

Deep Learning Models Performance Analysis for Cardiovascular Disease Using an ECG Based Dataset.

Saroj Kumari^{1*}, Meena Chaudhary², Raghav Mehra³

¹Research Scholar, Mangalayatan University, Aligarh, UP, India.

Email: saroj.cse10@gmail.com

²Professor, Mangalayatan University, Aligarh, UP, India.

Email: meena.chaudhary@mangalayatan.edu.in

³Professor, AIML, AIT-CSE, Chandigarh University, Mohali, Punjab, India.

Email: raghav.mehrain@gmail.com

Received: 26th Oct, 2025; Revised: 24th Dec, 2025; Accepted: 14th Jan, 2026; Available Online: 15th Feb, 2026

ABSTRACT

Because CVD is among the leading causes of death worldwide, early risk prediction is particularly important. This work employs a mixed dataset and differentiates between three types of variables, clinical/lifestyle, and ECG-based ones, aiming at analyzing four DLMs namely- MLP, LSTM, CNN, and ViT. The case for the relevance to PTB Using the ECG portion of the PTB-XL and combining it with UCI Heart Disease features; this resulted in 22 multimodal predictors down the first level. We evaluated the model performance using accuracy, precision, recall, F1-score, and AUC. The results indicate that all architectures perform well (Accuracy ≥ 0.97 , AUC ≥ 0.99). LSTM obtained the optimal overall balance (Accuracy = 0.99, Recall = 0.98, F1 = 0.98), though CNN had slightly lower recall, (Recall = 0.97), and ViT reached perfect accuracy (1.00) with slightly lower recall (0.90). SVM joint analysis of three modalities as an example, the hybrid method has shown the feasibility of deep learning in early CVD diagnosis and decision support by means of boosting robustness and clinical relevance as compared to single-modality studies.

Keywords: Cardiovascular Disease, Electrocardiogram (ECG), Deep Learning, Hybrid dataset, Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Vision Transformer (ViT).

How to cite this article: Kumari S, Chaudhary M, Mehra R, Deep Learning Models Performance Analysis for Cardiovascular Disease Using an ECG Based Dataset...Int J Drug Deliv Technol. 2026; 16(2): 87-97; DOI: 10.25258/ijddt.16.2.12

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

As a key diagnostic tool, the electrocardiogram (ECG) is commonly used in clinical practice because it assesses the electric activity of the heart and is non-invasive^{1, 2}. Its application involving a variety of cardiovascular disease makes it a cornerstone in the identification and treatment of heart diseases³. Nevertheless, the manual interpretation of ECG signals is usually takes a lot of time as it depends on human resources and is vulnerable to inter-observer variability and human errors⁴. This has motivated the exploration of automated ECG analysis techniques, particularly those leveraging the capabilities of machine learning (ML) and deep learning (DL)⁵. Recently, heart diseases (more popularly cardiovascular disease (CVD)) have become the primary cause of death worldwide⁶. With more than 18 million fatalities annually roughly 32% of all deaths globally CVD (as shown in Figure 1) remains the leading cause of mortality globally⁷. With prompt diagnosis and treatment, a sizable percentage of these fatalities could be avoided. Therefore, the medical and data science community have made it a top priority to reliably forecast CVD risk prior to the commencement of key events, such myocardial infarction.

Models like Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Vision Transformers (ViTs) are now capable to outperform than the traditional techniques in a variety of biomedical tasks because to recent developments in DL architectures. While LSTMs are well-suited for modelling sequential dependencies in physiological signals across time⁵, CNNs have demonstrated great accuracy in identifying arrhythmias and myocardial infarctions by learning local patterns in ECG data⁸ With their capacity to simulate long-range feature dependencies through self-attention mechanisms, ViTs which were first proposed for image recognition tasks⁹ have recently been applied to tabular and time-series medical data, providing enhanced performance¹⁰. However, the bulk of earlier studies on ECG-based CVD prediction focused mostly on examining the raw waveform data or, occasionally, a limited set of clinical factors. While these approaches sometimes work, they often miss the complexity and heterogeneity of cardiovascular disease. In fact, electrophysiological perturbations observed in ECG are far from being the only source for CVD. Instead, it's a multifactorial condition that is impacted by several factors such as demographics (age, sex), behavioral (drinking, smoking) and lifestyle (physical

activity) patterns, comorbidity (diabetes, obesity, hypertension).

