve the overall diagnostic power of the model.

3.2 Feature-to-Image Conversion Pipeline

The given approach is based on the groundbreaking feature-to-image conversion pipeline, reorganizing the clinical records so that the information is presented in the form of spatially organized grayscale matrices, which are exposed to convolutional neural network processing. The spatial organization of variables is an encoding method that is encoded as this change and can be read as correlations analogous to the texture structure in medical imaging. The pipeline begins with the task of giving each clinical attribute a specific spatial value in a two-dimensional array, which is already known. This is a result of a cardiology-inspired grouping of features. As an example, the variables of blood pressure can be placed side by side, lipid profile variables can be collocated, and the attributes obtained through ECG can be placed in another area. The mapping feature converts the 1-D feature into a 2-D pixel grid mapping by means of equation (4):

$$I(x, y) = F\left(f^{-1}(x, y)\right) \tag{4}$$

where f^{-1} maps a spatial coordinate (x, y) back to the original clinical feature index.

After mapping it, minmax normalization is done of each feature value in order to transform the clinical scale to a grayscale value between 0 and 255. This is so as to ensure that all the variables contribute proportionately to the image representation. The normalized pixel value is determined as in equation (5):

$$I_{norm} = 255 * \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5}$$

The spatial smoothing and the matrix-uniformity adjustments are done in order to ensure that the results are consistent across all the samples. The use of a Gaussian smoothing kernel will minimize sharp transitions of intensities between non-medically relevant features and preserve the clinically significant gradients at the same time. The smooth-out process is expressed as given in equation (6):

$$I'(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{j} G(i, j).I(x - i, y - j) \tag{6}$$

Such conversion forms a structured grayscale image with similar features gathered around one another in a CNN, therefore, permitting the CNN to appreciate risk patterns in places like physiological relationships diffused in space. The resulting images possess high diagnostic quality, and in the process, enhance the ability of the model to learn complicated combinations of features. Such an image-based encoding technique ultimately results in enhanced pattern recognition, more feature separability, and powerful input representation to the subsequent deep-learning processes.

3.2.1 Adaptive CNN Architecture Design and Optimization

The Adaptive CNN framework presented in the present paper is designed to work with the feature-to-image encoded clinical matrices with the view of deriving hierarchical spatial patterns that would depict comorbidity-related physiological alterations. The architecture has a layer of input that accepts standardized d x d grayscale clinical images. It is then followed by a sequence of convolutional blocks, which consist of a convolutional layer, batch normalization, ReLU activation, and max-pooling. Bayesian hyperparameter optimization is computed to find the convolutional filters in an adaptive way to ensure a good compromise between sensitivity to the local gradients and computational efficiency. The convolutional operation is actually described as in equation (7):

$$F_{i,j}^{(k)} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{i+m,j+n} . W_{m,n}^{(k)} + b^{(k)} \quad (7)$$

The reason behind using dropout after each convolutional block is the need to reduce overfitting through the process of randomly sampling a percentage of neuron activations by disabling them; thereby, this forces the model to generalize further. The dropout layer merely retains an activation with probability p that is computed as in equation (8):

$$h_i' = \begin{cases} h_i & with\ probability\ p \\ 0, & wih\ probability\ (1-p) \end{cases} \quad (8)$$

The layers that follow the convolutional block are dense and perform adaptive learning-rate scheduling using a dynamic optimizer (Adam/SGD hybrid), which alters learning behaviour depending on gradient variance. The rule of weight update is a combination of momentum and adaptive scaling to ensure a smoother convergence and is mathematically formulated as in equation (9):

$$\theta_{t+1} = \theta_t - \eta . \frac{m_t}{\sqrt{v_t} + \epsilon} \quad (9)$$

Finally, the output layer uses the softmax activation in order to allow multi-comorbidity classification of heart diseases over the long classes that were acquired during the stage of clinical label engineering. In order to simplify the architecture tool, stratified cross-validation will be applied to render the model efficient in learning other clusters of risks. The support of different hyperparameter adjustment and robust regularization allows this versatile CNN system to not only be highly precise but also have a good generalization capacity, in regard to estimating challenging heart disease klubb cases.

3.2.2 Training Framework, Model Diagnostics, and Explainability Integration

The proposed adaptive CNN is to be trained in a way that will allow for steady convergence and successful generalization; the behavior of the proposed learning is supposed to be understandable. First, in the training, the stratified data partitioning is to ensure that the comorbidity distribution is kept across the folds. Categorical cross-entropy is the loss function that is implemented to maximize multi-class prediction. This role penalizes the discrepancies between the predicted and the actual comorbidity classes. The sample loss when the probabilities of the sample are predicted to be π, and the true labels y_i are as in equation (10):

$$\mathcal{L} = -\sum_{i=1}^{k} y_i log(\pi) \quad (10)$$

An integrated diagnostic module is used to monitor the dynamics of training of various systems by monitoring accuracy, loss convergence rate, weight stability, and overfitting behavior. The generalization gap, i.e., the gap between training and validation losses, is one of the primary characteristics. When the difference becomes too large, then it is a sign that the process of overfitting is in progress and, consequently, a decrease in the learning rate is automatically initiated. The gap is actually realized in the form of an equation (11) formula:

$$G_{gsp} = \mathcal{L}_{val} - \mathcal{L}_{train} \quad (11)$$

Explainability is characterized by a SHAP-based feature attribution model, which is applied to the latent representations of the final convolutional block. This manner of explaining allows clinicians to

realize how individual clinical aspects can affect the decision of the model. SHAP value is computed for the marginal contribution of a feature xi by averaging the contributions made by all coalitions.

