SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I5P101

Volume 13 Issue 5, 1-12, May 2026

Original Article

A Hybrid Deep Learning Framework for Predicting Heart Disease: Combining CNN, LSTM, and Attention Mechanisms

Rasitha Banu Gul Mohamed[1], Sasikala[2], Maha Yousif Rizgalla sulieman[3], Hanan Abdullah Almaimani[4], Wafa Hetany[5], Faiza Abdalla Saeed Khiery[6]

[1,2]*Department of Public Health, HI Program, College of Nursing & Health Sciences, Jazan University, Jazan, Saudi Arabia.*
[3]*Department of Public Health, HEP Program, College of Nursing & Health Sciences, Jazan University, Jazan, Saudi Arabia.*
[4]*Master's in Public Health Epidemiology from Toledo University. Toledo, Ohio, USA, Lecturer, Department of Public Health Epidemiology, College of Nursing and Health Sciences, Jazan University, Jazan, Saudi Arabia.*
[5]*Department of Public Health, College of Nursing & Health Sciences, Jazan University, Jazan, Saudi Arabia.*
[6]*Department of Public Health, HIEP Program, College of Nursing & Health Sciences, Jazan University, Jazan, Saudi Arabia.*

[1]*Coresponding Author : rbanu@jazanu.edu.sa*

Received: 04 February 2026 Revised: 03 March 2026 Accepted: 02 April 2026 Published: 27 May 2026

Abstract - *Heart Disease (HD) continues to be among the top causes of death globally, and there is a need for precise and timely prediction models to aid clinical decision-making. This research suggests a new hybrid deep learning model by combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks to enhance the predictive accuracy of heart disease diagnosis. The model utilizes CNNs to extract spatial features from structured and time-varying clinical information, while LSTM and Bi-directional LSTM (Bi-LSTM) layers extract temporal dependencies and sequential patterns. Three configurations of architectures were tested: (i) CNN with 4 convolutional layers and a 2-layer LSTM with 64 units, (ii) CNN with 3 convolutional layers (32 filters) and a Bi-LSTM (64 units), and (iii) CNN with 5 convolutional layers (64 filters) and a single LSTM layer with 128 units. The performances were evaluated with regard to typical metrics accuracy, precision, recall, F1-score, and ROC-AUC score. Results showed that the Bi-LSTM-based setup obtained the best performance with an ROC-AUC value of 0.94 and high values for all the other measures. The results indicate that an equally balanced model structure with moderate complexity and bidirectional temporal learning yields the best predictive performance. The hybrid model holds the promise to be incorporated into clinical decision support systems to detect heart disease more accurately and in a timely fashion. This strategy is involved in the creation of intelligent diagnosis tools that can assist physicians in the identification of risky patients and in mitigating the burden of cardiovascular disease worldwide.*

Keywords - *Heart Disease Prediction, Hybrid Model, Convolutional Neural Networks, Long Short Term Memory, Attention Mechanisms.*

1. Introduction

Cardiovascular Disease (CVD) is the leading cause of death worldwide and claims 17.9 million lives annually. Conditions included are coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other respiratory diseases. They are a collection of cardiovascular diseases of the heart and blood vessels [1]. Of these, coronary HD is most common and usually most lethal. Early and precise prediction of heart disease is an important factor in avoiding related morbidity and mortality by means of early medical intervention and lifestyle change. Despite this, diagnosing and predicting heart disease pose a colossal challenge to the practice of medicine due to the heterogeneity of etiology, instability of presentation, and often insidious evolution of symptoms [2].

The conventional diagnosis of heart disease is highly dependent on a process involving clinical experience, patient history, physical examination, and some tests like Electrocardiograms (ECG), echocardiography [3], cardiac MRI, blood tests, and stress tests. Although they provide useful information, the disease is often not diagnosed early enough, nor are patients always correctly identified as being of the highest risk. In addition, the growing accessibility of large health data, e.g., Electronic Health Records (EHRs) [4], wearable sensor data, and imaging, has introduced new challenges in data analysis and interpretation, and intelligent tools to process and interpret high-dimensional, time-dependent data are needed.

During the past decade, the emergence of Artificial Intelligence (AI) [5] and DL [6] has been shown to have revolutionary power in medical diagnosis. These technologies facilitate computer-aided extraction of advanced representations and patterns from unprocessed clinical data to assist diagnostic efficiency and accuracy. DL methods, particularly CNNs [7] and Recurrent Neural Networks (RNNs) [8], such as LSTM networks [9], have already been found to be of monumental potential to employ in biomedical tasks since they have the capability of extracting spatial and temporal relationships among datasets.

Image processing and identifying spatial features are areas where CNNs have distinguished themselves [10]. In clinical diagnosis, CNNs have been used to a wide extent to deal with medical images such as MRI, CT scans, and echocardiograms with high potential to detect anomalous structures and patterns. They are programmed to learn spatial feature hierarchies automatically and dynamically from input through several convolutional layers, pooling layers, and Activation Functions (AFs). When used in the context of order or multivariate clinical information, CNNs may be able to detect patterns in the data that are not easily discovered by traditional statistical techniques or surface models of learning. LSTM networks, however, are designed RNNs to utilize to process temporal dependencies of sequence data [11].

LSTMs addressed the vanishing gradient problem that confronted RNNs to allow the model to learn long-term dependencies. This makes LSTMs extremely well-suited to time-series modeling like ECG signals, vital sign trends, or longitudinal patient records. Further, Bi-directional LSTM (Bi-LSTM) [12] networks are better than the normal LSTMs as they take the input sequence during both the forward pass and the reverse pass. This two-pass strategy enables the model to acquire dependencies that are possibly lost through considering just chronological order alone, enhancing the prediction accuracy of issues where earlier and later contextual information at points in time matters. Out of LSTMs' and CNNs' respective strengths being complementary to one another, hybrid structures incorporating combinations of both models have come forth as a more general approach towards tackling difficult biomedical prediction problems. Such hybrid models utilize the spatial learning ability of CNNs and the sequential modeling ability of LSTMs and are, therefore, most suitable for clinical data sets having both spatial and temporal content. In HD prediction, such models can input static patient information (e.g., demographics, lab values) and dynamic signs (e.g., time-series vital signs, ECG waveforms) concurrently, leading to more stable and efficient diagnostic outputs [13]. The model can potentially be incorporated into Clinical Decision Support Systems (CDSS) [14] in which it can help doctors make precise, data-driven inferences regarding patient health, leading to better clinical outcomes and resource utilization.

