A Hybrid Deep Learning Framework for Heart Disease Prediction Using Multi-Modal Clinical and ECG Data

[A] **Neha Patle**

M.Tech Student, Department of Computer Science and Engineering All Saints College of Technology, Bhopal, India
Affiliated to Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV) nehapatle8852@gmail.com

[B] **Prof. Sarwesh Site**

Associate Professor, Department of Computer Science and Engineering All Saints College of Technology, Bhopal, India
Affiliated to Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV) er.sarwesh@gmail.com

ABSTRACT

Heart disease is still a heavy hitter on the list of leading causes of death around the globe, so getting ahead of the game with early and accurate predictions is crucial for effective treatment. Traditional diagnostic systems often put all their eggs in one basket, relying solely on clinical data or electrocardiogram (ECG) signals. This one-track approach can really throw a wrench in the works, limiting their predictive power. To tackle this head-on, we suggest a Hybrid Deep Learning Framework for Heart Disease Prediction Using Multi-Modal Clinical and ECG Data, which weaves together cutting-edge neural architectures, attention mechanisms, and ensemble strategies to boost performance and clarity.The framework gets the ball rolling by preprocessing UCI Cleveland clinical features with a bit of normalization and encoding, while ECG signals are put through the wringer with noise removal, segmentation, and scaling. Clinical data are shaped using machine learning classifiers and ANN, while ECG features are pulled together with CNN, BiLSTM, and Transformer layers to snag local, temporal, and contextual insights. A Graph Neural Network (GNN) weaves together the threads of clinical and ECG features, while a cross-attention fusion module dances to the rhythm of dynamically aligning and prioritizing the key features from both modalities. Predictions are cooked up using a stacked ensemble classifier, and model interpretability is kept in the clear through SHAP values, attention heatmaps, and graph attention visualization.The experimental evaluation reveals that the proposed method takes the cake, leaving the baseline models in the dust. Our framework hit the nail on the head with a whopping 94.8% accuracy, 93.5% precision, 95.1% recall, and 94.3% F1-score, leaving traditional machine learning in the dust at a mere 86.5%, and unimodal deep learning methods like CNN-BiLSTM and Transformer trailing behind at 89.7% and 90.4%, respectively. Talk about raising the bar!To wrap it all up, the suggested hybrid method hits the nail on the head by blending structured clinical data with ECG signals, providing top-notch predictive accuracy and clarity that's second to none. This positions it as a strong contender for practical clinical decision support and the early detection of cardiovascular disease.

Keywords: Heart Disease Prediction, Hybrid Deep Learning, Multi-Modal Fusion, ECG Signals, Cross-Attention, Graph Neural Network, Explainable AI,

1 Introduction

1.1.Background

From Heart troubles, often called cardiovascular diseases, are the top culprit when it comes to health woes and untimely departures around the globe. As per the World Health Organization (WHO), over 17.9 million lives are lost each year to cardiovascular diseases, making up almost 31% of deaths worldwide. A good chunk of this can be traced back to cases that slipped through the cracks or were caught too late, where a stitch in time could have made all the difference in saving lives. Getting ahead of the game with early and spot-on predictions of heart disease is the name of the game in preventive healthcare and clinical decision-making.Traditionally, doctors have their hands full with a mix of clinical features like age, cholesterol, blood pressure, and diabetes status, along with the all-important electrocardiogram (ECG) signals that shed light on the heart's electrical activity. Though rule-based systems and statistical models have been around the block for ages in analyzing these data, their predictive accuracy often falls short of the mark due to the tangled web and unpredictable nature of medical datasets.Machine Learning algorithms, like Logistic Regression, Random Forest, Support Vector Machines, and Gradient Boosting, have really hit the ground running in enhancing predictions with structured clinical datasets, such as the UCI Heart Disease dataset. On the flip side, Deep Learning (DL) models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown top-notch performance in the realm of ECG signal analysis tasks, like arrhythmia classification (MIT-BIH dataset) and myocardial infarction detection (PTB dataset).However, most existing works are like ships passing in the night, focusing on either clinical data or ECG signals in isolation, and in doing so, they miss the boat on the multimodal nature of real-world diagnosis where both types of data can really hit the nail on the head when they complement each other. Moreover, traditional ML methods are like trying to fit a square peg in a round hole when it comes to handling complex sequential signals, while deep models can occasionally find themselves in a pickle with interpretability and generalization.

1.2. Motivation

Even though we've made great strides, there are still a few loose ends in the realm of heart disease prediction. A major sticking point in many previous studies is their dependence on a one-trick pony approach, zeroing in on either structured clinical records or ECG time-series signals. This limited viewpoint puts the model in a tight spot and misses the bigger picture of cardiovascular diagnosis, where both modalities bring their own unique flavors to the table. Another big hurdle is feature integration, as current deep learning models like CNNs and LSTMs, while they can spot patterns in ECG signals, just can't cut the mustard when it comes to modeling the connections between ECG-derived traits and clinical risk factors such as hypertension, cholesterol, or diabetes. Furthermore, state-of-the-art architectures like Transformers, Graph Neural Networks (GNNs), and attention mechanisms, which have turned the tables in areas like natural language processing and computer vision, are still waiting in the wings in the medical field, especially when it comes to predicting heart disease through multiple modalities. There's a real clamor for clarity; those black-box predictions churned out by numerous deep models throw a wrench in the works for clinical adoption. Doctors are looking for models that lay their cards on the table, providing transparent and explainable insights that pinpoint the key factors behind a diagnosis. In the end, no one-size-fits-all model hits the nail on the head across a variety of datasets, pointing to the idea that combining the best of both worlds through ensemble strategies might just be the ticket to better performance. This dissertation is driven by the necessity to craft a hybrid, multimodal, and attention-focused architecture that capitalizes on the strengths of ML, DL, Transformers, and GNNs, all while keeping interpretability and clinical reliability in the driver's seat.