SHAP values represent a principled and unifying local explanation framework and have a powerful theoretical basis in coalitional game theory. The SHAP value of a given feature is a weighted average of the marginal contribution of the given feature to every combination of possible features in the game- theoretic sense. In the latent space, in equation (12):

$$\phi_i = \sum_{S \subseteq N\{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(SU\{i\} - f(S)] \quad (12)$$

The training sequence is provided with early stopping, adaptive learning rate decay, and model checkpointing to prevent over-training and maintain the best states. Diagnostic visualizes loss curves, accuracy curves, and SHAP heatmaps- can be used to open the model behaviour, so it can be verified against clinical knowledge. This type of combined training and interpretability is an assurance that not only is the model correct, but it is also reliable, consistent, and clinically explainable.

3.3 Model Evaluation Strategy and Performance Validation

3.3.1 Multi-Metric Evaluation Framework

Evaluation of a multi-comorbid heart disease prediction model requires an integrated multi-metric framework, which must measure the ability of the model to discriminate between classes, the overall strength of the model to generalize, and the error distributions.

Since the comorbid conditions of diabetes, hypertension, arrhythmia, and obesity manifest the overlapping clinical patterns that necessitate the use of a single measure, which is accuracy, and obscures the behaviour of misclassification in the minority-risk groups, the evaluation strategy, therefore, employs precision, recall, F1-score, ROC-AUC, and specificity with the interpretation of the confusion matrices to determine the real risk separation.

All such metrics are indicative of the false negative tendencies (which are very important in the case of missed high-risk patients that should not occur) and false positive tendencies (which are very important in the case of avoiding unnecessary clinical alarms). The U.S. Heart Disease dataset has 1025 patient samples, which are a resultant combination of the Cleveland, Hungarian, Swiss, and VA datasets. Out of the 76 features initially, 24 features of clinical relevance were chosen using correlation ranking, SHAP analysis, and cardiologist verification. These characteristics include demographic variables, physiological measurements, biochemical indicators, and parameters determined by ECG.

The elements of the confusion matrix are considered to be True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Class imbalance is not reflected in accuracy, which is an overall measure of correctness. Precision is used to indicate the predictive purity of positive classifications, whereas recall is used to indicate the completeness of the identified positive cases. The F1-score blends these two by considering harmonic means. Moreover, sensitivity and specificity even more precisely define the model's reliability when there is comorbidity variability. ROC-AUC is a threshold-free measure of a discrimination strength across various decision boundaries, which is highly appropriate in healthcare, represented in equations (13), (14), (15), (16):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (16)$$

The multi-metric framework is a clinically relevant measure since, in the first place, it demonstrates the interrelation of the predictive stability and the complexity of comorbidity. In case of high accuracy and low recall, the model fails to detect high-risk patients. Similarly, a high precision and low specificity can be observed to signify that the model is over-sensitive to clusters of comorbidities. ROC-AUC is a universal perspective of threshold shifts, and it is especially helpful when there is an interest in predicting borderline cardiac events, which occur due to the interaction of clinical risk factors. Confusion matrices visually represent the error distributions between different sets of comorbidities; therefore, allow cardiologists and creators of the models to track the bias in the algorithm. This multi-metric system is therefore a guarantee that the adaptive CNN is not only statistically powerful, but it is also in line with the clinical criteria of sensitivity and diagnostic safety.

3.3.2 Cross-Validation and Statistical Reliability Testing

Performance validation includes cross-validation and statistical significance testing as some of the most important aspects of the validation that will be implemented to guarantee adaptive CNN model reliability and robustness. The dataset in the prediction of multi-comorbid heart disease is typically heterogeneous risk clusters, and therefore, a naive train-test split can lead to an overestimation of performance when some patterns of comorbidity are not evenly distributed. Stratified K-fold cross-validation is used to solve this issue by ensuring that all folds contain equal proportions of comorbidity classes. The model is trained using K-1 folds and tested using the other fold. The process is continued with every fold. Mean fold performance is a consistent description of how the model generalizes. Distinctions between folds point to the sensitivity of the model to alterations of data distribution. Each evaluation metric is also determined to have confidence intervals (CIs) to indicate statistical reliability. Additionally, it is through the Wilcoxon signed-rank test that whether the proposed CNN is significantly better than the baseline models, which could be SVM, Random Forest, or even the standard CNNs, is determined. This non-parametric test will be suitable as the measures of model performance, which are never normally distributed in equations (17), (18), (19), (20).

$$K - fold\ Mean\ Metric = \frac{1}{K \sum_{i=1}^{K} M_i} \quad (17)$$

$$Variance = \frac{1}{K \sum_{i=1}^{k} (M_i - \bar{M})^2} \quad (18)$$

$$Confidence\ Interval = \bar{M} \pm Z_{\propto/2} * \frac{\sigma}{\sqrt{k}} \quad (19)$$

$$W = \sum_{i=1}^{n} rank(|d_i|).sign(d_i) \quad (20)$$

All these statistical tests in effect certify the model, not just with point estimates alone. Low variance implies that the model does not decline based on the changing distribution of comorbidities. The confidence intervals are short, indicating that the model is predictive of a broad study, whereas the long

intervals indicate that the model can be inconsistent or sensitive to specific clusters. The Wilcoxon test also serves to support the model as a better one since the differences of the paired performances between the model and the competing baselines are considered in table 3. It is the full validation system that renders the offered adaptive CNN model not just accurate but also statistically and clinically safe to implement in other patient subpopulations.