This study will introduce a new hybrid deep learning framework that combines CNN & Bi-LSTM to predict heart disease from either structured (i.e., ECG) or sequential (i.e., Electrocardiograms) clinical datasets. Most existing models use CNNs separately from LSTM for the purpose of either deriving spatially invariant features or modelling the temporal relationships of cardiac arrhythmias. Unlike these studies, the present study develops a single unified architecture that integrates the spatial and temporal capabilities of both CNNs and LSTMs. Virtually all previous hybrid models have relied on using only unidirectionally learned temporal features, as well as limited feature fusion approaches that have prevented them from developing sophisticated capture of clinical interdependencies. Conversely, the framework will employ bidirectionally learned temporal features, thereby improving context knowledge of cardiac events. Additionally, we will demonstrate that the proposed model outperforms current state-of-the-art methodologies across all quantifiable evaluation metrics, thereby filling the previously identified gap in the literature.

Although the existing methods of CNN-based systems mostly focus on extracting features from spatial data, these same methods do not adequately replicate long-term temporal dependencies in sequential medical records. On the other hand, LSTM-based models will include representations of temporal relationships among medical records, but often do not adequately represent spatial correlations when modelling medical information that contains structured features. A relatively new approach that combines CNN and LSTM is not able to overcome the problem of unidirectional temporal learning and feature fusion strategies. As such, there is an obvious research opportunity for a unified approach to combining the spatial representation of medical records from CNNs with the temporal representations of medical records through bidirectional LSTMs in order to improve the ability to predict cardiac risk accurately.

Research Question and Problem Statement: How does a hybrid CNN-LSTM framework with bidirectional LSTM influence the predictive accuracy of heart disease diagnosis using structured and time-varying clinical data? The problem is stated as follows.

The prediction of heart disease is limited by conventional statistical and machine learning models, which inadequately consider temporal dependencies along with the correlations and dependencies among features. Therefore, there is a need for hybrid deep learning models capable of considering spatial features and examining temporal trends in clinical and time-varying data. This dissertation addresses these shortcomings, highlights, and evaluates optimized CNN-LSTM frameworks to provide supportive evidence to assist health care providers with early and accurate detection of heart disease, and ultimately change determinant aspects of clinical decision-making. The objectives of the study are as follows:

- To design and implement a hybrid DL model that effectively integrates spatial feature extraction (via CNNs) with temporal pattern recognition (via LSTM and Bi-LSTM networks) using structured and time-varying clinical data.
- To compare and analyze the performance of three varying CNN-LSTM architectural configurations with different numbers of convolutional layers, filter sizes, and LSTM/Bi-LSTM layers on classification metrics, including accuracy, precision, recall, F1-score, and ROC-AUC.
- To identify the optimal model configuration for HD classification tasks that optimizes predictive power and architectural simplicity.
- To illustrate the practicability and promise of applying the presented model to real-time medical decision support systems for automatic prediction of HD.
- To help extend the research frontier of medical AI and DL by investigating hybrid models that may be extendable to other illnesses and diagnostic tasks, including structured and sequential health data.

The rest of the paper is structured as follows. Section 2 gives a summary of the literature on HD prediction using ML and DL methods. It touches upon CNNs, LSTMs, and the application of hybrid models in medical diagnosis, mentioning areas where there are research gaps that this research intends to bridge. Section 3 summarizes the dataset, preprocessing techniques, model architecture formulation, and execution of the three hybrid model instances. Section 4 presents experimental results comparing the performance of the three configurations in standard classification measures. Section 5 concludes the findings, restates the contributions to the work, and outlines directions for extending the model's application to additional datasets, diseases, or deployment in real-world systems.

2. Related Works

Recent literature highlights the increasing utilization of hybrid deep learning solutions for heart disease and cardiovascular risk prediction. CNN-LSTM and Bi-LSTM models often enhanced with attention, SMOTE, or transformers are generally superior at capturing spatial and temporal information from clinical data, ECG, and heart sound signals. For example, Gray Wolf Optimization and Eurygaster Optimization, with feature selection, optimization with deep belief networks, both improve the efficiency and accuracy of models. It is concluded from the literature that hybrid deep learning frameworks hold promise for early and reliable diagnosis of cardiovascular disease, and can tackle the issues of unbalanced data, complex interactions of features, and sequential dependencies.

Bhavekar et al. (2022) devised a hybrid DL approach for the classification of HD. RNN and LSTM hybrid approaches have been applied to the classification of synthetic data through a variety of cross-validations. Additionally, a variety of ML and soft computing approaches will be employed to evaluate the system's performance. Deep hybrid learning outperforms both traditional DL and ML methods if employed separately [15].

Lilhore et al. (2025) introduce a novel CNN-Bi-LSTM hybrid model that combines the SMOTE with an attention-driven enhanced CNN-RNN architecture to enhance heart sound categorization, particularly from unbalanced datasets. In order to enhance the architecture by obtaining contextual information from both previous and subsequent time sequences, the study combines a Bi-LSTM network with an enhanced CNN to efficiently extract global and local characteristics [16].

Shi et al. (2022) suggest an ECG-based Attention mechanism enabled a Bi-LSTM (ABLSTM) model to automatically identify Congestive Heart Failure(CHF) in IoMT systems. This approach efficiently extracts complex characteristics of ECG signals and carries out detection by utilizing the model features that can store both short-term and long-term signal information, as well as the AM's adaptive learning capabilities in local features [17].

Albathan et al. (2024) suggest an innovative hybrid method for predicting Heart Attacks (HA) utilizing LSTM networks, robust data mining, and deep belief networks (DBNs) with attention mechanisms. The Kaggle, PhysioNet, and UCI datasets are used to predict HA. To produce a reliable and thorough prediction model, the hybrid model architecture employs DBNs for feature representation and selection, and LSTM networks for sequence learning [18].

Hao et al. (2025) developed a CNN and transformer hybrid model for the diagnosis and prognosis of HD. The Transformer's strong capacity for sensing global relations and CNN's power in identifying local features allow the model to identify HD risk factors from high-dimensional life history data effectively. This illustrates how well it handles unstructured and multi-dimensional data [19].

Shrivastava et al. (2023) employ CNN & Bi-LSTM to address the imbalance issue and build a hybrid model based on DL techniques for determining if a person has HD and offering awareness or a diagnosis based on that prediction. The study uses a CNN-BiLSTM for classification and an additional tree classifier for feature selection [20].

Rahman et al. (2023) predict CVD risk using an innovative self-attention-based transformer model, which integrates transformer networks with self-attention techniques. Self-attention layers are able to effectively model complex patterns in the data by generating representations and extracting contextual information. Interpretability is facilitated through self-attention mechanisms, which apply a particular degree of attention weight to every input sequence element. To achieve relevant data, this involves modifying the input and output layers, inserting more levels, and reconfiguring the attention processes. Moreover, this assists physicians in comprehending which data variables affected the predictions of the model [21].