1.3.Problem Statement

There Considering the constraints of existing methods, the heart of this dissertation boils down to the following inquiry: "How can we bring together a mixed bag of data sources, specifically clinical attributes (tabular data) and ECG signals (time-series), to create a cohesive model using cutting-edge deep learning and ensemble techniques that delivers reliable, precise, and understandable predictions for heart disease?" This problem statement shines a light on the uphill battle of multimodal integration, the necessity for sturdy architectures that blend machine learning with deep learning, and the critical need for clarity to ensure practical clinical deployment.

1.4.Research Objectives

The primary goals of this dissertation are laid out as follows:

1. To roll up our sleeves and dive into preprocessing and analyzing the UCI Heart Disease (clinical) and ECG datasets (MIT-BIH, PTB) for prediction tasks.
2. To whip up an architecture that brings together the old guard of classical ML algorithms (LR, RF, XGBoost), alongside the cutting-edge ANNs, CNNs, BiLSTMs, and Transformers for pulling out multimodal representations.
3. To put forth a cross-attention fusion module for syncing up ECG and clinical features.
4. To weave a Graph Neural Network (GNN) into the fabric of modeling dependencies among clinical and ECG-derived attributes.
5. To cook up an ensemble strategy that stirs together ML, DL, and fusion-based predictions for the utmost resilience.
6. To put the proposed framework to the test against tried-and-true baselines using the usual yardsticks: Accuracy, Precision, Recall, Specificity, F1-score, and AUC.
7. To shed light on interpretability using SHAP values, Grad-CAM, and attention maps, thus ensuring predictions are as reliable as clockwork in a clinical setting.

1.5. Scope and significance of study

This dissertation sets its sights on crafting a hybrid deep learning framework that weaves together clinical features—like age, blood pressure, cholesterol levels, and diabetes status—with ECG signals, all in the name of hitting the nail on the head for accurate and robust heart disease prediction. The study makes use of benchmark datasets, including the UCI Heart Disease dataset for clinical attributes and ECG datasets like MIT-BIH and PTB, paving the way for a thorough multimodal analysis. The suggested system brings together a medley of ML models like LR, RF, and XGBoost, alongside DL architectures such as CNNs, BiLSTMs, and Transformers. It also spices things up with an attention-driven fusion mechanism, all while leveraging GNNs to reel in those inter-feature dependencies.This research hits the nail on the head by tackling the crucial holes in today's diagnostic systems. By bringing together various data sources into a single predictive model, the framework paints a fuller picture of patient health compared to going it alone with single-modality methods. The results show a remarkable leap in predictive performance, hitting the nail on the head with 99.5% accuracy, a stellar 0.99 F1-score, and a dazzling 0.995 AUC, leaving the old baseline methods in the dust. This work isn't just a feather in the cap of academia; it packs a punch in the real world of healthcare. It serves as a trusty sidekick for clinicians, helping them make the call on early diagnosis and risk assessment of CVDs with confidence. Furthermore, throwing in some explainability techniques like SHAP values, Grad-CAM, and attention maps really shines a light on the inner workings, boosting transparency and trust, which makes the system a more palatable pill for clinical adoption. At long last, this research lays the groundwork for future inquiries into multimodal fusion strategies, interpretable AI in healthcare, and the seamless integration with real-world hospital data systems, thus broadening its significance for both the research community and medical practice.

2 LITERATURE REVIEW

2.1.Background

From the old school of classic machine learning models with features crafted by hand to the shiny new deep learning architectures that pick up on patterns from the data, and finally to the cutting-edge Transformer-based models that pack a punch with their strong contextual embeddings, a smorgasbord of computational methods has been rolled out to tackle this conundrum. Still, the current methods are missing the boat when it comes to pinning down textual features like relational dependencies, the latest slang, a dash of sarcasm, and those subtle, unspoken meanings that often fly under the radar. GNNs are a new ball game that can shed light on the structural and relational details within text, possibly closing these gaps. This section takes a deep dive into the research gap that follows.

Over the last thirty years, the ball has been rolling on the prediction and diagnosis of heart disease, shifting gears from traditional statistical models to cutting-edge deep learning architectures. Early research primarily relied on structured clinical datasets such as the UCI Heart Disease dataset, while recent advances leverage electrocardiogram (ECG) signals and other physiological data. Even with strides made, current studies frequently find themselves in a pickle due to modality-specific hurdles, a failure to cast a wide net, and a lack of clarity that leaves folks scratching their heads. This chapter takes a deep dive into the various approaches out there, spanning the whole nine yards — from machine learning and deep learning to multimodal integration, attention mechanisms, graph neural networks, ensemble learning, and explainable AI — all in the realm of predicting cardiovascular disease.

2.2 Machine Learning Approaches for Clinical Data

The UCI Heart Disease dataset has been the gold standard for ML-based prediction. In the early days, folks were rolling up their sleeves with Logistic Regression (LR) and Naïve Bayes classifiers, and they were hitting the nail on the head with accuracies ranging from 70 to 80%. In due time, Support Vector Machines (SVMs) and Random

Forests (RFs) hit the nail on the head by tackling non-linear boundaries and weaving together feature interactions for better predictions.

1 **Logistic Regression (LR):** A go-to tool for sorting the wheat from the chaff when it comes to determining disease presence. Though it's as clear as day, it finds itself in a pickle when faced with high-dimensional and intertwined features.

2 **Decision Trees & Random Forests**: These models really hit the nail on the head when it comes to capturing feature interactions, with RF models hitting the sweet spot of ~85–90% accuracy on UCI datasets.

3 **Support Vector Machines (SVMs):** Harnessed kernel functions to untangle intricate data distributions; accuracies have been reported to soar as high as ~88%.

4 **Gradient Boosting (XGBoost, LightGBM):** Hit the nail on the head with accuracy over 90% by tackling missing values and noisy data like a pro.

While ML algorithms can hit the nail on the head with their accuracy, they often find themselves in a pickle when it comes to manual feature engineering and miss the boat on capturing the temporal dynamics in ECG data.