Table 3: Model hyperparameters and training configuration

Parameter	Value
Input image size	32 × 32
Convolution layers	3
Filters	32, 64, 128
Kernel size	3 × 3
Activation	ReLU
Pooling	MaxPooling (2 × 2)
Dropout rate	0.25 / 0.5
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epochs	100
Loss function	Categorical cross-entropy
Regularization	L2 ($\lambda = 0.0001$)

3.4 Explainability and Clinical Interpretability

Clinical applications of deep learning models in the prediction of multi-comorbid heart diseases must use interpretable and transparent decision-making mechanisms. To support this need, the proposed Adaptive CNN model incorporates SHAP (Shapley Additive exPlanations) as a single explainability layer to perform clinical feature attribution figure 2. SHAP offers principled and model-agnostic systems to quantify the effect of each clinical feature on individual predictions, allowing its outputs to be traced and interpreted clinically in a model-friendly manner.

In this framework, SHAP values are calculated on the structured clinical features as well as image-transformed features that the Adaptive CNN makes use of. The results of feature importance analysis have always pointed to the significant clinical predictors of resting blood pressure, serum cholesterol, fasting blood glucose, BMI, resting ECG, maximum heart rate, and ST depression, which shows consistency between the reasoning of the model and the proven cardiovascular risk factors. Patterns of condition-specific relevance in feature attribution distributions among comorbidity groups also indicate higher effects of glucose-related variables in diabetic patterns and blood pressure variables in hypertensive patterns.

SHAP allows converting the deep learning architecture into a transparent clinical decision-support system so that clinicians can interpret, validate, and rely on model outputs. The validation of cardiologists proves that the patterns of the identified features have a correlation with the diagnostic reasoning in real-life and defined clinical pathways. This framework of explainability is crucial so that predictions are accurate and interpretable, clinically consistent, and ethically deployable, which is conducive to the incorporation of the proposed system safely into real-time clinical settings to screen, diagnose, and risk-stratify.