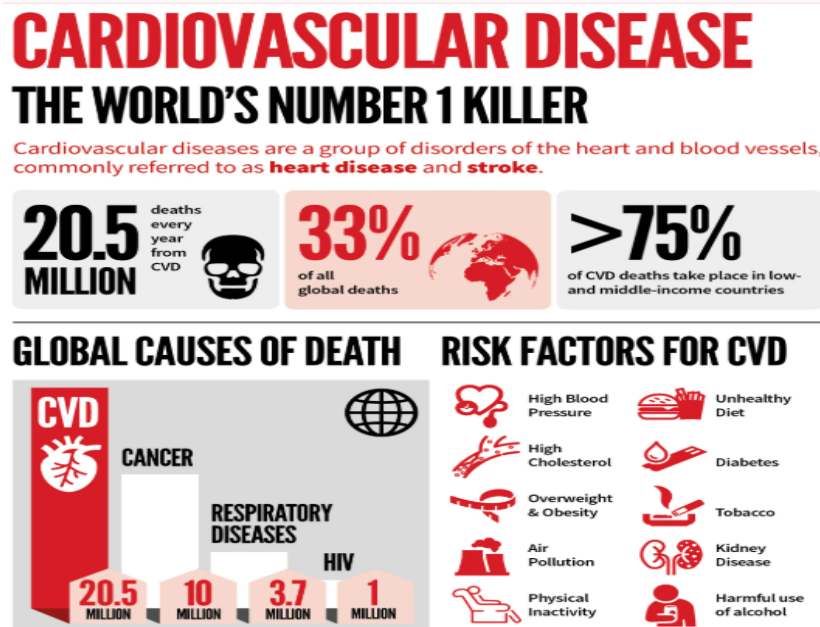


Figure 1: Global Overview of CVD and its Risk Factors¹¹

For instance, an ECG can diagnose electrical abnormalities, such as arrhythmias or ST-segment elevation, but it cannot autonomously diagnose risk factors that are attributable to metabolic or lifestyle components. As per research, the early identification and classification of CVD risk, lifestyle markers such as BMI, level of physical activity, dietary habits, sleep duration, and stress could be significantly possible^{12,13}. Similarly, comorbidities such as diabetes or chronic renal disease and genetic risk factors such as family members with heart disease often interact to accelerate the progression of the disease and impact patients' prognosis¹⁴. As a result, the models which rely solely on the ECG signals or the lack of clinical indicators may either be overfitted to some data patterns or lack of generality over large patient populations. As such, it is hypothesized that hybrid predictive models that integrate these disjoint factors into a single model will be well suited for learning from both static contextual (eg., demographic) and temporal physiological patterns using deep learning algorithms. This comprehensive approach allows for better modelling of patient heterogeneity and as well, improves the interpretability and clinical utility of predictions in applied settings.

Furthermore, time series such as ECG signals, tabular clinical information and categorical lifestyle features can all be combined along with one another through deep learning series such as CNNs, LSTMs, or Transformers. Recent evidence has shown that these multimodal fusion methods systematically improve prediction accuracy, especially in complex disorder such as coronary artery disease or heart failure^{15,16}. Taking this broader perspective, we propose to develop a hybrid dataset that integrates key clinical and lifestyle factors with ECG derived dynamic information to bridge the gap between signal-driven and context-aware

modelling in the current study. We propose a comprehensive comparison on the empirical performance of four types of deep learning models namely-MLP, LSTM, CNN, and ViT trained and tested on hybrid dataset, which includes clinical and lifestyle features in addition to ECG-based features to enable reliable prediction of cardiovascular risks.

LITERATURE SURVEY

In addition, heart diseases or CVD is still the principal reason of death globally by the findings of World Health Organisation (WHO) and is taking millions of lives per year annually (Figure 1)⁷. With the growing amount of healthcare data available from clinical, wearable, and lifestyle-related platforms, researchers can now examine complex predictive models for early CVD identification and prevention. Traditional approaches for risk scoring models like Framingham Risk Score¹⁷ that predicated on clinical factors such as age, cholesterol level, blood pressure and status of smoking, however, often suffer from lack of personalisation and inability to capture evolving physiologic status. Early diagnosis and prediction are vital for risk reduction and preventive measures against CVD. It is studies that several computational models have been proposed for modelling and prediction of cardiovascular events from medical data in history.