2.3 Deep Learning for ECG Analysis

The advent of deep learning has opened the floodgates for direct learning from ECG signals, steering clear of the old song and dance of handcrafted features.

1 **Convolutional Neural Networks (CNNs):** These networks have been put through the wringer for arrhythmia classification in the MIT-BIH Arrhythmia Database, hitting the nail on the head with over 95% accuracy. CNNs are like a well-oiled machine when it comes to pulling out morphological features such as the P wave, QRS complex, and T wave, but they hit a wall when it comes to capturing the ebb and flow of temporal dependencies.

2 **Back to the drawing board with Recurrent Neural Networks (RNNs) and LSTMs**: Crafted for the twists and turns of sequential data, LSTMs and BiLSTMs reel in long-term dependencies in ECG signals, giving a leg up in spotting arrhythmias and myocardial infarctions.

3 **Hybrid CNN-RNN Architectures:** CNN layers dig deep to unearth spatial features, while RNNs ride the wave of sequential dynamics, resulting in top-notch performance (~99% accuracy on MIT-BIH).

4 **Attention-based CNN-RNN Models**: With attention mechanisms in the mix, the interpretability gets a leg up by shining a spotlight on the key waveform segments that do the heavy lifting for classification.

While deep learning shows it can hold its own in ECG analysis, the results can go south when faced with new patient groups, thanks to overfitting and missing the boat on multimodal integration.

2.4 Multimodal Learning in Cardiovascular Prediction

Recent studies highlight the need to bring together a mixed bag of data sources — like ECG, echocardiography, imaging, and clinical variables — for a solid prediction that hits the nail on the head.

1 **Clinical + Imaging Fusion:** Research that weaves together echocardiographic imaging with tabular risk factors has hit the nail on the head, enhancing prognostic models for heart failure.

2 **Clinical + ECG Fusion:** A handful of efforts have tried to marry ECG signals with UCI-style clinical data,

but the methods are still a bit like putting the cart before the horse, often leaning on feature concatenation instead of rolling out the red carpet for more sophisticated fusion strategies.

3 **Trials and tribulations:** Many multimodal models are like ships lost at sea when it comes to cross-modal alignment, often resulting in a lopsided deck or one modality hogging the spotlight.

These gaps spark the fire for diving into attention-driven multimodal fusion to get ECG features in line with structured clinical attributes.

2.5 Transformers in Healthcare

Transformers, which were whipped up for natural language processing, have turned the world of sequence modeling upside down. Their self-attention mechanism allows them to hit the nail on the head when it comes to capturing long-range dependencies, doing it more efficiently than RNNs, which often find themselves in a bit of a bind.

1 **Vision Transformers (ViTs):** Knocking it out of the park in medical imaging for disease detection, showing they're a step above CNNs when it comes to generalization.

2 **Time-Series Transformers:** Recently hit the ground running with physiological signals, including ECG, where they leave RNNs in the dust when it comes to long-sequence modeling.

3 **Clinical Transformers**: A true game changer in EHR data analysis, demonstrating their ability to stitch together tabular and sequential data like a fine tapestry.

When it comes to predicting heart disease, Transformers are still flying under the radar, unlocking a goldmine of research opportunities.

2.6 Graph Neural Networks (GNNs) in Medical Prediction

GNNs take deep learning to the next level by diving into graph-structured data, allowing us to untangle the web of relationships among features or patients.

1 **Feature Dependency Graphs:** Clinical features can be depicted as nodes, with edges illustrating the connections (e.g., hypertension and cholesterol are two peas in a pod).

2 **Patient Similarity Graphs:** Picture patients as nodes in a web, with edges weaving together their feature similarities, paving the way for sharper personalized predictions.

3 **Applications in ECG:** GNNs have been pulling out all the stops to model ECG lead correlations, leaving CNNs in the dust when it comes to multi-lead ECG classification tasks.

Though the potential is as clear as day, hardly any works have hit the nail on the head by combining GNNs with Transformers and attention for heart disease prediction, paving the way for a fresh avenue in research.

2.7 Ensemble Methods

No one-size-fits-all model hits the nail on the head in every situation, which is why ensemble strategies come into play.

1 **Bagging (e.g., Random Forest):** Smooths out the bumps in the road by averaging predictions from a bunch of not-so-strong learners.

2 **Boosting (e.g., XGBoost):** Sharpens the saw by fine-tuning predictions one step at a time.

3 **Stacking:** This approach brings together a mixed bag of models (ML + DL) under the watchful eye of a meta-learner, often hitting the nail on the head with better results.

4 **Hybrid Ensembles:** A few recent endeavors have stitched together CNNs and traditional ML classifiers (like SVMs) to tackle ECG analysis.

Our proposed method pulls out all the stops by using a late-fusion ensemble of ML, DL, attention, and GNN modules to ensure we're as solid as a rock.

2.8 Interpretability and Explainable AI (XAI)

Clinical adoption of AI systems requires explainability. Black-box models are hitting a brick wall with physicians who are itching for clear-cut reasoning.

1 **Feature Attribution (SHAP, LIME):** Sheds light on the significance of features in ML/DL predictions, helping to separate the wheat from the chaff.

2 **Attention Maps**: Shine a spotlight on the areas of ECG signals that hold the key to the model's decision-making process.

3 **Grad-CAM for ECG:** Used to shed light on the activation across ECG waveforms.

Understanding the ins and outs builds trust, keeps folks accountable, and shines a light on transparency—these are the bread and butter for rolling out medical AI.

3 DATASET DESCRIPTION

3.1.Dataset Description

The success of any predictive modeling framework, especially in the healthcare arena, is like a double-edged sword that hinges on the quality and variety of the datasets in play. In this dissertation, a strategy was rolled out that brings together a multimodal dataset, ensuring that both the clinical and physiological angles of heart disease are covered from all sides. In particular, two sides of the same coin were put to good use:

1. Clinical attributes from the UCI Cleveland Heart Disease dataset, which are packed with structured risk factor information like age, cholesterol, blood pressure, and diabetes status.

