Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 15 April 2026
Revised 21 May 2026
Accepted 24 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 906–930

KEYWORDS:
Heart disease prediction, Causal inference, LiNGAM, Graph Attention Network, Large Language Model, Explainable AI, Clinical NLP, Multimodal learning, Biomedical graph, Risk stratification.

Causal-GAT-LLM: A Hybrid Framework for Explainable Heart Disease Prediction using LiNGAM and Language Models

Amit Kumar[1*], V. K Jain[2], Ajay Kumar[3]

[1] Research Scholar, Computer Science & Engineering, Teerthanker Mahaveer University, Moradabad (U.P.), India
[2] Vice Chancellor, Teerthanker Mahaveer University, Moradabad (U.P.), India
[3]Professor, Department of General Medicine, TMMC&RC, Teerthanker Mahaveer University, Moradabad (U.P.), India
Corresponding Author: amitakg84@gmail.com ORCID:https://orcid.org/0009-0004-7946-3119

ABSTRACT: Heart disease remains a leading cause of mortality worldwide, necessitating accurate and interpretable prediction systems capable of handling complex, multimodal healthcare data. This study proposes a novel hybrid framework Causal Graph Attention Network with Large Language Model and Linear Non-Gaussian Acyclic Model (Causal-GAT-LLM-LiNGAM) for robust and explainable heart disease prediction. The approach begins by applying the LiNGAM algorithm to structured clinical data to uncover directed causal relationships among variables such as blood pressure, cholesterol, and smoking history. These causal graphs form the backbone of a Graph Attention Network (GAT), which learns to emphasize the most influential features through attention mechanisms. Simultaneously, a domain-specific Large Language Model (LLM), such as BioBERT or GPT-based variants, is employed to extract semantic embeddings and causal cues from unstructured clinical text, including physician notes and discharge summaries. These text-derived features are integrated into the graph to enrich node and edge representations. By utilizing attention weights and LLM-generated narratives, the combined model not only improves predictive accuracy but also offers human-interpretable reasons for risk assessment. Evaluated on benchmark datasets such as MIMIC-III and UCI Heart Disease, the proposed framework outperforms traditional GNNs and black-box models in terms of accuracy, AUC, and clinical interpretability. This research demonstrates the synergistic potential of causal inference, graph learning, and large language models in developing next-generation diagnostic tools for personalized cardiovascular healthcare.

1. INTRODUCTION

Heart disease is a severe health issue of great concern to the population as well as the major cause of premature deaths among the global population killing millions annually. The heart structure and functioning may be impaired by various cardiovascular disorders, such as heart valve disease, arrhythmias, coronary artery disease, and heart failure. Majority of these conditions are deadly besides critical health problems such as heart attack and stroke. In this way, early diagnosis and proper diagnosis of heart disease should be performed. This extreme condition needs proper treatment that is anchored on a fast and reliable diagnostic process [1]. Diagnosis of heart disease using the traditional approach involves physical assessment, history, and clinical assessment. However, these techniques may not always be sufficient for accurately identifying cardiac disease, especially in its early stages. To enhance the diagnosis and prognosis of heart disease, doctors are increasingly using data mining and machine learning techniques to examine enormous volumes of clinical data [2].

Several machine learning methods are frequently employed, including hybrid and ensemble learning. Forecasting cardiac illness is always more accurate with combined machine learning models than with solo algorithms and traditional methods [3]. These

models successfully address the shortcomings of individual methods and utilize their complementary advantages since they combine two or more algorithms. As an example, neural networks can be more accurate in predicting complex non-linear interactions whereas decision trees are interpretable and easy to see which features are most important. The combination of the approaches will enable the creation of a model that will be rather precise and understandable to the medical staff. This general manner of approach improves the generalizability of other populations and sets of data since it minimizes overfitting [4]. The hybrid models can be useful in combining the different types of data, including numerical, categorical, and visual data, and thereby a better and more accurate determination of the risk profile of a patient can be made. Such hybrid machine learning methods can be synergistically combined to predict heart disease with a considerable enhancement of accuracy and performance of the predictive analytics [5].