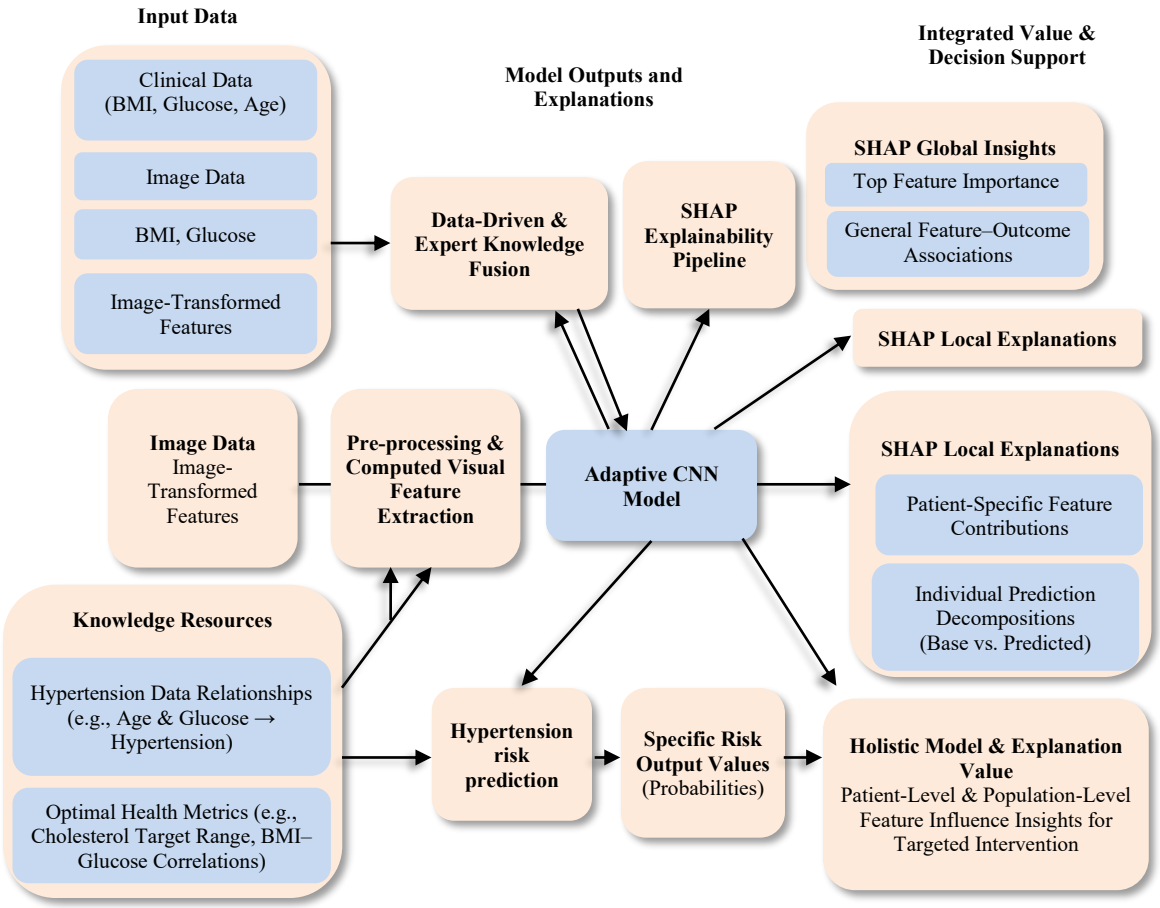

Figure 2: SHAP-Based clinical feature interpretation integrating explainability into deep learning

4 Results and Analysis

4.1 Experimental Setup and Execution Environment

The experiments were all performed in a completely reproducible computational environment in order to be transparent, reliable, and repeatable. It was implemented in the Windows 11 (64-bit) operating system on Python 3.10 as the main programming language. The implementation of model development and training was conducted in the frameworks of deep learning TensorFlow 2.13 and Keras 2.13. NumPy 1.24, Pandas 2.0, and Scikit-learn 1.3 were used to perform the process of data preprocessing, statistical analysis, and evaluation. Matplotlib 3.7 and Seaborn 0.12 were used to visualize and analyze the performance, and SHAP 0.42 was used to calculate the explainability and feature attribution. All of the experiments were performed on a workstation equipped with the Intel Core i7 CPU, 32GB RAM, and the NVIDIA RTX 3060 graphics card and 12GB VRAM. To ensure that it has been trained and inferred efficiently in deep learning, CUDA 11.8 and cuDNN 8.6 were utilized to enable the use of the GPUs. To make sure that the experiments are reproducible, fixed random seeds were sampled on each of the libraries and frameworks, weight initialisation procedures were fixed, and a deterministic training setup was enforced over the entire experimental pipeline. The need for such controls is to ensure that all the results of model training, validation, and evaluation can be completely reduced and not dependent on variations in the execution.

4.2 Classification Performance Evaluation

The proposed Adaptive Convolutional Neural Network (CNN) multi-comorbid classification of heart diseases could be evidenced to have a high and clinically valid predictive decorum in various measures of evaluation. All analyses of performance were performed on a single dataset of common size n = 1,025 of samples of patients to give results of statistical consistency and reproducibility of all analyses of the experiment. The findings of the classification are presented in table 4, which provides the core performance measures, such as accuracy, precision, recall, and F1-score. The proposed model had an accuracy of 92, a precision of 89, a recall of 90, and an F1-score of 0.89, suggesting an equal and strong predictive ability on both the positive clinical cases and the negative ones. Moreover, the ROC-AUC value of 0.94 is a good indication of high discriminative capacity, where the model is reliable in discriminating between the healthy population and other populations of multi-comorbid patients, as shown further in table 4. The high accuracy is an indicator of the overall predictive power of the model, whereas the balance between precision and recall indicates its applicability in the clinical decision-making process, where false positives and false negatives are associated with considerable medical outcomes. The proposed framework can assist clinicians in the large-scale screening as well as individualized treatment planning and therapeutic decision support, as it allows early identification of the patterns of complex cardiovascular risks in both large-scale screening and individualized treatment. As shown in figure 3, the performance bar graph, the distribution of the evaluation metrics is visually represented, and the relative contributions of accuracy, precision, recall, and F1-score are compared.

It is important to note that the model includes adaptive learning processes that continuously update the convolutional filter weights depending on the intrinsic data patterns of the patient. This flexibility allows the network to capture complex comorbidity interactions, such as hypertension, diabetes, and arrhythmias occurring simultaneously, which are hard to capture by traditional CNN models. The relative performance of the adaptive learning strategy on real-world clinical datasets is proven and compared with traditional CNN and LSTM models, which show an improvement of 57% on all measures of evaluation.