Saheed et al. (2024) propose a novel, effective, real-time technique for diagnosing Cardiovascular Disease (CVD) using Supervised ML (SML) hyperparameter tuning. To make feature selection easier, the study uses an Attention Mechanism (AM) in conjunction with a new adaptation of the BiLSTM model. The SageMaker cloud environment is used to outsource the model's deployment at scale. The proposed model's simplicity is ascribed to its lower computing complexity [22].

Ramesh et al.(2024) proposed the Convolutional LSTM (O-SBGC-LSTM), which is improved by the Eurygaster Optimization Algorithm (EOA), as a way to tune hyperparameters for the early identification and prevention of diabetes illness. This method explores the co-occurrence relationship between the spatial and temporal domains and captures discriminative features in spatial configuration and temporal dynamics [23]. Table 1 provides the related studies pertaining to the present study.

Table 1. Studies in the literature for HD prediction

Author (Year)	Method	Strengths	Limitations
Bhavekar et al. (2022)	Hybrid DL using RNN and LSTM with cross-validation on synthetic data	Effective combination of DL models; improves classification; robust evaluation	Use of synthetic data may not generalize well to real-world clinical scenarios
Lilhore et al. (2025)	CNN-BiLSTM with SMOTE and attention mechanism	Handles unbalanced datasets; combines local/global features; uses contextual time sequence information	Increased model complexity; SMOTE may introduce noise
Shi et al. (2022)	ECG-based Attention-enabled BiLSTM (ABLSTM)	Captures both short- and long-term dependencies; strong local feature learning via attention	Specific to ECG signals, computationally intensive in IoMT scenarios
Albathan et al. (2024)	Hybrid model with LSTM, Deep Belief Networks (DBNs), and attention	Uses multiple datasets; strong feature representation and sequence learning	The model may be too complex for real-time use; it requires fine-tuning
Hao et al. (2025)	CNN + Transformer for diagnosis and prognosis	Effectively processes high-dimensional, unstructured data; strong global and local feature extraction.	High computational cost; transformer training data-dependent
Shrivastava et al. (2023)	CNN-BiLSTM hybrid with tree-based feature selection	Addresses class imbalance; adds interpretability via a tree classifier; combined temporal and spatial features	Potential overfitting; interpretability is still limited compared to rule-based systems
Rahman et al. (2023)	Transformer with self-attention for CVD prediction	High interpretability; strong contextual learning; adaptable attention layers	Requires large data for effective training; complex for smaller devices or edge deployment
Saheed et al. (2024)	BiLSTM + attention + SML with hyperparameter tuning on SageMaker	Lightweight and real-time capable; cloud scalability; attention mechanism simplifies feature selection	Dependent on cloud environment; not benchmarked against other scalable real-time systems
Ramesh et al. (2024)	O-SBGC-LSTM	Captures both spatial and temporal features; optimization improves accuracy.	Primarily applied to diabetic illness, complex optimization may not generalize easily to HD.

Recent progress in HD forecasting has more and more utilized hybrid deep learning models that integrate models including CNNs, LSTMs, Bi-LSTMs, attention mechanisms, and transformers to achieve better diagnostic performance. These approaches are ideally adapted for extracting both spatial and temporal features in intricate biomedical signals such as ECG and heart sounds. Transformer and attention models are employed to improve the interpretability and efficiency of the model by concentrating on important data and long-distance dependencies. For addressing dataset imbalance issues that occur in the majority of medical data, algorithms like SMOTE are applied for over-sampling minority classes. Some works also prioritize real-time deployment capability along with interpretability for applications in Internet of Medical Things (IoMT) and cloud environments. Still, some research gaps remain.

Most models are calibrated using synthetic or benchmark data, which restricts the applicability of the models to various real-world clinical situations. High computational expense also restricts the applicability in low-resource or real-time settings. One uniform set of evaluation metrics does not exist, and hence, directly comparing across models is impossible. Most existing methods also utilize single-modal data, while effective multi-modal data fusion is not well-explored. End-to-end deployment pipelines covering all the stages from data acquisition to clinical decision support are rarely considered. Explainability, although partially augmented by attention mechanisms, is still not fully integrated to the extent of clinical validation necessities. Lastly, the majority of studies target binary classification rather than disease progression or staging modeling, which is essential in preventive cardiology. Overcoming the above limitations can lead to more robust, scalable, and clinically useful AI-based diagnostic tools.

3. Methodology

A hybrid DL method is suggested in this study that employs CNN, LSTM, and Attention Mechanisms (AM) for the prediction of HD. This study investigates a novel hybrid DL approach that leverages the strengths of both CNN and LSTM models to better predict HD. The new system is meant to handle structured clinical data that entails time-related attributes, learning spatial as well as sequential attributes to make intelligent predictions in HD probability. The hybrid architecture features early feature extraction using a CNN module, subsequent LSTM or Bi-LSTM layers extracting temporal dependencies from data. Three model settings were constructed and used in experimentation of the proposed method's efficiency:

- CNN-LSTM Configuration I: Four-layer CNN followed by two-layer Bi-LSTM with 64 units. The first setting targets greater spatial features and average sequential learning capacity.
- CNN-BiLSTM Configuration II: A three-layer 32-filter CNN and a single 64-unit Bi-LSTM layer. This configuration explores the influence of bidirectional sequence learning along with shallow CNN depth.
- CNN-LSTM Configuration III: A more intricate CNN with five convolutional layers (with 64 filters) and one LSTM layer of 128 units. This configuration explores the influence of model intricacy on prediction accuracy.

Data preprocessing, feature extraction, design of the hybrid model, training, evaluation, and comparison against existing standard models are conducted.

3.1. Dataset Description

In this research, publicly accessible datasets for HD, predominantly the Cleveland HD dataset [24] from the UCI ML Lab, are used. The dataset comprises roughly 300 to 1000 patient records, and each record has 13 to 15 clinical features pertaining to heart health. Age, gender, type of chest pain, resting Blood Pressure(BP), serum cholesterol, fasting blood glucose, resting ECG results, peak heart rate attained, exercise-induced angina, and ST depression (oldpeak) are among the most important variables. The target variable is binary in nature, indicating the presence (1) or absence (0) of HD. This dense set of clinical parameters provides a suitable foundation for building and validating predictive models. Table 2 presents the most important features in the HD dataset, which are helpful in predicting HDs.