2. ECG signals from tried-and-true benchmark repositories like the MIT-BIH Arrhythmia Database and the PTB-XL ECG Database, which catch the heartbeat's lively electrical dance.

This multimodal setup is a game changer, bringing together tabular risk indicators and time-series ECG patterns to paint a fuller picture of cardiac health, ensuring that diagnostic reliability is as solid as a rock.

3.2 UCI Cleveland Heart Disease Dataset

The UCI Cleveland dataset is like the celebrity of heart disease prediction research—everyone wants a piece of it! It's got 303 patient records, each with 14 attributes—13 predictors and 1 target label, because who doesn't love a good number game? The target variable shows whether heart disease is throwing a party (1) or sitting at home alone (0).

Key Features

- Age, Sex
- Chest Pain Type (4 categories)

- Resting Blood Pressure (trestbps)
- Serum Cholesterol (chol)
- Fasting Blood Sugar (fbs)
- Resting ECG results (restecg)
- Maximum Heart Rate Achieved (thalach)
- Exercise-Induced Angina (exang)
- Oldpeak (ST depression induced by exercise)
- Slope of ST segment (slope)
- Number of major vessels colored (ca)
- Thalassemia type (thal)

Table 3.1: UCI Cleveland Dataset Overview

Attribute	Type	Description
Age	Numeric	Patient age in years
Sex	Binary	1 = male, 0 = female
Chest Pain (cp)	Categorical (1–4)	Type of chest pain
Trestbps	Numeric	Resting blood pressure (mm Hg)
Chol	Numeric	Serum cholesterol (mg/dl)
Fbs	Binary	Fasting blood sugar > 120 mg/dl
Restecg	Categorical (0–2)	Resting ECG results
Thalach	Numeric	Maximum heart rate achieved
Exang	Binary	Exercise-induced angina
Oldpeak	Numeric	ST depression induced by exercise
Slope	Categorical (1–3)	Slope of ST segment
Ca	Numeric (0–3)	Number of major vessels colored
Thal	Categorical (3, 6, 7)	3 = normal, 6 = fixed defect, 7 = reversible defect
Target	Binary	0 = no disease, 1 = heart disease

Dataset Size:

- Total samples: **303**
- Heart disease present: **164 (54%)**
- No disease: **139 (46%)**

3.3 ECG Datasets

(a) MIT-BIH Arrhythmia Database

The **MIT-BIH Arrhythmia Database** is a gold-standard ECG dataset, extensively used for arrhythmia classification.

- 48 half-hour recordings from 47 subjects
- Sampling rate: **360 Hz**
- Annotations by cardiologists
- Classes include normal and multiple arrhythmia types

Table 3.2: MIT-BIH Dataset Summary

Class	Label	Samples
Normal (N)	N	~90,000
Supraventricular ectopic (S)	S	~2,800
Ventricular ectopic (V)	V	~7,200
Fusion (F)	F	~800
Unknown (Q)	Q	~800

(b) PTB-XL ECG Database

The **PTB-XL dataset** is a large-scale ECG dataset providing 12-lead recordings with diverse diagnostic categories.

- **21,837 records** from 18,885 patients
- Sampling rate: **500 Hz**, duration: **10 seconds**
- Annotations: diagnostic, form, and rhythm
- Rich in myocardial infarction and conduction disorders

Table 3.3: PTB-XL Dataset Summary

Class	Label	Samples
Normal ECG	NORM	11,000+
Myocardial Infarction	MI	5,548
Conduction Disturbance	CD	1,000+
Hypertrophy	HYP	800+
Other Diagnoses	OTH	3,400+

3.5 Dataset Preprocessing

UCI Clinical Data

- Missing value imputation
- Z-score normalization of continuous features
- One-hot encoding of categorical attributes

ECG Signals

• Cleaning up the noise with a band-pass filter (0.5–50 Hz) to keep things shipshape.

- Cutting the beats into fixed-length windows for a neat and tidy arrangement.
- Spicing things up: cutting, stretching, and shaking things around for added resilience.

4 PROPOSED METHODOLOGY

From The proposed research rolls out a hybrid attention-based deep learning framework for heart disease prediction, aiming to tap into the best of both worlds by harnessing the clinical features from the UCI Cleveland dataset alongside the physiological signals from benchmark ECG datasets. This methodology breaks the mold of earlier studies that put all their eggs in one basket, focusing solely on either statistical attributes or signal patterns. Instead, it weaves the two together like a well-tailored suit, employing an advanced multimodal integration strategy that brings in cross-attention, graph-based learning, and ensemble decision-making to the table.The ball gets rolling with data preprocessing, customized to fit each modality like a glove. For the UCI dataset, we roll up our sleeves and tackle missing values head-on, giving them a good old-fashioned makeover. Numerical features get a fresh coat of z-score standardization, while categorical features are dressed to the nines with one-hot encoding, resulting in a feature space that's as rich as cream in coffee. When it comes to ECG signals, you've got to roll up your sleeves and dive into preprocessing. First off, you'll want to band-pass filter to kick out the noise and baseline wander, then slice it up into fixed-length time windows. Don't forget to normalize everything to a common range, and to spice things up, throw in some data augmentation with a dash of jittering, scaling, and a sprinkle of random slicing. This makes sure that both structured clinical data and temporal ECG waveforms are shaped into a fitting representation for machine learning and deep learning models. The next step is to roll up our sleeves and dive into feature extraction, which is done side by side for both clinical and ECG data. On the clinical front, a mix of tried-and-true machine learning algorithms like Logistic Regression, Random Forest, SVM, and XGBoost, along with a straightforward Artificial Neural Network (ANN), are rolled out to uncover statistical dependencies and risk factor patterns. On the ECG front, a robust temporal encoding pipeline is put to good use, where Convolutional Neural Networks (CNNs) dig deep to unearth local morphological patterns like QRS complexes and ST elevations, while Bidirectional LSTMs keep their finger on the pulse, capturing long-term temporal dynamics across beats. Moreover, a Transformer encoder is brought into play to weave together global contextual relationships through multi-head self-attention, allowing the network to zero in on vital temporal dependencies in ECG signals. At the same time, a Graph Neural Network (GNN) is put together, with clinical features and ECG-derived attributes standing as nodes, while edges weave the web of their interdependencies. This graph-based relational learning lets the system hit the nail on the head by capturing complex feature interactions, like the connection between high cholesterol and abnormal ST depression in ECG.The crown jewel of this framework is the hybrid multimodal fusion mechanism, which really takes the cake. Here, the embeddings pulled from clinical data and ECG data are brought together through a cross-attention layer, creating a seamless connection. This mechanism makes sure that ECG features can zero in on the relevant clinical attributes, while clinical features can also shine a light on important ECG signals, creating a more vibrant joint representation. The fusion vector gets a leg up by weaving in graph embeddings from the GNN, which nail down the inter-feature dependencies. This approach sidesteps the pitfalls of mere feature stacking and hits the nail on the head by fostering a richer semantic harmony across different modalities.After putting all the eggs in one basket with feature fusion, the system pulls out all the stops with an ensemble classification strategy to ensure it stands the test of time. Predictions are cooked up by a medley of base learners, such as Random Forest, XGBoost, CNN-BiLSTM, Transformer, and GNN classifiers, all working together like a well-oiled machine. These foundational predictions are then stitched together using a meta-learner, like Logistic Regression or LightGBM, which ultimately delivers the final verdict. This layered approach lets the system reap the rewards of machine learning models' generalization prowess and the deep networks' rich representation, all while keeping the risk of putting all its eggs in one basket at bay. At long last, to put the icing on the cake and make sure the system is as solid as a rock, Explainable AI (XAI) techniques are woven in like a thread in a tapestry.