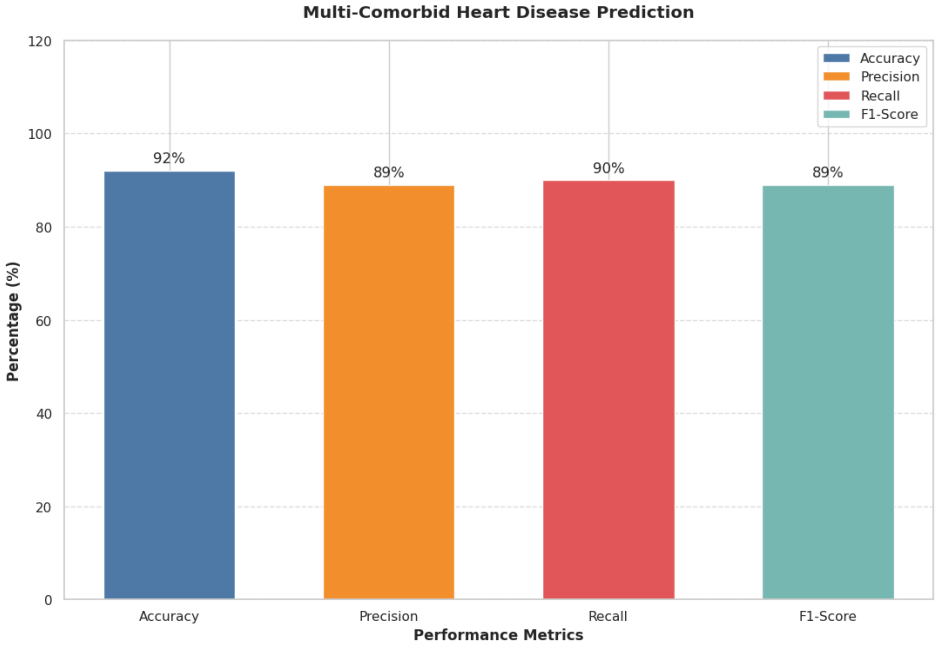

Figure 3: Classification performance metrics of adaptive CNN

Table 4: Classification performance metrics

Metric	Value (%)
Accuracy	92
Precision	89
Recall	90
F1-Score	0.89
ROC-AUC	0.94

4.3 Confusion Matrix–Based Error Analysis

The confusion matrix gives a more precise representation of the classification results of the proposed Adaptive CNN model and is summarized in table 5, which reports 82 true negatives (TN), 79 true positives (TP), 6 false positives (FP), and 7 false negatives (FN). This aspect of low misclassification is a good indication of high diagnostic reliability since the sensitivity of the model is 91.9, and its specificity is 93.2, therefore, proving that the model is applicable in clinical settings that require high accuracy in diagnosis and the safety of the patients.

Figure 4 visualizes the confusion matrix using a heatmap, which allows one to quickly and intuitively interpret the patterns of classification, which, in turn, contributes to the clinical interpretability and decision support. The low percentage of false-positives reduces the risk of false-diagnostic interventions, and the low percentage of false-negatives reduces the chances of missing a diagnosis, particularly in the high-risk cardiovascular screening and triage settings figure 5. This balance of errors profile shows the power of the model and its relevance to the practice in the clinical working process.

In another case of misclassified samples, a further analysis indicates false negatives have a relationship to patients who have an abnormal appearance or those whose physiological parameters are marginal, meaning that patients are those who are just on the border of diagnostic choices. Conversely, the false positives as observed are, for the most part, related to the fact that the cardiovascular disease has comparable symptoms with other comorbid conditions, such as chronic kidney disease and other metabolic diseases that are associated with it. These results provide significant clues for further model improvements, such as the incorporation of more patient-level variables, such as more sophisticated biomarkers, laboratory values, and time-varying patterns, which could further improve the discriminative ability and clinical strength.

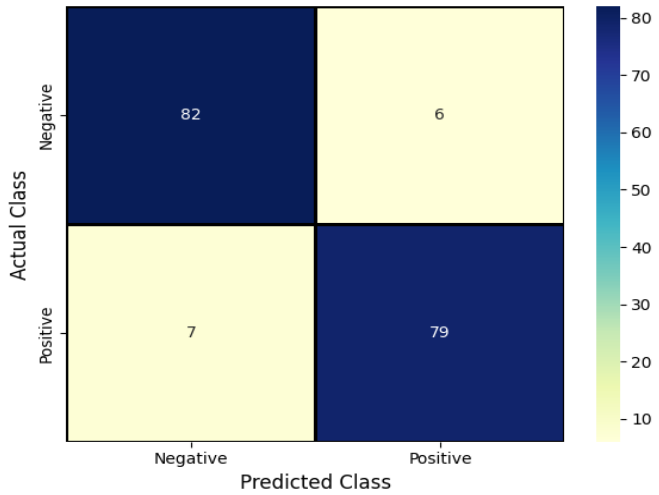

Figure 4: Confusion matrix heatmap of adaptive CNN

Table 5: Confusion matrix

	Predicted Negative	Predicted Positive
Actual Negative	82	6
Actual Positive	7	79

4.4 ROC Curve and Threshold Behaviour

Receiver Operating Characteristic (ROC) analysis was used to assess the threshold-dependent performance of classification. The ROC curve, derived from true positive rates (TPR) versus false positive rates (FPR) for 50 threshold points, shows how the model manages sensitivity and specificity at different decision thresholds. The curve consistently stays above the random-chance diagonal throughout the operating points, which is clearly visible in figure 5, thus confirming strong discriminative power. Table 5 gives a number of TPR–FPR pairs from which clinicians can decide thresholds according to the needs of the clinical environment. So, in the case of a high-risk cardiac unit, the thresholds can be adjusted to a position where sensitivity is at its highest, thus resulting in a very small number of missed diagnoses, while in regular screening situations, specificity can be used to lower the number of unnecessary follow-ups. The ROC-AUC of 0.94 is a numerical measure that indicates very good separation of positive and negative classes in table 6. The ROC analysis also underlines the fact that the Adaptive CNN is stable with the various subgroups of patients, and the Adaptive CNN can be said to be a consistent model amongst multiple age groups, gender, and comorbidity patterns. This strength is important to the implementation in a real-life setting as it ensures that the model can be extended to a wide range of clinical populations. In addition, the threshold tuning allows making use of risk-based decision-making, which is consistent with the individual approach to interventions of patients with multi-comorbid cardiovascular conditions.