There are also several works using structured clinical data such as UCI Heart Disease dataset which consist of among other properties age, sex, cholesterol level, resting blood pressure, exercise induced angina and chest pain type¹⁸. Using this dataset to tell beforehand the presence of CVD has been attempted and accomplished by classical machine learning methods. For instance, on the Cleveland subset of the UCI data, **Gudadhe et al.**¹⁹ observed that Support Vector Machine (SVM) and decision tree-based models

could achieve high accuracy. These approaches are restricted by feature engineering and linear or independent assumptions between variables, and they often only consider a limited number of features. The advent of large, easily accessible ECG databases set like PTB-XL⁴ made deep learning models to directly learn patterns from raw ECG signals. In ECG waveforms, CNNs and LSTMs are especially good at identifying temporal and spatial relationships. **Strodthoff et al.**²⁰ achieved state-of-the-art results in multi-label ECG classification on the dataset of PTB-XL using CNNs. Likewise; **Raghunath et al.**²¹ reported that deep learning model can detect subtle signs of LF from raw ECGs even for asymptomatic patients. While ECG models' performance is high, they often omit contextual information that is important for the CVD risk prediction at the individual level, e.g., demographics, comorbidities, life style. Recent work has shown that adding signal-based data to the clinical features has improved the generalisation and interpretability of the models. To improve prediction accuracy of CVD, **Zhang et al.**²², they suggested a multimodal model using neural network and ECG waveforms and clinical demography features. This approach is consistent with the real-world situation, where lifestyle factors (every day smoking, alcohol consumption), metabolic markers (diabetes, body mass index) and genetic disorders have a major effect on CVD risk and are not only determined by the ECG patterns. Moreover, to classify subjects with HF, **Ali et al.**²³ introduced a hybrid model that was built by integrating EHR data and variables obtained from ECG. Their results support the notion that multimodal input performs better than single-modality models. Originally developed for use in image classification⁸, Vision Transformers (ViTs) have recently been modified for use with sequential and time-series biological data. By applying self-attention methods and segmenting input data into patches (or tokens), ViTs enable the model to capture long-range dependencies. ViTs have been applied to ECG signals in cardiac research by considering segments as 1D patches²⁴. ViTs show potential performance and robustness, especially when paired with clinical metadata, even so research is yet in its infancy stages²⁵⁻²⁷. There is little research combining clinical data, ECG signals, and lifestyle information into a unified predictive framework, despite the fact that each modality provides distinct insights. The majority of earlier models either concentrate on structured clinical data or ECG waveform classification. Few researches investigate how deep learning model performance can be improved by using a fully hybrid dataset, such one built from PTB-XL ECG signals and enhanced with clinical and lifestyle variables from datasets like UCI. The potential of deep multimodal learning was highlighted by recent research by **Lin et al.**²⁸, which showed that deep learning framework trained on 12-lead ECGs could predict cardiovascular mortality. By assessing and contrasting the performance of MLP, LSTM, CNN, and ViT models on a fused dataset intended to capture the intricacy of actual CVD risk variables; this study seeks to close that gap.

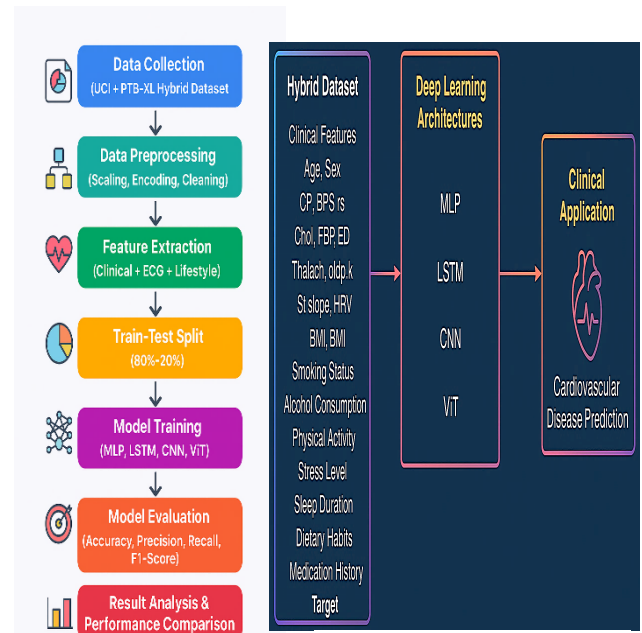


Figure 2 (a): Architecture and Performance Evaluation

Figure 2 (b): Training of Hybrid Dataset on DL Models for CVD

RESEARCH METHODOLOGY

In order to assess the effectiveness of many architectures like-MLP, LSTM, CNN, and ViT for CVD prediction, we created a hybrid dataset and put a deep learning pipeline into place. The four primary parts of the methodological framework are data composition, Pre-processing of data, Model Selection and architecture, and Evaluation as shown in Figure 2.