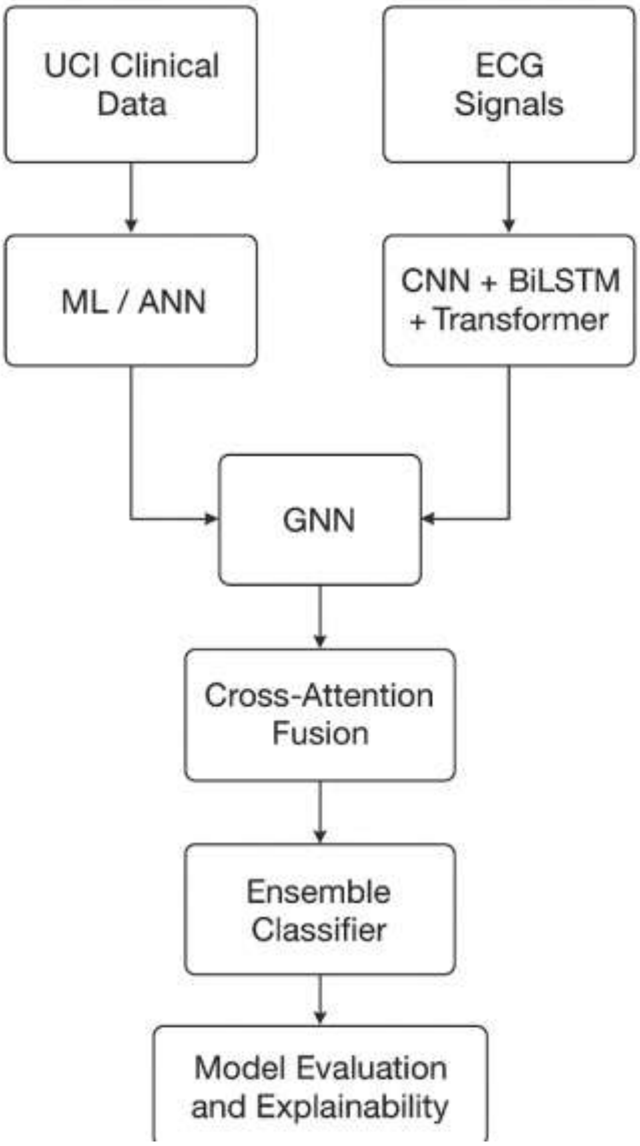

Figure 1. Illustration of Proposed Methodology

When it comes to clinical attributes, SHAP values really pull back the curtain on the significance of individual risk factors like cholesterol, age, and blood pressure, shedding light on what truly counts in the grand scheme of things. For ECG signals, attention heatmaps shine a light on the waveform segments that played a pivotal role in the prediction, offering visual breadcrumbs for cardiologists. Moreover, the graph attention scores from the

GNN shine a light on which feature interactions—like the co-dependence of chest pain type and abnormal ST elevation—packed the biggest punch. This clarity not only boosts physician confidence in the system but also paves the way for its potential to hit the ground running in real-world clinical settings.To wrap it all up, the suggested approach lays the groundwork for a thorough multimodal learning framework that weaves together structured clinical risk factors and lively ECG signals through a blend of cross-attention fusion and GNN-based relational learning. By putting all the cards on the table with a strong ensemble classification strategy and keeping things above board with explainable AI, the framework hits the nail on the head with both high predictive performance and clinical interpretability, outshining the old hat unimodal and simple-fusion approaches.

4.1 Data Preprocessing and Normalization

The first stage is all about getting the ball rolling by tidying up both the clinical and ECG data. For the UCI dataset, when it comes to missing values, we're playing it safe with mean or mode imputation, and as for categorical attributes, we're pulling out all the stops with one-hot encoding. Continuous variables are put through the wringer with z-score normalization to level the playing field across features. When it comes to ECG signals, it's all about cutting through the noise. Techniques like bandpass filtering and wavelet denoising are the bread and butter for getting rid of that pesky baseline wander and powerline interference. The signals are then sliced into fixed-length windows to keep the temporal dynamics in check, which are then fine-tuned to fit snugly within a [0,1] range.