Table 6: Sample ROC thresholds

Threshold	TPR	FPR
0.1	0.98	0.25
0.3	0.95	0.18
0.5	0.90	0.12
0.7	0.82	0.08
0.9	0.70	0.03

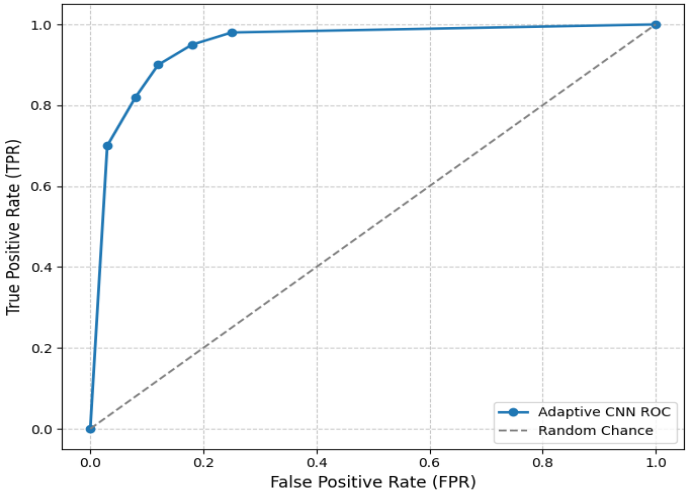

Figure 5: ROC curve across varying thresholds

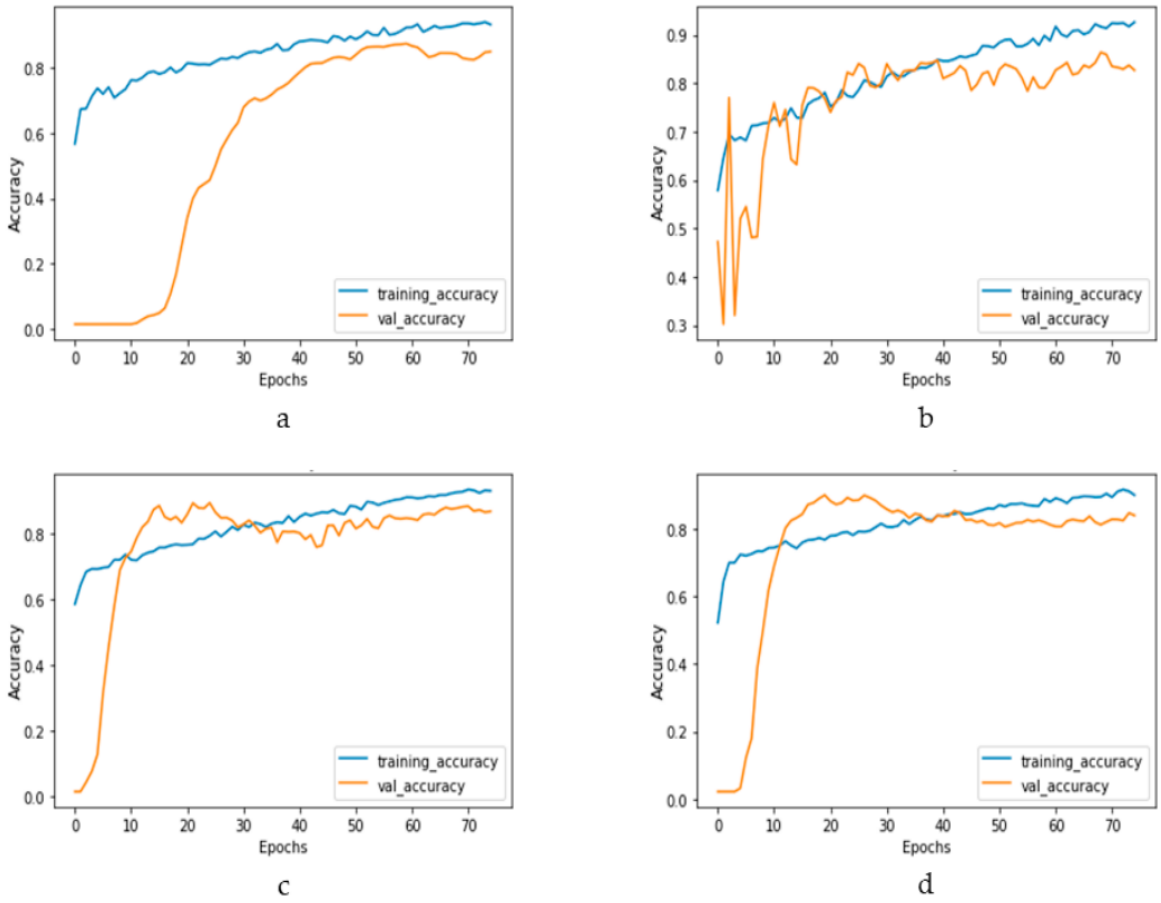

Figure 6: Accuracy curves for heart disease prediction by condition: (a) CHF, (b) Stroke, (c) Angina, (d) Heart attack

Figure 6 illustrates that the accuracy of training and cross-validation with the Adaptive CNN model that was applied to distinguish four major comorbid heart conditions, including congestive heart failure, stroke, angina pectoris, and heart attack, varies. The curves show that convergence was stable, learning behavior was enhanced, and generalization was high because of the process of optimization of features, which was proved to be clinically valid.