Table 2. HD dataset-Key attributes and their description

Attribute	Description
Age	Age in years
Sex	Gender (0 for female, 1for male)
Chest Pain Type	Type of chest pain (0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic)
Resting BP	Resting BP when admitted to the hospital
Serum Cholesterol	Serum cholesterol level in mg/dl
Fasting Blood Sugar	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
Resting ECG Results	Results from the resting ECG (0, 1, 2 indicating different abnormalities)
Maximum Heart Rate Achieved	The highest heart rate achieved during the exercise test
Exercise-Induced Angina	Exercise-induced angina (1 = yes; 0 = no)
ST Depression	Depression in the ST segment observed after exercise relative to rest, indicating ischemia

3.2. Data Preprocessing

Prior to training the model, some preprocessing had already been done in certain sections for maintaining the quality of the data as well as for enhancing the performance of the model. Missing values of the dataset were dealt with by replacing the mean value for numeric attributes and the mode value for categorical attributes. In order to normalize the data and for faster convergence of the model, the numerical characteristics are normalized using Min-Max scaling with 0 to 1 range values. Category features were mapped by one-hot encoding to machine-readable values. The dataset was divided into training and test sets, with 20% going toward testing and 80% going toward training. Furthermore, 5-fold cross-validation was used in training to improve the robustness and generalizability of the model so that the model would be able to perform well on various subsets of data.

3.3. Hybrid Model Architecture

The hybrid DL model that has been proposed combines CNN, LSTM, and Attention mechanisms to better predict HD based on extracting spatial and sequential patterns from clinical data. First, raw tabular data is reshaped into a 2D format to be used with CNNs. The CNN module uses 1D convolutional layers to capture local interaction features, followed by ReLU activations and 1D max-pooling operations to downsample and promote valuable patterns. The obtained features are fed into the LSTM module, which excels at capturing sequential dependencies and intricate feature interdependencies with respect to time. To avoid overfitting, dropout regularization is added after the LSTM layer. Then, an attention mechanism is utilized to provide dynamic weights to features to allow the model to concentrate on the most relevant clinical attributes affecting HD prediction. This self-attention layer enhances model performance and interpretability. Finally, the output layer, having a sigmoid Activation Function (AF), takes the learned feature representations and passes them to the fully connected dense layers to predict the probability of HD. The hybrid technique performs better than traditional ML models in the predictive ability of CNNs, LSTMs, and attention. The overall workflow of the hybrid architecture is shown in Figure 1.

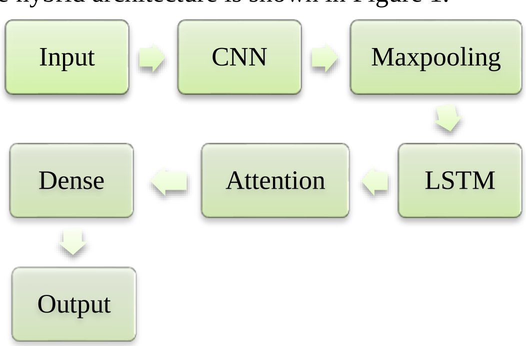

Fig. 1 Overall workflow of the proposed hybrid model

3.3.1. Convolutional Neural Network (CNN)

In the suggested framework, the clinical tabular information was initially reshaped into a 2D form suitable for processing by CNNs.This redesign enables the model to take advantage of spatial interactions between features. The CNN module starts with 1D convolutional layers, which are designed to automatically learn local feature patterns and correlations between adjacent clinical attributes. The workflow for the CNN is shown in Figure 2.

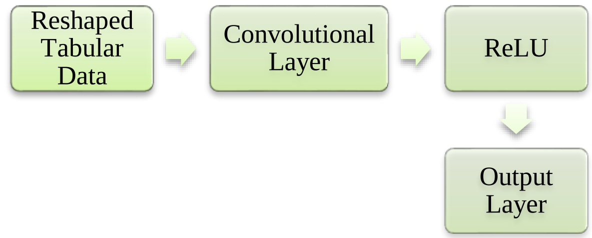

Fig. 2 Workflow of CNN

Mathematically, the convolution operation on an input vector x and a kernel w can be formulated.

$$y[i] = (x * w)[i] = \sum_{k=0}^{k=1} x[i+k].w[k] \quad (1)$$

In equation (1), K is the kernel size, i is the position index, and y[i] is the output feature map. After convolution, the Rectified Linear Unit(ReLU) AF is defined as

$$ReLU(z) = \max(0, z) \quad (2)$$

Is used element-by-element to give the model non-linearity, which enables it to learn intricate representations. The feature maps are downsampled using 1D max-pooling layers after activation, which lowers their dimensionality while keeping the most important data.

Max pooling over a window size p selects the maximum value within each window, formally.

$$y[i] = \max(x[i], x[i+1], \ldots . x[i+p-1]) \quad (3)$$

This convolution-pooling pipeline effectively extracts robust, localized patterns from clinical data, which are then passed to the subsequent LSTM methods for further temporal dependency modeling [25].

3.3.2. Long Short-Term Memory (LSTM)

Following feature extraction through the CNN module, the feature representations are input into an LSTM network. Long-term dependency learning is possible using LSTMs, a unique kind of RNN, and sequential relationships between input features, which are critically relevant for learning how clinical attributes may interact to influence HD risk over time. The vanishing gradient issue is addressed with LSTMs, which employ gates to control information flow in contrast to conventional RNNs. An LSTM unit has three primary gates: the input gate, forget gate, and output gate. The operations in an LSTM cell at time step t are mathematically expressed as:
Forget gate :

$$f_{t=} \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

Input Gate :

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\hat{C}_t = \tanh(W_c . [h_{t-1}, x_t] + b_C \quad (6)$$

Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (7)$$

Output gate:

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = O_t * \tanh(C_t) \quad (9)$$

In the above equations, x_t is the input at time t, h_{t-1} is the hidden state from the previous time step, W_f, W_i, W_C, W_o are the weight biases, b_f, b_i, b_C, b_o are the bias terms, σ is the sigmoid AF, tanh is the hyperbolic tangent AF, and *=is the element-wise multiplication [26].

3.3.3. Attention Mechanism

Following the feature representations through the LSTM, a self-attention mechanism is utilized to further the methods on the most important clinical features that define HD.

The attention mechanism allows the technique to allocate varying weights to various areas of the input data rather than treating all features uniformly. In other words, it dynamically adjusts each attribute in making the detection.