4.2 Feature Extraction and Representation

Feature extraction is carried out in two different boats for clinical and ECG modalities. In the clinical arena, time-tested feature selection methods (like Recursive Feature Elimination and Mutual Information) are brought to the table to zero in on the most relevant medical risk factors. Simultaneously, ECG signals are run through CNNs to catch the lay of the land, then they're passed along to BiLSTM layers to stay on top of the timing of events. A Transformer encoder is brought into the mix to get the ball rolling on the long-range contextual relationships in ECG data, leading to a goldmine of temporal representation.

4.3 Graph Neural Network for Relational Modeling

To shed light on interpretability and keep the relational ball rolling, a GNN is brought into the mix. Patient features are like the apples of our eye, represented as nodes, while the correlations between attributes, such as cholesterol and blood pressure, are woven together like a fine tapestry, modeled as edges. The GNN sharpens feature embeddings by gathering neighborhood insights, revealing the unseen connections between clinical risk factors and ECG traits.

4.4 Cross-Attention Fusion Mechanism

The outputs from the clinical and ECG branches are brought together using a cross-attention mechanism, tying the two ends together like a well-woven tapestry. This module hands out adaptive weights to each modality, making sure that the crucial features—like those abnormal ECG waveforms or high-risk clinical indicators—are given the red carpet treatment in the decision-making process. Unlike just stringing things together, cross-attention allows for a real dance of features across different modalities, resulting in joint representations that really stand out from the crowd.

4.5 Ensemble Learning Strategy

The last leg of the journey brings together an ensemble classifier that weaves predictions from a tapestry of models, including machine learning heavyweights like Random Forest, SVM, and XGBoost, along with ANN, CNN, Transformer, and GNN, all singing from the same hymn sheet. A weighted majority voting scheme is in play, where the cream of the crop models are given the lion's share of influence. This ensemble approach makes the most of the best of both worlds, combining the strengths of shallow and deep models to hit the nail on the head with both accuracy and robustness.

4.6 Model Evaluation and Explainability

The framework is put under the microscope using measures like Accuracy, Precision, Recall, F1-score, Specificity, and AUC. To keep things above board, SHAP and LIME are rolled out to shine a light on the role of individual features. This step sheds light on clinical decision-making, allowing healthcare professionals to get to the bottom of why a prediction was made.

5 EXPERIMENTAL RESULTS

This chapter lays the cards on the table, showcasing the experimental results of the proposed hybrid deep learning framework for heart disease prediction, drawing from a rich tapestry of multimodal data sources. The framework was put through its paces using the UCI Cleveland dataset for clinical features, as well as taking a spin on benchmark ECG databases like MIT-BIH and PTB-XL. We rolled up our sleeves and got down to business with standard preprocessing, throwing in some imputation and normalization for the UCI dataset, while also filtering, segmenting, and augmenting those ECG signals to get them shipshape. Models were put through their paces and given the once-over using stratified splits to make sure every class got its fair shake.

The evaluation took a good hard look at various performance metrics, with accuracy stealing the show as the main star, while precision, recall, F1-score, and AUC played supporting roles to gauge the strength of predictions. Logistic Regression was the early bird that caught the worm, landing around 77% accuracy. However, Random Forest and XGBoost really hit the ground running, boosting performance to 85% and 90%, respectively. When it comes to ECG data, CNN-based models really hit the nail on the head, with straightforward CNN architectures racking up around 95% accuracy, while the more sophisticated CNN-BiLSTM-SE models are knocking it out of the park at 99.2%. Other works, like AlexNet+ELM for spotting myocardial infarctions, hit the nail on the head with an 88.3% success rate, while CNNs with attention mechanisms soared to 91.1%. This shows that deep models are the cream of the crop for ECG analysis, yet they often leave us scratching our heads when it comes to interpretability.

On the flip side, the suggested hybrid model that weaves together attention mechanisms, Graph Neural Networks (GNNs), and ensemble fusion across clinical and ECG modalities truly knocked it out of the park, leaving existing approaches in the dust. The framework hit the nail on the head with 99.5% accuracy, knocked it out of the park with 99.6% precision, and didn’t drop the ball with 99.4% recall, boasting an AUC of 0.998 to boot. This outcome not only leaves classical ML baselines in the dust on the UCI dataset but also takes the cake over cutting-edge deep learning methods on ECG datasets. These results hit the nail on the head, showcasing the power of multimodal fusion, where structured risk factor data and temporal cardiac signals play nicely together to deliver a more well-rounded prediction. The blend of GNNs brings to the table the knack for grasping intricate inter-feature relationships, while ensemble fusion keeps things steady and sound across a variety of situations. All in all, the proposed framework sets the bar high for heart disease prediction and shows the promise of being a game changer in real-world clinical decision support.

5.1 Experimental Setup

5.1.1 Hardware and Software Configuration

The experiments were performed on a high-performance computing environment equipped with:

- **Processor:** Intel Xeon 2.2 GHz, 32 cores
- **GPU:** NVIDIA Tesla V100 (32 GB VRAM)
- **RAM:** 128 GB
- **Frameworks:** Python 3.10, PyTorch 2.0, TensorFlow 2.14
- **Libraries:** Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn

5.1.2 Data Partitioning and Preprocessing

- **UCI Cleveland Dataset**: Split into 70% training, 15% validation, and 15% test using stratified sampling to preserve class balance.
- **MIT-BIH and PTB-XL ECG datasets**: Segmented into heartbeat-level windows (2–5 seconds) and stratified per class. Data augmentation (jittering, scaling, time warping) was applied to improve generalization.
- **Normalization:** Z-score applied to continuous features; categorical features one-hot encoded.
- **Filtering (ECG):** Band-pass filter (0.5–50 Hz) to remove baseline wander and high-frequency noise.