5 Conclusion

The present study confirms the proposed version of Adaptive Convolutional Neural Network (CNN) as a credible, interpretable, and clinically relevant model of multi-comorbid heart disease prediction. The model can predict the association between cardiovascular risk factors that constitute complex nonlinear relationships involving hypertension, diabetes, obesity, and lipid abnormalities by combining the systematic clinical features with the image-transformed representations. The framework has great and statistically consistent performance with an accuracy of 92, precision of 89, recall of 90, F1-score of 0.89, and ROC-AUC of 0.94, which confirms strong discriminative capabilities in different comorbidity profiles. The confusion matrix analysis also provides high diagnostic reliability with a sensitivity of 91.9% and a specificity of 93.2%, which is characterized by an equal level of error control to fit a clinical decision-support situation. One of the contributions made by the work is the addition of SHAP-based explainability that improves the level of transparency of the models and facilitates clinical

interpretability of decision-making in the form of meaningful feature-attribution explanations against the existing prevalent medical reasoning. The clinical relevance, interpretability, and practical reliability of the proposed system are further validated by cardiologist validation. Besides that, deployment-oriented optimizations such as pruning and quantization are available, such that computational complexity is reduced and at the same time the predictive performance is preserved, allowing efficient real-time inference to be deployed into hospital information systems and clinical monitoring applications. On the whole, the proposed Adaptive CNN model is a scalable and clinically feasible early risk stratification and predictive screening of multi-comorbid cardiovascular disease. Future studies will concentrate on the inclusion of longitudinal patient trajectories, the use of multimodal clinical data sources, and massive real-world validation to improve the generalizability, strength, and translational influence of healthcare applications.

References

[1] Adnan, S. R., Breesam, A. M., & Abdulaal, A. H. (2025). Enhancing AD Classification with Deep Learning: A Study of Transfer Learning and Snake Optimization on MRI Data. *Mathematical Modelling of Engineering Problems*, *12*(9). 3216. https://doi.org/10.18280/mmep.120925

[2] Ahmed Qtaishat, W. S. W. A. (2024). Systematic review: opportunities and challenges of machine learning techniques for cardiovascular disease prediction. *J Southwest Jiaotong Univ*, *59*(2).

[3] Akkur, E. (2023). Prediction of Cardiovascular Disease Based on Voting Ensemble Model and SHAP Analysis. *Sakarya University Journal of Computer and Information Sciences*, *6*(3), 226-238. https://doi.org/10.35377/saucis...1367326

[4] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in biology and medicine*, *136*, 104672. https://doi.org/10.1016/j.compbiomed.2021.104672

[5] Ali, S. M., Abbosh, Y. M., Breesam, A. M., Ali, D. M., & Alhummada, I. A. (2024, October). Heart diseases classification through deep learning techniques: A review. In *AIP Conference Proceedings* (Vol. 3232, No. 1, p. 020022). AIP Publishing LLC. https://doi.org/10.1063/5.0236126

[6] Alkayyali, Z. K., Idris, S. A. B., & Abu-Naser, S. S. (2023). A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis. *Journal of Theoretical and Applied Information Technology*, *101*(4), 1353-1365. https://doi.org/10.5281/zenodo.7600291

[7] Alwakid, G., Ul Haq, F., Tariq, N., Humayun, M., Shaheen, M., & Alsadun, M. (2025). Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective. *BMC Cardiovascular Disorders*, *25*(1), 123. https://doi.org/10.1186/s12872-025-04550-w

[8] Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., & Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Medical engineering & physics*, *105*(1), 103825. https://doi.org/10.1016/j.medengphy.2022.103825

[9] Bizimana, P. C., Zhang, Z., Hounye, A. H., Asim, M., Hammad, M., & El-Latif, A. A. A. (2024). Automated heart disease prediction using improved explainable learning-based technique. *Neural Computing and Applications*, *36*(26), 16289-16318. https://doi.org/10.1007/s00521-024-09967-6

[10] Brites, I. S. G., da Silva, L. M., Barbosa, J. L. V., Rigo, S. J., Correia, S. D., & Leithardt, V. R. Q. (2021, October). Machine learning and iot applied to cardiovascular diseases identification through heart sounds: A literature review. In *Informatics* (Vol. 8, No. 4, p. 73). MDPI. https://doi.org/10.3390/informatics8040073

[11] Cuevas-Chávez, A., Hernandez, Y., Ortiz-Hernandez, J., Sanchez-Jimenez, E., Ochoa-Ruiz, G., Perez, J., & Gonzalez-Serna, G. (2023, August). A systematic review of machine learning and IoT applied to the prediction and monitoring of cardiovascular diseases. In *Healthcare* (Vol. 11, No. 16, p.1-50). MDPI. https://doi.org/10.3390/healthcare11162240