In a self-attention layer, each input feature will get to talk to every other feature so as to come up with an attention score. Formally, given an input matrix X, the attention method comes up with three matrices: Query (Q), Key (K), and Value (V) through learned linear transformations.

$$Q = XW^Q, \text{K=X}W^k, V = XW^V \quad (10)$$

In equation (10) $W^Q, W^K, W^{V,}$ they are trainable weight matrices. The attention scores are calculated with the scaled dot product.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})\text{V} \quad (11)$$

In equation (11) d_k is the key vectors' dimensionality. The attention scores are standardized across all features thanks to the softmax function.

This mechanism allows the model to selectively "pay attention" to notable clinical features such as type of chest pain, BP, and cholesterol while predicting, improving interpretability, and predictive accuracy.

By learning what features have the most influence, the model not only improves but also provides insight into what is most important for HD [27].

3.3.4. Fully Connected Layers

Finally, after the feature representation has been extracted and transformed by the CNN, LSTM, and Attention modules, the last operation of the hybrid model is the use of fully connected (dense) layers. The dense layers are employed to further continue abstractions and blend the learned features in an attempt to assist with classifying the final output. Two heavy layers are employed, both of which are preceded by the ReLU (Rectified Linear Unit) AF that introduces non-linearity to the model. The ReLU function is given by:

$$ReLU(z) = \max(0, z) \quad (12)$$

And allows the method to learn complicated patterns through sparse activation and fast gradient propagation.

The dense layer of output has only a single neuron with sigmoid activation. The sigmoid function gives any value between 0 and 1, and it is the perfect function to be used for binary class problems, such as the presence of HD (1) or its absence (0). Sigmoid function can be represented mathematically as:

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (13)$$

In equation (10), z is the input to the neuron. A threshold is then imposed over this probability (usually 0.5) to yield the final class output. This application of dense layers, along with an output sigmoid unit, provides a safe, interpretable, and probabilistic output in medical diagnosis.

4. Results and Analysis

The effectiveness of the suggested hybrid DL model was assessed by comparing it with the preliminary models, such as Logistic Regression (LR), Random Forest (RF), XGBoost, standalone LSTM, and standalone CNN models. A straightforward yet effective baseline consisted of machine learning models like LR, a common linear classifier with wide applicability for binary classification tasks.

Additionally, RF was used, an ensemble technique that builds a collection of decision trees and averages their predictions to improve prediction and prevent overfitting. Additionally, XGBoost, a powerful gradient boosting algorithm with high efficiency and strong performance on structured data, was incorporated as a standard against more sophisticated ML methods.

For the DL component, individual models were constructed using LSTM and pure CNN networks according to their respective capabilities. While the single LSTM retrieved sequential dependencies among clinical characteristics, the single CNN was used to extract local patterns of features. A thorough understanding of how much performance is improved by combining CNN, LSTM, and attention mechanisms can be obtained by comparing the hybrid model to existing techniques.

The experiments were carried out utilizing Python with Tensorflow and Keras libraries on High-Performance Computing (HPC) hardware with NVIDIA GPUs. GPU acceleration allowed for efficient training of the operation of deep learning models. Random seeds were maintained consistently to ensure reproducibility. The experiments will use the same hardware and software to maintain consistency and the reliability of results.

Data from the HD and cardiovascular disease data sets were used in the tests. HD is predicted using a number of health indicators from the dataset. Class 0 showed the absence of an illness, whereas class 1 indicated its presence. These characteristics were used in the classification process. The test data set is used to evaluate the algorithm's efficacy after it has been trained using the suggested hybrid approach, utilizing the training data set. Data sets are divided 80:20 between training and testing. Table 3 provides the list of hyperparameters used in this experiment.

Table 3. Hypermeters used in this study

Component	Hyperparameter	Value/Setting
CNN Module	Number of Conv1D Layers	2
	Filters per Conv1D Layer	64, 128
	Kernel Size	3
	AF	ReLU
	MaxPooling1D Pool Size	2
LSTM Module	Number of LSTM Units	128
	Dropout Rate	0.3
Attention Mechanism	Attention Type	Self-Attention
Fully Connected Layers	Number of Dense Layers	2
	Neurons /Dense Layer	64, 32
	AF(Dense)	ReLU
Output Layer	AF	Sigmoid
Training Parameters	Optimizer	Adam
	Learning Rate	0.001
	Batch Size	32
	Number of Epochs	100
	Loss Function	Binary Cross-Entropy
Validation	Cross-validation	5-Fold

Table 4. Performance analysis –proposed hybrid model

Model	Accuracy	Precision	Recall (Sensitivity)	F1-Score	ROC-AUC Score
Logistic Regression	0.82	0.80	0.81	0.80	0.85
Random Forest	0.85	0.84	0.86	0.85	0.88
XGBoost	0.87	0.86	0.88	0.87	0.90
Standalone CNN	0.84	0.82	0.85	0.83	0.87
Standalone LSTM	0.83	0.81	0.84	0.82	0.86
Proposed Hybrid model	0.90	0.89	0.91	0.90	0.93

Table 4 displays the performance of several models for predicting HD in terms of five evaluation metrics: Accuracy, Precision, Recall (Sensitivity), F1-Score, and ROC-AUC Score. The Proposed Hybrid approach (CNN + LSTM + Attention) performs better than all baseline models for all metrics.

The most accurate among the hybrid models has an accuracy of 90%, i.e., it does it correctly more often when determining whether there is HD or not, higher than the others. XGBoost at 87% and Random Forest at 85% lag behind, with the worst being Logistic Regression at 82%.

The precision of 89% in the hybrid model denotes its high potential to accurately predict positive HD cases, reducing false positives. The other models, such as XGBoost (86%) and Random Forest (84%), are relatively low.

The hybrid model's recall of 91% is the best among all, i.e., it captures nearly all the actual cases of HD. This is very crucial in medical conditions where a false negative can be disastrous. Balancing precision and recall, the hybrid model's F1-score of 90%, again the highest among all, is a sign that recall and precision are being carefully balanced.

The ROC-AUC score of 0.93 of the hybrid model indicates better discriminative power, i.e., it can differentiate between positive and negative cases well. XGBoost (0.90) and Random Forest (0.88) also indicate good values, but slightly less.