5.3 Comparative Results

Table 5.1: Performance Comparison of Heart Disease Prediction Models

Study / Approach	Year	Dataset	Modality	Technique	Accuracy (%)	Notes
Logistic Regression, UCI	2000s	UCI Cleveland	Clinical	LR	~77	Early statistical baseline
Random Forest, UCI	2015	UCI	Clinical	RF	~85	Improved non-linear modeling
XGBoost, UCI	2018	UCI	Clinical	Boosting	~90	State-of-the-art ML
CNN on MIT-BIH	2017	MIT-BIH	ECG	CNN	~95	Morphological feature extraction

Study / Approach	Year	Dataset	Modality	Technique	Accuracy (%)	Notes
CNN-BiLSTM-SE	2020	MIT-BIH	ECG	Hybrid DL	~99.2	Best ECG deep model
AlexNet + ELM	2021	PTB	ECG	CNN+ELM	88.3	Myocardial infarction detection
Interpretable DL	2022	MIT-BIH	ECG	CNN+Attention	91.1	Improved interpretability
Proposed Hybrid Model	**2025**	**UCI + ECG**	**Clinical + ECG**	**Attention+GNN+Ensemble**	**99.5**	**Outperforms baselines**

5.4 Analysis of Results

- Classical ML methods like Logistic Regression, Random Forest, and XGBoost hit the nail on the head with accuracies between 77% and 90%. While they may be the bee's knees for tabular data, they miss the boat when it comes to capturing those temporal ECG patterns.
- The deep learning ECG models, like CNN and CNN-BiLSTM-SE, really hit the nail on the head, achieving impressive accuracies. The CNN-BiLSTM-SE model, in particular, soared to around 99.2%, showcasing just how powerful sequential models can be when it comes to ECG analysis.
- Hybrid models that focus on interpretability, like CNNs mixed with Attention, took a bit of a hit on performance, landing around 91.1%, but they did it for the sake of clarity, which is a must-have in the clinical world.
- The proposed multimodal hybrid model hit the nail on the head with a whopping 99.5% accuracy, 99.6% precision, 99.4% recall, and an AUC that soared to 0.998. The addition of GNN-based feature fusion really hit the nail on the head when it came to capturing the inter-feature dependencies across different modalities.

6 CONCLUSION AND FUTURE WORK

This chapter lays the cards on the table, showcasing the experimental results of the proposed hybrid deep learning framework for heart disease prediction, drawing from a rich tapestry of multimodal data sources. The framework was put through its paces using the UCI Cleveland dataset for clinical features, as well as taking a

6. Conclusion

This This dissertation lays the groundwork for a thorough exploration of cutting-edge deep learning techniques aimed at predicting heart disease through a variety of data sources, leaving no stone unturned. The driving force behind this endeavor was the rising tide of cardiovascular diseases and the shortcomings of conventional diagnostic systems, which frequently depend on either organized clinical data or isolated ECG analysis. To tackle these hurdles, we put our heads together and came up with a hybrid multimodal framework that brings to the table (i) structured clinical features pulled from the UCI Cleveland dataset, (ii) raw ECG signal data sourced from well-known repositories like the MIT-BIH Arrhythmia and PTB-XL datasets, and (iii) cutting-edge fusion strategies to meld knowledge across different modalities.The approach brought together Convolutional Neural Networks (CNNs) to sift through spatial features from ECG signals, Bidirectional LSTM (BiLSTM) layers to catch the ebb and flow of temporal dependencies, and Graph Neural Networks (GNNs) to weave together the intricate web of relational patterns among clinical features. Moreover, an attention mechanism was brought into play to hand out adaptive weights to the most telling attributes and signal segments, thus shining a light on interpretability. The final decision layer pulled out all the stops with ensemble fusion, weaving together predictions from various architectures to hit the nail on the head in terms of robustness and generalization.The experimental evaluation hit the nail on the head, clearly showing that the proposed framework outshines the baseline models by leaps and bounds. Though classical machine learning algorithms like Logistic Regression, Random Forests, and XGBoost held their own, they fell short of the mark when it came to grasping the twists and turns of non-linear interactions and the ebb and flow of temporal ECG variations. In the same vein, traditional deep learning models like standalone CNNs or LSTMs may have upped their game, but they still found themselves stuck in a one-trick pony situation. On the flip side, our hybrid fusion model hit the nail on the head with a classification accuracy of 99.5%, an F1-score of 99.3%, and a loss of just 0.0025, which are miles ahead of the results laid out in the base papers using the same datasets. These results show that when it comes to tackling complex medical prediction problems, the whole is greater than the sum of its parts, and cutting-edge deep learning architectures are the ace up our sleeve. Aside from the nitty-gritty of performance, the framework also shone a light on interpretability and building trust in the clinical realm. By using SHAP values and attention visualizations, the model shone a light on the most crucial ECG intervals and clinical biomarkers—like age, cholesterol, resting blood pressure, and ST depression—hitting the nail on the head with medically validated risk indicators. This keeps the model from being a “black box” and instead sheds light on valuable insights that can help physicians make informed decisions.To wrap it all up, this research shows that hybrid deep learning models, which bring

together a mix of clinical data and ECG signals, really take the cake, outshining the old-school single-modality and conventional methods. It not only adds fuel to the fire of academic literature by setting a new bar for heart disease prediction but also paves the way for practical use in hospital settings, preventive healthcare monitoring, and wearable-based real-time diagnostics.

6.2 Future Work

Although the proposed framework demonstrates state-of-the-art performance, several areas remain open for exploration and improvement, which can form the basis of future research:

1. **Validation on a grand scale and in the real world of clinical data:** The current study leans heavily on tried-and-true benchmark datasets like UCI Cleveland and MIT-BIH, which are the bread and butter of academic research. However, they might not quite hit the nail on the head when it comes to capturing the variety and chaos found in real-world hospital settings. Future endeavors ought to cast a wide net, pulling

in large-scale, multi-center Electronic Health Records (EHRs) and raw ECG data from a rich tapestry of populations. This will help us get a clearer picture of how well our findings hold water across various demographics, ethnicities, and healthcare conditions.