This paper suggests Causal-GAT-LLM, a new hybrid framework to jointly promote Causal Graph Attention Networks (Causal-GAT), Large Language Models (LLM) and Linear Non-Gaussian Acyclic Models of Feature Mapping (LiNGAfM) to advance the accuracy and interpretability of heart disease prediction. The technique initiates with structured clinical and patient health data, LiNGAfM is applied to highlight underlying causal associations among features, which can be used to disentangle features and interpret them. These causally sensitive attributes are then cascaded using a Graph Attention Network which learns complex interactions between patient attributes by dynamically weighting informative features. A pre-trained LLM is added to augment the semantic knowledge with contextual reasoning on textual health notes and medical narratives to provide a suitable match between the structured and unstructured data. This multimodal combination makes it easier to understand the results of prediction, and it is interpretable. The resulting system does not only show good classification on benchmark data, but also gives human interpretable explanations of model decisions, as well as causal attributions, which is essential in clinical AI systems to create more explainable systems.

Overall, our contributions are as follows

- To develop a new hybrid framework, Causal Graph Attention Network with Large Language Model and Linear Non-Gaussian Acyclic Model (Causal-GAT-LLM-LiNGAM), for robust and explainable heart disease prediction.
- The approach starts with the application of the LiNGAM algorithm to structured clinical data to uncover directed causal relationships among variables such as blood pressure, cholesterol, and smoking history.
- Finally, these causal graphs form the backbone of a GAT, which learns to emphasize the most influential features through attention mechanisms. Simultaneously, a domain-specific Large Language Model (LLM), such as BioBERT or GPT-based variants, is employed to extract semantic embeddings and causal cues from unstructured clinical texts, including physician notes and discharge summaries. These text-derived features are integrated into the graph to enrich node and edge representations.

This article is organized as follows for the remainder: Section 2 reviews the literature; Section 3 explains the proposed approach; Section 4 presents the findings and discussion; and Section 5 discusses the conclusion and next steps.

2. SURVEY

Heart disease is still one of the world's leading causes of death, and early detection and precise diagnosis are essential to both its prevention and successful treatment. Because data mining, machine learning, and artificial intelligence have advanced so quickly, researchers are using computer models more and more to advance the accuracy of heart illness prediction. To categorize patient cases and find important risk factors in clinical data, methods such as ensemble learning, neural networks, decision trees, and support vector machines have been used. In this review, the recent developments in heart disease prediction are discussed with attention to current methods, emerging tendencies, and challenges, and future opportunities of further development.

Biswas et al. [6] working a variety of feature selection strategies to determine the most crucial characteristics for their suggested machine learning model, which will be used to detect cardiac conditions early on. Three feature subsets (SF1, SF2, and SF3) were generated using the mutual information, ANOVA, and chi-square methods. The best model-feature combination was found using six machine learning techniques. With accuracy of 94.51%, sensitivity of 94.87%, specificity of 94.23%, log loss of 0.31, and AUC of 94.95, the random forest perfect based on the SF3 feature subset was the best of them all. The model may be a useful tool for the early clinical identification of heart disease, and the results show that it has an outstanding cost-performance ratio.

Alshraideh et al. [7] provided an extensive method to improve the validity of the forecast, since it involves the selection of features, data pre-processing, and model development. Particle Swarm Optimization (PSO) was used with them in the selection of features, and their performance was compared to other AI algorithms. They were able to discover the 58 most important characteristics through PSO. The practice could assist physicians to make superior decisions as the practice encourages early diagnosis, diagnosis, and treatment of heart disease. When applied to the JUH data of 486 cardiac patients, the method demonstrated an accuracy of 94.3, which is equivalent to the state-of-the-art techniques. On the other hand, other algorithms on a different dataset that we were working on had an accuracy of 85-90 percent.

According to Fatima et al. [8], a hybrid approach to cardiac disease prediction, which utilizes genetic algorithms (GAs) and convolutional neural networks (CNNs), was proposed. CNNs are effective with regard to identifying the hierarchical patterns and distilling complex features in medical data, whereas GAs enhance the work of the model by choosing features in the most optimal way and improving the degree of prediction. The proposed solution was trialed on popular heart disease-related datasets with assistance of such key presentation metrics as accuracy, precision, recall, and processing speed. The experimental findings have revealed that CNN-GA hybrid model is more efficient as compared to the traditional ones and is scalable and reliable in regard to the prediction of heart disease. The article demonstrates how artificial intelligence (AI) can be used in the medical diagnosis area to apply machine learning and evolutionary algorithms to enhance the standard of care given to patients in a global setting.

Abdullah et al. [9] claim that the machine learning algorithms are accurate and effective substitutes to the detection and categorization of individuals with heart disease. Early detection of heart failure would be necessary to prevent its occurrence as it raises the survival rates of the patients. To make the forecast more accurate, the study used sophisticated data preparation methods and a specific model of the Enhanced Multilayer Perceptron (EMLP). The suggested classification algorithm could achieve an accuracy rate of 92 per cent on CDC cardiac disease dataset which is greater compared to the traditional techniques. As the results showed, the EMLP model is more accurate, recalls high precision, and F1-score than the current algorithms, thereby, the early process of cardiovascular disease prognosis and prediction are improved significantly.