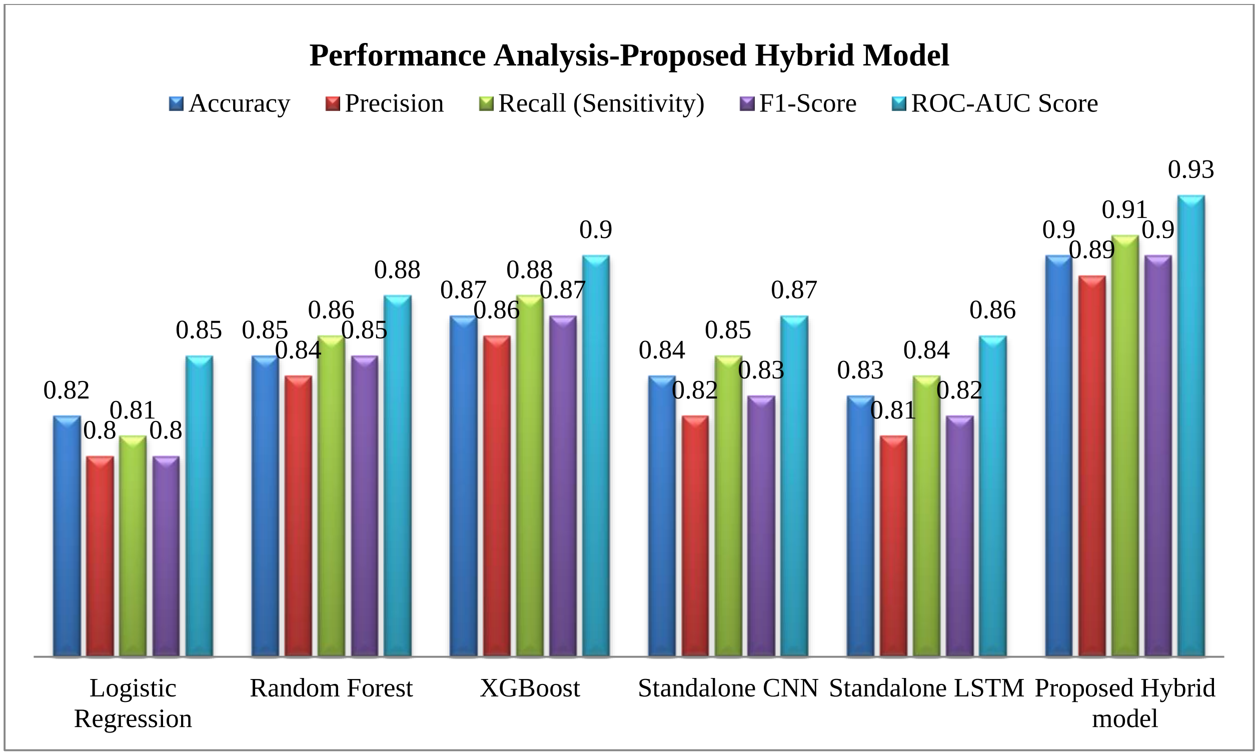

Fig. 3 Performance analysis – proposed hybrid model

From Figure 3, it is found that the LR method shows moderate predictive power in this study. It provides an accuracy of 0.82, with precision and recall values of 0.80 and 0.81, respectively. Its ROC-AUC value of 0.85 attests to a reasonable capacity to separate cases of HD from non-disease cases. It lags behind the more sophisticated models, however, and thus testifies to the weakness of linear decision boundaries in handling complex medical data.

Random Forest model performs much better with the benefit of ensemble learning via decision aggregation over multiple decision trees. It obtained a recall of 0.86, an accuracy of 0.85, and a precision of 0.84. With an ROC-AUC of 0.88, it shows good classification ability and balanced performance, and hence it is a good traditional machine learning model for HD detection.

XGBoost even surpasses Random Forest with a 0.87 accuracy and ROC-AUC of 0.90. Precision and recall are both higher than 0.86 and have an excellent ability to classify HD correctly, as well as have high sensitivity at the same time. This explains the strength of gradient boosting in detecting complex patterns in the data.

The CNN model, which was trained on tabular clinical information, has very good performance with an accuracy of 0.84 and an ROC-AUC of 0.87. It is better at generalizing than Logistic Regression, but not quite as good as XGBoost in terms of precision and F1-score. Its greatest strength is in learning spatial hierarchies within structured input features, but it has very bad single-model performance.

Table 5. Performance analysis with different hyperparameter settings –proposed hybrid Method

Method	Hyperparameter Setting	Accuracy	Precision	Recall (Sensitivity)	F1-Score	ROC-AUC Score
Proposed Hybrid Model	CNN (4 conv layers) + LSTM (2 layers, 64 units)	0.90	0.89	0.91	0.90	0.93
Proposed Hybrid Model	CNN (3 conv layers, 32 filters) + Bi-LSTM (64 units)	0.91	0.90	0.92	0.91	0.94
Proposed Hybrid Model	CNN (5 conv layers, 64 filters) + LSTM (128 units)	0.89	0.88	0.90	0.89	0.92

The LSTM model also performs similarly to CNN, but with a lower accuracy of 0.83 and ROC-AUC of 0.86. Its precision and recall indicate that it can learn sequential dependencies quite well. Lack of spatial feature extraction limits its independent performance in this case.

The optimal overall performance is achieved by the hybrid model that integrates CNN, LSTM, and an attention mechanism. With an accuracy of 0.90, a recall of 0.91, and an ROC-AUC score of 0.93, it outperforms all other models. The integration leverages the local pattern capture ability of CNN, the temporal relationship modeling ability of LSTM, and the dynamic feature weighting ability of the attention mechanism to develop a strong, interpretable, and highly accurate predictive system for HD detection.

Table 5 shows a comparative performance evaluation of various configurations of the proposed hybrid deep learning model for predicting HD, with emphasis on the impact of changing CNN and LSTM architectures on classification metrics.

CNN (4 conv layers) + LSTM (2 layers, 64 units): This setup provides robust performance with 0.90 accuracy and ROC-AUC of 0.93, which reflects a well-balanced model. The four convolution layers enable the network to learn very detailed local information, and the two LSTM layers function to learn sequential patterns. With 0.89 precision and 0.91 recall, this combination is highly sensitive with good generalization.

CNN (3 conv layers, 32 filters) + Bi-LSTM (64 units): This model shows an accuracy rate of 0.91, precision of 0.90, recall of 0.92, and ROC-AUC of 0.94. The bidirectional LSTM improves the learning capability of the model both from the past and future contexts, which enhances the temporal representation of features. The use of a slightly smaller number of CNN layers and filters prevents overfitting, leading to an optimal feature learning-generalization trade-off.

CNN (5 conv layers, 64 filters) + LSTM (128 units): This architecture indicates comparatively low performance with accuracy 0.89 and ROC-AUC 0.92. Perhaps the deeper CNN and the high LSTM layer introduced too much model complexity and resulted in overfitting with small clinical data. However, its recall value of 0.90 ensures that it is still performing well in the sense of identifying true positives.

The second setup (CNN with 3 conv layers and Bi-LSTM) is the best, achieving a balance of accuracy, precision, and interpretability. It highlights the advantage of architectural optimization and the strength of bidirectional temporal modeling in clinical prediction tasks.