Dataset Composition

By combining complementing data from two publicly accessible sources namely- PTB-XL ECG dataset²⁹ and the UCI Heart Disease

dataset¹⁸, 1375 patient records were created in order to create a hybrid dataset that is indicative of actual cardiovascular disease with 22 features.

While the UCI dataset contributed crucial demographic and clinical variables like sex, age, blood pressure, cholesterol level, and other known CVD predictors, the PTB-XL dataset offered comprehensive ECG-derived features like heart rate variability, ST-segment characteristics, and resting ECG outcomes. In order to harmonise patient profiles across both datasets, data records were merged based on similar or common characteristics (e.g., age, sex, and ST slope). This allowed for the creation of enriched, multifactorial examples that accurately depict the intricate aetiology of cardiovascular disorders^{14, 15}. A more thorough

assessment of deep learning models on diverse inputs from the signal, clinical, and lifestyle domains was made possible by this fused dataset. The dataset integrates features from three distinct categories as provided in Table1.

Table 1: Description of Features

Category	Feature	Description
Clinical Features: These features are commonly found in real-world Electronic Health Record (EHR) systems and are frequently used in CVD diagnostics	Age	Age of Patient
	Sex	Gender (1 = Male, 0 = Female)
	Cp	Type of Chest Pain (1–4)
	BPS resting	Resting Blood Pressure
	Chol	Serum Cholesterol (mg/dl)
	FBP	Fasting Blood Sugar (>120 mg/dl)
	Rest ECG	Resting ECG Outcomes
	Thalach	Maximum Heart Rate Achieved
	oldpeak	ST depression
	ST slope	Slope of ST segment
Lifestyle Aspects Cardiovascular risk is greatly influenced by lifestyle decisions and behaviours.	HRV	Heart Rate Variability (ms)
	BMI	Body Mass Index
	Smoking Status	1 = Smoker, 0 = Non-Smoker
	Alcohol Consumption	1 = Regular, 0 = None
	Physical Activity	Level (1= Sedentary, 2= Moderate, 3= Active)
	Stress Level	1= Low, 2 = Medium, 3= High
Sleep Duration	Average Hours Per Day	

	Dietary Habits	1 = Poor, 2 = Average, 3 = Good
	SpO2 %	Blood Oxygen Saturation
	Respiratory Rate	Breaths Per Minute
	Medication History	1 = Yes, 0 = No
Target Variable The risk of CVD was represented by a binary outcome variable	Target	Heart disease presence (1 = Yes, 0 = No)

Pre-processing of Data

To guarantee consistent model training dynamics, Using Min-Max Scaling all numerical features normalized to [0,1] range. One-hot encoding has been used for categorical data like sex, smoking status, and physical activity. In order to replicate clinical data inconsistencies, missing values (if any) were simulated and imputed using the mode for categorical features and the median values for numerical ones. To assess model performance in a controlled and repeatable way, the complete dataset was portioned into

training (80%), validation (10%), and testing (10%) subsets at random.

Deep Learning Models Architecture

We used the TensorFlow / Keras and PyTorch frameworks to implement four cutting-edge DL models namely-MLP, LSTM, CNN, and ViT in order to assess the predictive power of various neural architectures for CVD diagnosis. To guarantee a fair and consistent comparison, the same dataset used to train each model. Description of every model is given below in Table2:

Table 2: DL Model & Architectural Framework

Model	Architecture
Multilayer Perceptron (MLP) It is a common feed-forward design appropriate for tabular data, served as the baseline classifier ³⁰	The network was made up of: Input: 25 neurones make up the input layer, which corresponds to the input features. Three fully connected dense layers with 128, 64, and 32 neurones each make up the hidden layers. Activation Function: ReLU (Rectified Linear Unit), activation function applied for all hidden layers, this activation function adds non-linearity. Dropout Rate: To reduce overfitting, dropout layers are applied at a rate of 0.3 after each dense layer. Output: The output layer consists of a single neurone that produces a binary classification output (0 or 1) by sigmoid activation. When temporal correlations are not dominating, the MLP design works well for learning intricate feature interactions ³¹ .