According to Wang et al. [10], the suggested LGAP model combines Long Short-Term Memory (LSTM) networks, Graph Neural Networks (GNNs), and Multi-Head Attention mechanisms that are effectively used to analyze time-series data. An investigation of the behavioural patterns, as well as the relationship between the two, could be learned using this model with the help of the relational graph data along with the data on the patient related time-series, that is, the activity logs, physiological measurements and medical records. The practice is holistic making the prognosis of cardiovascular disease risk more accurate and universal. The experiment results conducted on such datasets like PhysioNet and NHANES prove the fact that LGAP is more effective than the traditional models, particularly, in personalized health management and predicting accuracy. The development of LGAP has given an alternative way of anticipating the cardiovascular diseases and forming the treatment plan of the patients.

Kumar and Thakur [11] proposed using stacked collective machine learning to increase the accuracy of cardiac disease estimate. The procedure starts by using an unsupervised learning method called a Variational Autoencoder (VAE) to find important patterns in the data. Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, Naive Bayes, and Quadratic Discriminant Analysis are among the classifiers that are combined to produce final predictions. SVMs divide the data, Decision Trees produce different decisions, KNN identifies related situations, and Naive Bayes forecasts likelihood. Each model is unique. Two such ensemble techniques that use the benefits of several models to increase accuracy are bagging and AdaBoost. To extract more intricate patterns from the input, a Dense Neural Network (DNN) is stacked with the VAE itself. To train a meta-learner that will be used for the final classification, all of the estimates from the individual perfect are combined. The arranged ensemble technique of heart disease prediction (HDPM) outperformed the other methods in terms of efficacy and stability, achieving an accuracy of 92.3 and a precision of 89.3 during 10 training epochs.

To precisely identify and categorize cardiovascular disease (CVD), Khan et al. [12] suggested two deep learning models: Blending-based Cardiovascular Disease Detection Network (BlCVDD-Net) and Ensemble-based Cardiovascular Disease Detection Network (EnsCVDD-Net). BlCVDD-Net is a combination of LeNet, GRU, and Multilayer Perceptron (MLP) in a blending technique whereas EnsCVDD-Net is a combination of LeNet and Gated Recurrent Unit (GRU) in an ensemble method. SHAP offers a complete explanation of how different features affect the diagnosis of CVD. There are several performance measures used to gauge the performance of the networks. With a precision of 91%, recall of 85%, accuracy of

88%, and F1-score of 88%, EnsCVDD-Net outperforms baseline models and consumes 777 seconds. In a similar vein, BlCVDD-Net outperforms top deep learning models, achieving 91% accuracy, 91% F1-score, 96% precision, and 86% recall while executing at a noticeably quicker 247-second rate. Table 1 displays the review of the literature for the suggested strategy.

Table 1: Literature survey for proposed method

Ref.	Techniques / Models Used	Dataset / Features	Key Results
Biswas et al. [6]	Feature Selection (Chi-Square, ANOVA, MI); ML Models (LR, SVM, KNN, RF, NB, DT)	UCI / SF1, SF2, SF3 feature subsets	RF with SF3 achieved 94.51% accuracy, 94.87% sensitivity, 94.23% specificity, and an AUC of 94.95.
Alshraideh et al. [7]	RF, SVM, DT, NB, and KNN with Particle Swarm Optimization (PSO) for feature selection.	JUH dataset (486 patients / 58 features)	PSO-RF achieved 94.3% accuracy; others: 85–90%
Fatima et al. [8]	Hybrid CNN + Genetic Algorithms (GA)	Standard cardiac datasets	CNN-GA model outperformed conventional methods; high accuracy and computational efficiency
Abdullah et al. [9]	Enhanced Multilayer Perceptron (EMLP) with data refinement	CDC heart disease dataset	EMLP achieved 92% accuracy, superior precision, F1-score, and recall
Wang et al. [10]	LGAP: LSTM + GNN + Multi-Head Attention for time-series & relational data	PhysioNet, NHANES	LGAP showed superior accuracy & generalization; enables personalized health management
Kumar & Thakur [11]	Stacked Ensemble Model with VA, KNN, SVM, DT, NB, QDA, DNN; meta-learning approach	Not specified (generic clinical data)	Ensemble model (HDPM) reached 92.3% accuracy, 89.3% precision after 10 epochs
Khan et al. [12]	EnsCVDD-Net (LeNet + GRU), BlCVDD-Net (LeNet + GRU + MLP) with SHAP for interpretability	Cardiovascular disease dataset	BlCVDD-Net: 91% accuracy, 96% precision; EnsCVDD-Net: 88% accuracy; explainable AI used.