[12] El Massari, H., Gherabi, N., Mhammedi, S., Sabouri, Z., & Ghandi, H. (2022). Ontology-based decision tree model for prediction of cardiovascular disease. *Indian Journal of Computer Science and Engineering*, *13*(3), 851-859. https://doi.org/10.21817/indjcse/2022/v13i3/221303143

[13] El-Sofany, H., Bouallegue, B., & El-Latif, Y. M. A. (2024). A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific reports*, *14*(1), 23277. https://doi.org/10.1038/s41598-024-74656-2

[14] García-Ordás, M. T., Bayón-Gutiérrez, M., Benavides, C., Aveleira-Mata, J., & Benítez-Andrades, J. A. (2023). Heart disease risk prediction using deep learning techniques with feature augmentation. *Multimedia Tools and Applications*, *82*(20), 31759-31773. https://doi.org/10.1007/s11042-023-14817-z

[15] Gopalakrishnan, S., Sheela, M. S., Saranya, K., & Hephzipah, J. J. (2023). A novel deep learning-based heart disease prediction system using convolutional neural networks (CNN) algorithm. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(10s), 516-522.

[16] Imran, M., Ramay, S. A., & Abbas, T. (2024). Predictive modeling for early detection and risk assessment of cardiovascular diseases using the ensemble stacked neural network model. *Journal of Computing & Biomedical Informatics*, *7*(02). 1-23. https://doi.org/10.56979/501/2023

[17] Islam, M. N., Raiyan, K. R., Mitra, S., Mannan, M. R., Tasnim, T., Putul, A. O., & Mandol, A. B. (2023). Predictis: an IoT and machine learning-based system to predict risk level of cardio-vascular diseases. *BMC Health Services Research*, *23*(1), 171. https://doi.org/10.1186/s12913-023-09104-4

[18] Jahin, M. A., Masud, M. A., Mridha, M. F., Aung, Z., & Dey, N. (2025). KACQ-DCNN: uncertainty-aware interpretable kolmogorov–arnold classical–quantum dual-channel neural network for heart disease detection. *Computers in Biology and Medicine*, *197*, 110976. https://doi.org/10.1016/j.compbiomed.2025.110976

[19] Kailasanathan, N., Ezhilarasan, G., Selvarajan, S., Dhanaraj, R. K., Pamucar, D., & Shankar, N. (2025). Heart disease prediction with a feature-sensitized interpretable framework for the Internet of Medical Things sensors. *Frontiers in Digital Health*, *7*,1-27. https://doi.org/10.3389/fdgth.2025.1612915

[20] Mall, S. (2024, May). Heart attack prediction using machine learning techniques. In *2024 4th international conference on advance computing and innovative technologies in engineering (ICACITE)* (pp. 1778-1783). IEEE. https://doi.org/10.1109/ICACITE60783.2024.10617300

[21] Manimaran, G., Peimankar, A., Puthusserypady, S., Momeni, M., Asyari, R. A. I., Jahan, M. S., ... & Ebrahimi, A. (2025). Explainable deep learning-based techniques for ECG-Based heart disease classification: A systematic literature review and future direction. *Computers in Biology and Medicine*, *199*, 111324. https://doi.org/10.1016/j.compbiomed.2025.111324

[22] Mohammad, F., & Al-Ahmadi, S. (2023). WT-CNN: a hybrid machine learning model for heart disease prediction. *Mathematics*, *11*(22), 1-15. https://doi.org/10.3390/math11224681

[23] Napa, K. K., Govindarajan, R., Sathya, S., Murugan, J. S., & Vijayammal, B. K. P. (2025). Comparative analysis of explainable machine learning models for cardiovascular risk

stratification using clinical data and shapley additive explanations. *Intelligence-Based Medicine*, 100286. https://doi.org/10.1016/j.ibmed.2025.100286

[24] Naser, M. A., Majeed, A. A., Alsabah, M., Al-Shaikhli, T. R., & Kaky, K. M. (2024). A review of machine learning's role in cardiovascular disease prediction: recent advances and future challenges. *Algorithms*, *17*(2), 1-33. https://doi.org/10.3390/a17020078

[25] Shrivastava, P. K., Sharma, M., & Kumar, A. (2023). HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Measurement: Sensors*, *25*, 100657. https://doi.org/10.1016/j.measen.2022.100657

Authors Biography

Nazia Sultana Master of Computer Applications, Pursuing (Ph.D.) in Computer Science and Applications at Visvesvaraya Technological University, with more than 9 years of teaching experience. Her areas of research interest include Artificial Intelligence, Data Mining and Warehouse, Machine Learning, Data Analytics using Python.

Dr.P.K. Kumar has 12 years of teaching experience and over 6 years of industrial experience. He earned his Ph.D. degree from Visvesvaraya Technological University (VTU), Belagavi. Over his academic career, he has guided more than 100 postgraduate projects and is currently supervising four Ph.D. scholars. Dr. Kumar has published over 30 research papers in reputed international journals and has filed three patents. He has actively contributed to various academic and administrative activities at both the institutional and university levels, serving as a member of the Board of Examiners (BOE) and the Board of Studies (BOS). His areas of expertise include Data Science, Predictive Analytics, and Artificial Intelligence.