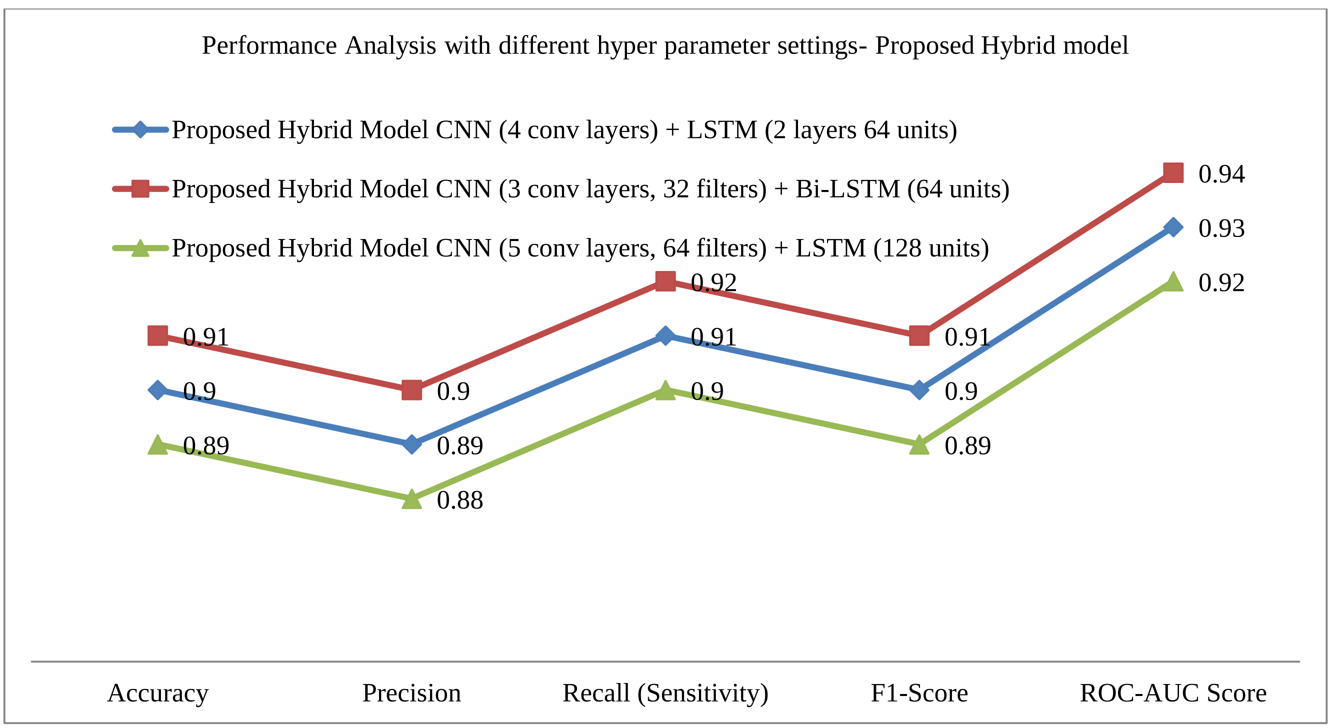

Fig. 4 Performance analysis with different hyperparameter settings

Figure 4 shows how well three varied configurations of an envisioned hybrid deep learning model mixing CNN and LSTM (or Bi-LSTM) layers perform. Among the versions, the version with 3 convolutional layers (32 filters) and 1 Bi-LSTM layer (64 units) performed better in all the performance metrics, acquiring the best performances in Recall (0.92), F1-

Score (0.91), and ROC-AUC Score (0.94), accompanied by competitive Accuracy (0.91) and Precision (0.90). This indicates that adding a Bi-LSTM layer instead of a regular LSTM improves the model's capacity to learn sequential dependencies, leading to improved classification performance. The 4 convolutional layers and 2-layer LSTM (64 units) model also showed robust performance, albeit slightly less than the Bi-LSTM model. On the other hand, the setting of 5 convolutional layers (64 filters) with a deeper LSTM (128 units) produced the worst performance in all the metrics, suggesting that more depth and complexity do not necessarily mean superior results and might contribute to the model's worse generalization. In general, the results demonstrate the strong performance of a mid-depth CNN combined with a Bi-LSTM for this classification task.

The better performance of the hybrid model is attributable to the synergy of CNN, LSTM, and attention mechanisms, which, when combined, capture the spatial and temporal features in the clinical data. The performance among the basic ML models that were tried, e.g., Logistic Regression, Random Forest, and XGBoost, despite being acceptable, did not model the sequential dependency or the complex interactions; the accuracy and recall were lower. The standalone CNN or LSTM models selectively capture spatial or temporal information, respectively, and fail to cohesively model them together. In this sense, the hybrid model supports a bidirectional LSTM and an attention mechanism that together use past and future sequences, giving variable importance weights to different features, which can enhance the hybrid model's sensitivity and discrimination. This enables it to detect complex patterns in patient data that traditional process models or componentwise DL systems potentially miss. Compared to studies reported to date, which have only used either CNN-LSTM or attention-based models, we report a higher ROC-AUC (0.94), recall (0.92), and precision (0.90), and emphasize this architectural itemization to report robustness in identifying positive cases of heart disease, while at the same time minimizing false negative cases. The middle ground of the architecture chosen gave it enough complexity to generalize appropriately to the clinical setting without overfitting to the training data (as deeper networks performed worse in this study). This study gave weight to the optimization of the architecture and its ability to employ a bidirectional temporal learning capability related to clinical prediction tasks.

5. Conclusion

The study presents a powerful hybrid DL approach to predicting HD that combines the feature engineering, the strength of CNNs, with the sequential learning, LSTM, and Bi-LSTM layers. Through controlled experimentation involving three model configurations, the study establishes that the architecture consisting of three convolutional layers (32 filters) followed by a Bi-LSTM layer (64 units) outperforms in predictive accuracy. This configuration works better when it comes to evaluation measures like ROC-AUC, F1-score, accuracy, and recall, indicating a balance of correctly classifying positive and negative HD instances. The more sophisticated models, like deeper stacks of CNN or larger LSTM units, performed worse, as would be expected by overfitting and the greater challenge in training. The retention of information by Bi-LSTMs from past and future time steps was helpful in extracting the sequence of clinical information, like patient history and time-series health records.