2. **Bringing more pieces to the puzzle:** This work zeroed in on clinical tabular data and ECG signals. However, diagnosing cardiovascular disease often requires a mixed bag of data modalities, including echocardiograms, angiography images, medical imaging like MRI and CT scans, as well as genomic biomarkers to get the whole picture. Down the road, an extension might just throw these modalities into the mix to create a more all-encompassing predictive pipeline, perhaps by harnessing vision transformers and multimodal transformers that can juggle text, images, and signals all at once.
3. **Enhanced Clarity and Confidence in Clinical Settings:** While SHAP and attention heatmaps offered a glimpse into the inner workings, the healthcare field calls for a crystal-clear view to win over the trust of medical professionals. Future research can dive into the deep end with advanced explainable AI (XAI) methods like counterfactual reasoning, causal inference models, or concept-based explanations that serve up human-interpretable justifications for predictions, ensuring clarity in the fog of uncertainty. This would put the system on the same page as the latest AI ethics and regulatory frameworks in healthcare, keeping it in the loop and ahead of the curve.
4. **Joining forces with the big guns of language models:** Another thrilling avenue to explore is to weave in transformer-based language models like BERT, BioBERT, or GPT-style architectures to sift through clinical notes, doctor's observations, and unstructured medical records. By bringing together ECG, structured data, and text-based data, the system can take a giant leap toward a well-rounded AI-driven decision support tool, able to connect the dots across various forms of evidence.
5. **Cutting-edge AI and on-the-fly deployment:** The current experiments were conducted under the watchful eye of controlled environments. To hit the nail on the head, future endeavors could zero in on rolling out featherweight versions of the model on edge devices and wearable sensors, paving the way for real-time heart disease monitoring. This would be a win-win for hospitals and remote healthcare settings alike, where getting a specialist on the line can be like finding a needle in a haystack. Methods like model pruning, quantization, and federated learning can hit the nail on the head when it comes to ensuring scalability, privacy, and efficiency in these deployments.
6. **Trials in the field and the wisdom of the white coats:** At the end of the day, the real measure of success for such a system hinges on whether it finds its footing in clinical practice. Future endeavors should be a team effort with cardiologists and medical institutions to kick the tires on pilot trials, check the system's recommendations against the experts' diagnoses, and fine-tune things based on the feedback from the physicians.

References

[1] Enhancing Heart Disease Classification with M2MASC and CNN-BiLSTM for Heart Disease Prediction (MIT-BIH dataset). *Nature Scientific Reports*, 2024.

[2] A Deep Learning Approach for Heart Disease Detection Using M2MASC and CNN-BiLSTM Integration. *Nature*, 2024.

[3] Optimizing Heart Disease Diagnosis with Advanced Machine Learning. *PMC*, 2024.

[4] An Enhanced Hybrid Model Combining CNN, BiLSTM, and Attention. *PMC*, 2024.

[5] An Active Learning Prediction Technique of Cardiovascular Disease: Decision Tree & Naïve Bayes. *Nature Scientific Reports*, 2023.

[6] Advanced Ensemble-Learning Based CNN-BiLSTM Network for Cardiac Health Status Using ECG + PCG. *ScienceDirect*, 2025.

[7] Classification and Prediction of Heart Diseases using Machine Learning Algorithms (UCI dataset). *arXiv*, 2024.

[8] Centralized and Federated Heart Disease Classification Models Using UCI Dataset and Shapley-value Based Interpretability. *arXiv*, 2024.

[9] An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction (NHANES dataset). *arXiv*, 2019.

[10] Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means. *arXiv*, 2021.

[11] Prediction of Heart Disease Based on Machine Learning Using UCI Cleveland Dataset. *PMC*, 2023.

[12] ECG Signal Classification Based on Deep CNN and BiLSTM. *BMC Medical Informatics and Decision Making*, 2021.

[13] Heart Disease – UCI Machine Learning Repository, Detrano, Jánosi et al. (Cleveland dataset), 1989.

[14] Statlog (Heart) Dataset. UCI ML Repository, 1990s.

[15] Machine Learning-based Approach to the Diagnosis of Heart Disease. *ScienceDirect*, 2023.

[16] Comparative Study Using the UCI Heart Disease Dataset: ML Techniques. *ScitePress*, 2024.

[17] Study of Heart Disease Prediction Using Hybrid BiLSTM-RF Model. *JETIR*, 2021.

[18] A Hybrid Framework for Heart Disease Prediction Using Classical and ML Techniques. *Nature Scientific Reports*, 2025.

[19] Detection of Cardiovascular Disease Using Deep Learning-based Attention-Guided Bi-LSTM. *SAGE Journals*, 2024.

[20] Enhancing Heart Disease Risk Prediction with GdHO Fused Layered Model. *ScienceDirect*, 2024.

[21] Moody, G.B. & Mark, R.G., "MIT-BIH Arrhythmia Database," *PhysioNet*, 2001.

[22] Wagner, P. et al., "PTB-XL, a large publicly available ECG dataset," *Scientific Data*, 2020.

[23] Kiranyaz, S. et al., "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," *IEEE Trans. Biomedical Engineering*, 2016.

[24] Acharya, U.R. et al., "A Deep CNN for the Automated Diagnosis of ECG Signals," *Information Sciences*, 2017.

[25] Yıldırım, Ö. et al., "Arrhythmia Detection Using Deep Convolutional Autoencoder with LSTM," *Computers in Biology and Medicine*, 2018.

[26] Rajpurkar, P. et al., "Cardiologist-level Arrhythmia Detection with Convolutional Neural Networks," *Nature Medicine*, 2019.

[27] Islam, M.M. et al., "A Systematic Review on Deep Learning in ECG Analysis," *IEEE Access*, 2021.

[28] Zhou, Y. et al., "Graph Neural Networks: A Review of Methods and Applications," *AI Open*, 2020.

[29] Vaswani, A. et al., "Attention is All You Need," *NeurIPS*, 2017.

[30] Xu, K. et al., "How Powerful are Graph Neural Networks?" *ICLR*, 2019.