<p>Long Short Memory (LSTM) Long-range dependencies in sequential data are intended to be captured by LSTM networks, a specialised type of recurrent neural networks (RNNs). Despite being primarily static tabular data, the dataset can benefit from temporal modelling because ECG-derived metrics like heart rate variability and ST slope show time-dependent behaviour ³².</p>	<p>The components of the LSTM architecture were: Input: The input was moulded into time-series sequences with fixed-length windows using certain features that were extracted from the ECG. LSTM Layers: Two 64-unit stacked LSTM layers that can recognise temporal changes. Dropout and Recurring Dropout: To enhance generalisation and avoid overfitting, set at 0.2. Dense layers: For binary classification, a last dense layer with sigmoid activation was employed.</p> <p>In order to replicate how a doctor could interpret ECG patterns over time, this architecture was created to mimic both short- and long-term temporal dependencies ³³.</p>
<p>Convolution Neural Network (CNN) In this case, we used a 1D CNN to process ECG-derived features and extract localised patterns across feature sequences, while CNNs are usually applied to picture or time-series data. This is comparable to filtering ECG waveforms to identify signal segments that</p>	<p>The CNN architecture consisted of: 1D Convolutional Layers: ReLU activation, kernel size of 3, and two convolutional layers with 32 and 64 filters, respectively. Max Pooling: To minimise feature dimensionality and computation, a pooling layer of size two was employed. Dense layers and flattening: Before arriving at the output sigmoid neurone, the output was flattened and moved through layers that were all fully linked. Dropout: To prevent overfitting, 0.3 was used.</p> <p>CNNs were thought to be useful for analysing small feature groupings, particularly input related to ECGs, and they are particularly good at collecting local correlations ³⁶.</p>

<p>point to anomalous cardiac events. 34, 35</p>	
<p>Vision Transformer (ViT) A more recent development, the ViT architecture uses the transformer encoder mechanism, which was first created for natural language processing, to analyse structured and visual input ⁸. ViT models global inter-feature connections without the need of recurrence or convolution, in contrast to CNNs or LSTMs ³⁷.</p>	<p>The architecture of ViT includes: Patching input: Every patient record (i.e., 25 features) was divided into embeddings or fixed-size "patches" (e.g., 5 features per patch). Transformer encoder layers: To develop deep contextual representations, feed-forward and multi-head self-attention layers were included. Positional encoding: Used to keep feature embeddings in order. Output: For binary classification, the final class token embedding was run through a fully linked layer. ViT is especially effective in heterogeneous datasets because it provides a non-sequential, attention-based method that may capture intricate feature dependencies ³⁸.</p>
<p>Model Optimization Procedure The Binary Cross-Entropy loss- function, was used to train all models. We used the Adam optimiser^{39, 40} with a 32 batch size and 0.001 learning rate. To prevent overfitting, a validation-based checkpoint mechanism and early halting with</p>	$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$ <p>Where, y_i is the true label and \hat{y}_i is the predicted probability for sample i</p> <p>Grid search was used for hyperparameter tweaking over the following to guarantee peak performance: The rate of learning is [0.01, 0.001, 0.0001] Rates of dropout: [0.2, 0.3, 0.5] Size of batch: [16, 32, 64] The number of units or layers varies depending on the model design.</p>

patience of 10 epochs were employed. A maximum of 100 epochs were used to train each model.

RESULTS AND EVALUATION

On the hybrid ECG, clinical and lifestyle dataset, the experimental results show that all four DL models MLP, LSTM, CNN, and ViT performed remarkably well, attaining high accuracy, precision, recall, F1-score, and AUC values (≥ 0.94 across all measures), as indicated in Table 3 and Figure 3. This illustrates how well the dataset was designed and how well multimodal features were integrated for the prediction of CVD.

Table 3: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-score	AUC
MLP	0.97	0.96	0.92	0.94	0.99
LSTM	0.99	0.97	0.98	0.98	0.99
CNN	0.98	0.94	0.97	0.95	0.99
ViT	0.98	1.00	0.90	0.95	1.00