2.1 Research gap

The lack of personalized, real-time, and interpretable models capable of effectively integrating diverse patient data including genetic information, wearable sensor outputs, and electronic health records represents a significant research gap in heart disease prediction, despite notable advances in machine learning and medical diagnostics. Current models often struggle to integrate multi-modal data sources, generalize across populations, and handle data imbalance. Moreover, most research relies on static datasets and overlooks the potential of continuous learning systems that can adapt over time to new clinical knowledge and patient-specific risk factors.

2.2 Disadvantage of proposed model

- The dataset may have a skewed class distribution (e.g., more healthy than diseased cases), leading to biased predictions and poor presentation on minority classes.
- The model can employ very few clinical characteristics, and thus, it might miss other important characteristics like genetic information, lifestyle, or ECG patterns.

- High model complexity or insufficient regularization can lead to overfitting, especially when the training data is limited or lacks diversity.
- Healthcare professionals might have difficulty in understanding or believing the predictions of complex models, including deep learning models, because they lack transparency.

3. PROPOSED SYSTEM

3.1 Overview

This section presents a new hybrid model, the Causal Graph Attention Network with Large Language Model and Linear Non-Gaussian Acyclic Model (Causal-GAT-LLM-LiNGAM), this is a reliable and comprehensible forecast of heart disease. To find directed causal relationships between variables like blood pressure, cholesterol, and smoking history, the technique first applies the LiNGAM algorithm to structured clinical data. These causal graphs serve as the foundation for a GAT, which uses attention mechanisms to learn which properties are most important. Simultaneously, a domain-specific Large Language Model (LLM), such as BioBERT or GPT-based variants, is employed to extract semantic embeddings and causal cues from unstructured clinical text, including physician notes and discharge summaries. These text-derived features are integrated into the graph to enrich node and edge representations. The architectural design of the suggested method is illustrated in Figure 1.

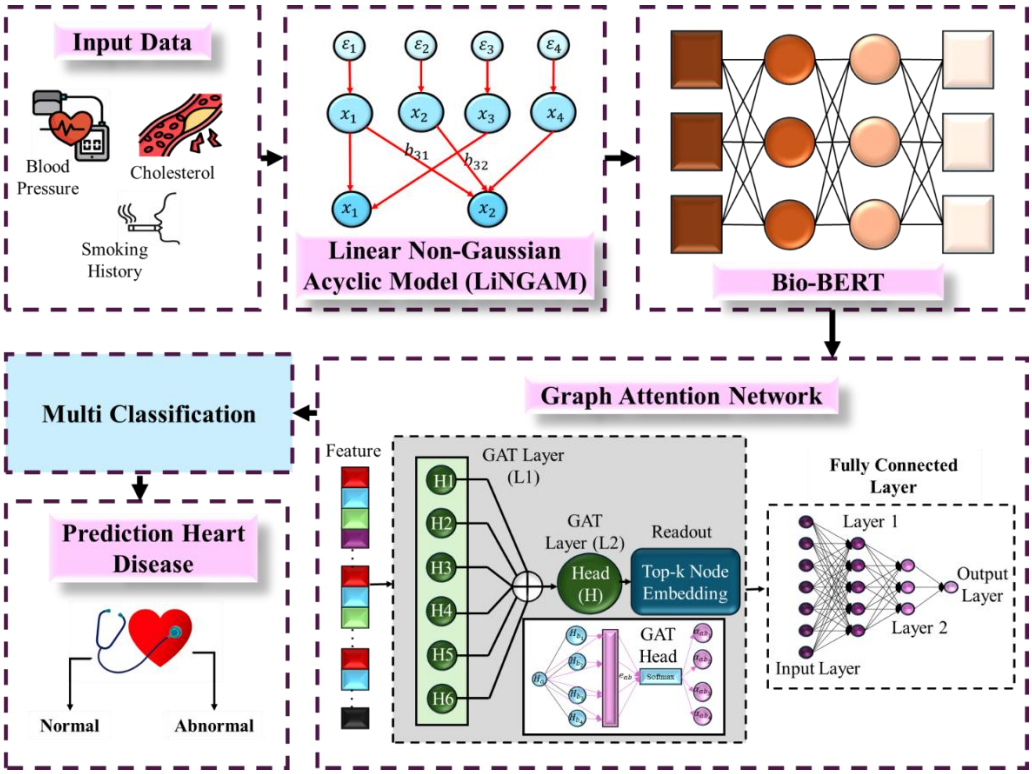

Figure 1: Architecture diagram for suggested method.