The findings highlight the value of architectural optimization over the simple scaling of depth or size of the model. The suggested hybrid model is important to apply in real life in medicine since it is a concrete and scalable addition to the early diagnosis and treatment of HD. Research in the future can explore the application of attention mechanisms, multimodal fusion of data modalities like images, blood work, and ECG, and explainability methods to enhance the usefulness and transparency of the model for professionals. In general, the research here creates a basis for advanced deep learning solutions for the reduction of the impact of cardiovascular disease through smart data-driven predictive frameworks.

References

[1] Saleh Mahdi Muhammed, Ghassan Abdul-Majeed, and Mahmoud Shuker Mahmoud, “Early Prediction of Chronic Heart Disease based on Electronic Triage Dataset by using Machine Learning,” *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, Al-Muthana, Iraq, pp. 131-136, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Elham Nasarian et al., “AI Framework for Early Diagnosis of Coronary Artery Disease: An Integration of Borderline SMOTE, Autoencoders and Convolutional Neural Networks Approach,” *arXiv preprint*, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Muhammad Raheel Khan et al., “Comprehensive Analysis of Cardiovascular Diseases: Symptoms, Diagnosis, and AI Innovations,” *Bioengineering*, vol. 11, no. 12, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Benjamin A. Satterfield, Ozan Dikilitas, and Iftikhar J. Kullo, “Leveraging the Electronic Health Record to Address the COVID-19 Pandemic,” *Mayo Clinic Proceedings*, vol. 96, no. 6, pp. 1592-1608. 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Serdar Bozyel et al., “Artificial Intelligence-Based Clinical Decision Support Systems in Cardiovascular Diseases,” *Anatolian Journal of Cardiology*, 28, no. 2, pp. 74-86, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rohit Bharti et al., "Prediction of Heart Disease using a Combination of Machine Learning and Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, pp. 1-11, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[7] Mohammad Farah, and Saad Al-Ahmadi, "WT-CNN: A Hybrid Machine Learning Model for Heart Disease Prediction," *Mathematics*, vol. 11, no. 22, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[8] Zakaria K. D. Alkayyali, Syahril Anuar Bin Idris, and Samy S. Abu-Naser, "A Systematic Literature Review of Deep and Machine Learning Algorithms in Cardiovascular Diseases Diagnosis," *Journal of Theoretical and Applied Information Technology*, 101, no. 4, pp. 1353-1365, 2023. [Google Scholar] [Publisher Link]
[9] Tawsifur Rahman et al., "Deep Learning Technique for Congenital Heart Disease Detection using Stacking-based CNN-LSTM Models from Fetal Echocardiogram: A Pilot Study," *IEEE Access*, vol. 11, pp. 110375-110390, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[10] R. Archana, and P.S. Eliahim Jeevaraj, "Deep Learning Models for Digital Image Processing: A Review," *Artificial Intelligence Review*, vol. 57, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[11] V.K. Sudha, and D. Kumar, "Hybrid CNN and LSTM Network for Heart Disease Prediction," *SN Computer Science*, vol. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[12] Muthu Ganesh Veerabaku et al., "Intelligent Bi-LSTM with Architecture Optimization for Heart Disease Prediction in WBAN through Optimal Channel Selection and Feature Selection," *Biomedicines*, vol. 14, no. 11, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[13] Md Maruf Hossain et al., "Cardiovascular Disease Identification using a Hybrid CNN-LSTM Model with Explainable AI," *Informatics in Medicine Unlocked*, vol. 42, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[14] Abdulwahab Ali Almazroi, "A Clinical Decision Support System for Heart Disease Prediction using Deep Learning," *IEEE Access*, vol. 11, pp. 61646-61659, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[15] Girish S. Bhavekar, and Agam Das Goswami, "A Hybrid Model for Heart Disease Prediction using Recurrent Neural Network and Long Short Term Memory," *International Journal of Information Technology*, vol. 14, pp. 1781-1789, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[16] Umesh Kumar Lilhore et al., "An Attention-Driven Hybrid Deep Neural Network for Enhanced Heart Disease Classification," *Expert Systems*, vol. 42, no. 2, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[17] Xin Shi et al., "Congestive Heart Failure Detection based on Attention Mechanism-Enabled Bi-Directional Long Short-Term Memory Model in the Internet of Medical Things," *Journal of Industrial Information Integration*, vol. 30, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[18] Mubarak Albathan, "Hybrid LSTM and Deep Belief Networks with Attention Mechanism for Accurate Heart Attack Data Analytics," *International Journal of Computer Science & Network Security*, vol. 24, no. 10, pp. 1-16, 2024. [Google Scholar] [Publisher Link]
[19] Ran Hao et al., "A Hybrid CNN-Transformer Model for Heart Disease Prediction Using Life History Data," *2025 8th International Symposium on Big Data and Applied Statistics (ISBDAS)*, Guangzhou, China, pp. 833-837, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[20] Prashant Kumar Shrivastava et al., "HCBiLSTM: A Hybrid Model for Predicting Heart Disease using CNN and BiLSTM Algorithms," *Measurement: Sensors*, vol. 25, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[21] Atta Ur Rahman et al., "Enhancing Heart Disease Prediction using a Self-Attention-Based Transformer Model," *Scientific Reports*, vol. 14, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[22] Yakub Kayode Saheed et al., "Modified Bi-Directional Long Short-Term Memory and Hyperparameter Tuning of Supervised Machine Learning Models for Cardiovascular Heart Disease Prediction in Mobile Cloud Environment," *Biomedical Signal Processing and Control*, vol. 94, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[23] B. Ramesh, and Kuruva Lakshmanna, "A Novel Early Detection and Prevention of Coronary Heart Disease Framework Using Hybrid Deep Learning Model and Neural Fuzzy Inference System," *IEEE Access*, vol. 12, pp. 26683-26695, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[24] Dhadkan Shrestha, "Advanced Machine Learning Techniques for Predicting Heart Disease: A Comparative Analysis Using the Cleveland Heart Disease Dataset," *Applied Medical Informatics*, vol. 46, no. 3, pp. 91-102, 2024. [Google Scholar] [Publisher Link]
[25] Padmakumari Pitchal, Shanthi Ponnusamy, and Vidivelli Soundararajan, "Heart Disease Prediction: Improved Quantum Convolutional Neural Network and Enhanced Features," *Expert Systems with Applications*, vol. 249, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[26] D. Cenitta et al., "Ischemic Heart Disease Prognosis: A Hybrid Residual Attention-Enhanced LSTM Model," *IEEE Access*, vol. 13, pp. 4281-4289, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[27] Li Li, Xi Chen, and Sanjun Hu, "Application of an End-to-End Model with Self-Attention Mechanism in Cardiac Disease Prediction," *Frontiers in Physiology*, vol. 14, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]