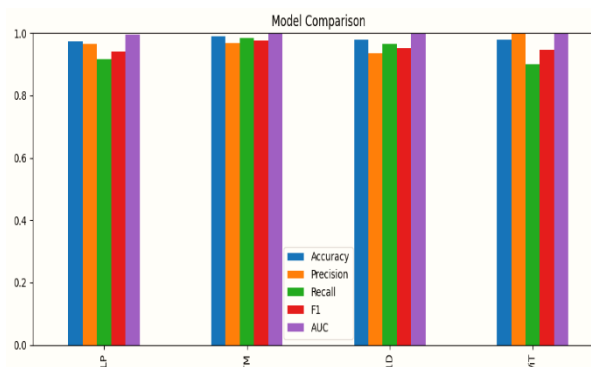


Figure 3: Model Performance Chart

With an F1-score of 0.94 and an accuracy of 97%, the Multilayer Perceptron (MLP) demonstrated its capacity to understand intricate feature relationships. But compared to other models, its recall (0.92) was marginally poorer, indicating that MLP might overlook a small number of positive instances (patients at risk for CVD). At 99% accuracy, 0.98 F1-score, and 0.99 AUC, the LSTM network produced the best overall performance. Its high sensitivity in identifying CVD cases is indicated by its superior recall (0.98), which makes it particularly appropriate for clinical settings where reducing false negatives is crucial. This is a demonstration of the capability of LSTM for predict sequential dependencies by temporal parameters derived from ECG, such as heart rate variability. It was shown that the CNN had a good performance in solving the problem to extracting the local feature patterns from the ECG data and clinical indicators fast, which leads to an accuracy of 98% and a recall of 0.97. Strong diagnostic sensitivity is ensured by its high recall, while a little inclination towards false positives is shown by its somewhat poorer precision (0.94) when compared to LSTM. The Vision Transformer (ViT) recognised actual positive CVD patients flawlessly and without false alarms, demonstrating exceptional precision (1.00).

Its significantly poorer recall (0.90) suggests that it failed to notice a tiny percentage of positive cases. However, ViT represent good discriminative capacity across thresholds with an AUC of 1.00, which makes it a viable option for practical implementation. All things considered, the LSTM model performed the best because it balanced all measures, especially its high recall and F1-score. Although they both performed competitively, CNN and ViT displayed minor recall and precision trade-offs.

The findings demonstrate that deep learning models trained on hybrid datasets that combine clinical, lifestyle, and ECG variables can outperform conventional machine learning techniques in CVD risk assessment, achieving nearly flawless predictive accuracy, as documented in earlier research. The superiority of these models is demonstrated when contrasted with previous research as provided in Table 4.

Table 4: Comparison with Earlier Studies

Study	Year	Dataset Used	Model Applied	Performance	Best Model
Kachue et al. ⁴¹	2018	MIT-BIH Arrhythmia (ECG)	CNN (Transfer Learning)	Accuracy = 0.86	CNN
Yildirim et al. ³³	2018	PhysioNet Dataset	BiLSTM + Wavelet Features	Accuracy = 0.86	BiLSTM
Oh et al. ⁴²	2018	ECG Data (Variable Length Beats)	CNN + LSTM	F1 Score = 0.93	CNN + LSTM
Giraldina et al. ²⁵	2024	ECG + cardiac MRI + clinical features	Self-supervised multimodal learning and distillation	Balanced accuracy ↑ +3.5%	Learns cross-modal representation
Nam et al. ²⁷	2024	Printed + ECG images	VizEC GNet with multimodal attention and distillation	Precision ↑ +3.5%, Recall ↑	Novel architecture handles ECG
Our Study	2025	Hybrid dataset: ECG-derived + clinical + lifestyle (n ≈ 1375)	MLP, LSTM, CNN, ViT (multimodal fusion)	Accuracy: 0.97–0.99; F1: 0.94–0.98; AUC: 0.99–1.0	LSTM best balance; ViT perfect precision; hybrid data boost performance.

CONCLUSION

The results of this investigation are very important from a clinical standpoint. The prediction models are guaranteed to reflect the complex character of cardiovascular disease (CVD) by the integration of clinical, ECG-derived, and lifestyle factors. Our hybrid dataset reflects real-world patient records, where cardiologists must simultaneously take into account biological, behavioural, and lifestyle risk factors, in contrast to previous research that solely relied on ECG signals or restricted clinical criteria. Because it may reduce false negatives, the LSTM model which has 99% accuracy and 0.98 recall is very useful in a clinical situation. In actuality, failing to notice a high-risk patient could result in a potentially fatal incident, such a myocardial infarction that goes unnoticed. Cardiologists may find that LSTM-based systems serve as trustworthy early warning tools by guaranteeing the identification of almost all true positive

CVD cases, which could lead to better outcomes through prompt therapies.

In contrast, the CNN and MLP models show strong and understandable performance with balanced F1-scores (0.94–0.95) and accuracies of 97–98%. These models have the potential to be more computationally efficient than LSTM and ViT, which makes them attractive options for use in wearable technology or mobile health applications where real-time prediction and resource efficiency are crucial. Together, these findings imply that clinical decision support systems (CDSS) in hospitals and telemedicine platforms may use deep learning models trained on hybrid datasets. Such systems could assist in prioritising high-risk individuals for additional diagnostic tests (such as echocardiography or angiography) and longitudinally monitoring lower-risk patients by offering risk classification of patients. This study shows that combining clinical data (such as age, blood pressure, cholesterol, etc.), ECG-derived features

(such as ST slope, HRV, Thalach, Oldpeak, etc.), and lifestyle factors (such as smoking, drinking, stress, sleep, etc.) into a single predictive framework produces better results than using data from multiple sources. The complex nature of cardiovascular health is well captured by the hybrid dataset, which makes the derived models both accurate and therapeutically meaningful.

The findings support the notion that deep learning architectures, particularly LSTM and ViT, are capable of producing extremely accurate predictions for the risk stratification of CVD. Such models can be integrated into healthcare systems to facilitate proactive intervention methods, individualised risk assessment, and early detection by utilising hybrid datasets. This will ultimately lower the morbidity and mortality rates linked to cardiovascular disease.

FUTURE WORK

While the current study demonstrates the potential of hybrid deep learning models integrating ECG, clinical, and lifestyle features for accurate cardiovascular disease (CVD) prediction, there remain several promising directions for future research and practical enhancement:

Extension to Large-Scale, Multi-Center Datasets: A hybrid dataset of 1,375 records combined from PTB-XL and UCI sources was used in this analysis. To evaluate the generalisability of the suggested models across populations, ethnic groups, and healthcare systems, future research should validate them on larger, multi-center, and demographically diverse datasets. Such validation would allow domain adaptation for global CVD risk prediction and assist in identifying biases resulting from sampling, class imbalance, or regional health disparities.

Interpretable and Explanatory Deep Learning (XAD): Black-box models, such as LSTM and ViT, can present interpretability issues in clinical settings despite their high predictive accuracy. To visualise how features like blood pressure, HRV, and stress contribute to prediction outcomes, future research should use Explainable AI frameworks like Grad-CAM, SHAP (SHapley Additive exPlanations), or attention heatmaps. In addition to facilitating model integration into clinical decision support systems (CDSS), such interpretability can increase physician trust.

Attention Optimisation and Multimodal Fusion: Independent deep learning architectures were used in the current study for comparative analysis. Creating hybrid or ensemble architectures, like CNN-LSTM, LSTM-Transformer, or ViT-attention fusion networks, to collaboratively learn spatial, temporal, and contextual dependencies would be a logical next step. The interpretability and resilience of multimodal fusion may be further improved by sophisticated attention mechanisms (such as transformer encoders and cross-modal attention)

REFERENCE

1. Friedman DJ, Green MA, Peterson SL. Principles of electrocardiographic interpretation. 3rd ed. New York: McGraw-Hill; 2025.

2. Hong S, Zhou J, Shang Z, Liu Y. Opportunities and challenges of deep learning methods for ECG analysis. *IEEE Rev Biomed Eng.* 2020;13:204–217.

3. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation.* 2000;101(23):e215–e220.

4. Nechita M, Popescu M, Ionescu L, Marinescu R. Challenges in manual ECG interpretation: A review of human factors and diagnostic variability. *J Electrocardiol Cardiac Inform.* 2024;61(2):123–131.

5. Su P, Zhang Y, Chen M, Wang L. Cardiovascular disease: The global burden and recent trends. *J Med Syst.* 2023;47(1):1–10.

6. World Health Organization. Cardiovascular diseases (CVDs). WHO; 2021. Available from: [https://www.who.int/news-room/factsheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/factsheets/detail/cardiovascular-diseases-(cvds))

7. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. *Comput Methods Programs Biomed.* 2018;161:1–13.

8. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. *Proc Int Conf Learn Representations (ICLR).* 2021.

9. Wang H, Wang Z, Chen Y, et al. TransTab: Learning with tabular data using transformers. *Proc Adv Neural Inf Process Syst (NeurIPS).* 2022.

10. World Heart Federation. Infographic on physical activity and cardiovascular disease. World Heart Federation; 2025. Available from: <https://world-heart-federation.org/resource/infographic-physical-activity/>

11. Gaziano JM. Reducing the growing burden of cardiovascular disease in the developing world. *Health Aff.* 2007;26(1):13–24.

12. Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *Lancet.* 2004;364(9438):937–952.

13. Bots ML, Peters SA, Woodward M. The role of comorbidities in cardiovascular disease. *Eur J Prev Cardiol.* 2015;22(2 Suppl):21–28.

14. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Deep learning for ECG classification: The PhysioNet/Computing in Cardiology Challenge 2020. *PhysioNet Challenge.* 2020.

15. Zhou JT, Du J, Li H, et al. Multimodal representation learning for clinical prediction. *ACM Trans Knowl Discov Data.* 2021;15(5):1–26.

16. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J. Factors of risk in the development of coronary heart disease: Six-year follow-up experience. *Ann Intern Med.* 1961;55(1):33–50.
17. Wilson PWF, Kannel WB, D’Agostino RB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97(18):1837–1847.
18. Detrano R, Janosi A, Steinbrunn W, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am J Cardiol.* 1989;64(5):304–310.
19. Gudadhe M, Wankhade S, Dongre S. Decision support system for heart disease based on support vector machine and artificial neural network. *Proc Int Conf Comput Commun Technol (ICCCT).* 2010:741–745.
20. Strodthoff N, Wagner P, Schaeffter D, Samek W. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *Physiol Meas.* 2020;41(10):104003.
21. Raghunath S, Ulloa-Cerón E, et al. Deep neural networks can predict new-onset atrial fibrillation from the electrocardiogram and enhance its diagnostic utility. *Circulation.* 2022;145(5):398–408.
22. Zhang Y, Li X, Chen M, et al. A multimodal deep learning model for early detection of heart disease using ECG and clinical data. *IEEE Access.* 2020;8:49161–49170.
23. Ali M, Khan S, et al. Hybrid predictive model for early detection of heart failure using EHR and ECG signals. *Comput Biol Med.* 2022;142.
24. Fan A, Zhang H, et al. ECG-ViT: Learning longitudinal ECG representations using vision transformer for cardiovascular risk stratification. *Proc IEEE Eng Med Biol Soc (EMBC).* 2022.
25. Giralda E, Liu J, Basha A. Self-supervised multimodal representation learning for cardiovascular disease prediction. *Proc AAAI Conf Artif Intell.* 2024;38(13):14441–14449.
26. Mohsen A, Shah A. ECG-DiaNet: Multimodal ECG and clinical risk factor integration for systemic disease prediction. *Sensors.* 2025;25(3):678.
27. Nam Y, Kim HJ, Choi J. VizECGNet: Vision transformer with multimodal knowledge distillation for printed ECG analysis. *IEEE J Biomed Health Inform.* 2024;28(5):2403–2415.
28. Lin Y, Chen Z, Huang J, et al. Deep learning-based survival prediction for cardiovascular mortality using 12-lead ECGs: Development and validation of ECG-Surv. *Lancet Digit Health.* 2024;6(1):e34–e44.
29. Wagner P, Strodthoff N, Bousseljot RD, et al. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data.* 2020;7(1):1–15.
30. Haykin S. *Neural networks and learning machines.* 3rd ed. Upper Saddle River: Pearson; 2009.
31. Badirli A, Klambauer G, et al. Exploring the power of KANs: Overcoming MLP limitations in complex data analysis. *Appl Comput Eng.* 2024.
32. Faust O, Acharya UR, Molinari R, Acharya RB. Automated detection of atrial fibrillation using long short-term memory network with ECG signals. *Comput Biol Med.* 2018;102:327–335.
33. Yildirim O, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med.* 2018;102:411–420.
34. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–444.
35. Acharya UR, Raghavendra U, Tan JH, Hagiwara Y, Sudarshan VK, Koh JEW. Automated characterization of coronary artery disease and myocardial infarction using decomposition and deep convolutional neural network. *Pattern Recognit Lett.* 2017;94:274–282.
36. Zhang Z, Chen C, Li W, Wang J. Interpretable deep learning for automatic diagnosis of 12-lead ECG signals. *IEEE Trans Instrum Meas.* 2021;70:1–12.
37. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. *Proc IEEE Int Conf Comput Vis (ICCV).* 2021:10012–10022.
38. Raghu M, Zhang C, Kleinberg J, Bengio S. Vision transformers for medical image analysis. *arXiv.* 2021;arXiv:2106.10270.
39. Kingma DP, Ba J. Adam: A method for stochastic optimization. *Proc Int Conf Learn Representations (ICLR).* 2015.
40. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. *Nat Med.* 2019;25(1):65–69.
41. Kachuee M, Fazeli S, Sarrafzadeh M. ECG heartbeat classification: A deep transferable representation. *Proc IEEE Int Conf Healthc Inform (ICHI).* 2018:443–444.
42. Oh SL, Ng EYK, Tan RS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med.* 2018;102:278–287