The Fig.1 presents a hybrid deep learning perfect of predicting heart diseases based on causal thinking, language modelling and graph neural networks. The Input Data block serves to accumulate any different types of clinical data such as ECG signals, vascular data, speech or textual input. LiNGAM module (i.e., the Linear Non-Gaussian Acyclic Model) involves causal inference based on directional association among clinical variables (e.g., x1 ->x2x_1 ->x_2x1 ->x2) that are estimated by some causal coefficients (bijb_{ij}bij) under the assumption of non-Gaussian. Simultaneously, unstructured textual data are translated into semantic features by means of BioBERT, a transformer-based language model that was trained using biomedical text. This can be done by means of clinical notes, or patient narratives, among others. These features and structured data pass through a GAT, which estimates topological relationships (by passing attention-based messages) between patient features. The GAT uses several layers through which both heads compute weighted feature embeddings and it improves feature importance representation through Top-k Node Embedding. The learned embeddings are then relayed into a Fully Connected Layer that consists of input,

output and hidden (Layer 1 and Layer 2) layer that facilitates the downstream tasks like classification. Multi Classification block works using these outputs to forecast the existence of a heart disease based on Normal and Abnormal cases. Those are causal discovery, language semantics, and relational learning which give a comprehensible and interpretable pipeline consisting of causal-discovery-based heart disease prediction that is accurate.

3.1 Dataset

MIMIC-III Dataset

The Medical Information Mart for Intensive Care III (MIMIC-III) is a publicly existing critical care database that is often utilized to predict cardiac illness due to its large and diverse clinical data. The de-identified electronic health records of over 60,000 intensive care unit patients admitted to Beth Israel Deaconess Medical Center between 2001 and 2012 are also present. The collection contains diagnostic codes, vital signs, test results, prescription drugs, clinical notes, and demographic data. To predict cardiac sickness using machine learning and deep learning algorithms, relevant features are obtained and preprocessed, such as heart rate, blood pressure, cholesterol, ECG measurements, and comorbidities [13]. The multimodality of the data enables a detailed analysis of cardiovascular risk variables and allows the expansion of comprehensible and reliable models based on both structured and unstructured data. MIMIC-III is a benchmark dataset that can be used to further the research on predictive modeling of heart disease due to its wide use and the quality of annotations.

UCI Heart Disease

The UCI Heart Disease dataset, available through the UCI Machine Learning Repository, is frequently used as a benchmark for developing and evaluating predictive models for heart disease diagnosis. Clinical information about patients is collected from various sources, including the VA Long Beach, Hungarian, Cleveland, and Swiss databases. Although the dataset contains 76 variables, most analyses focus on 13 to 14 key aspects, including age, sex, resting blood pressure, serum cholesterol, maximum heart rate achieved, kind of chest discomfort, ECG, and stress test results [14]. These characteristics assist identify if heart disease is present or not and are suggestive of cardiovascular health. Usually binary or multiclass, the target adjustable shows the presence or severity of heart disease. Due to its structured format, manageable size, and well-defined clinical features, the UCI Heart Disease dataset serves as an ideal testbed for evaluating machine learning algorithms, feature selection technique, and explainable AI models aimed at improving heart disease prediction and risk stratification.

3.2 Preprocessing

Preprocessing of both structured and unstructured data is essential to ensure reliable and consistent input to the hybrid Causal-GAT-LLM-LiNGAM framework. The initial stage in handling missing values in organized clinical data is to impute them using methods like mean imputation, in which the mean of the available values is used to replace each missing entry x_i:

$$x_i^{imp} = \begin{cases} x_i, & if\ x_i \neq NaN \\ \frac{1}{N}\sum_{j=1}^{N} x_j, & if\ x_i = NaN \end{cases} \quad (1)$$

Next, the features are standardized to a common scale using Z-score normalization, defined as:

$$x_i^{std} = \frac{x_i - \mu}{\sigma} \quad (2)$$

Where σ and μ stand for the feature's mean and standard deviation, respectively. Categorical variables are transformed using one-hot encoding, converting a category c_i into a binary vector:

$$OneHot(c_i) = [0,.....,1,....,0] \in \square^k \quad (3)$$

Where k is the number of unique categories. For unstructured text data, such as physician notes and discharge summaries, preprocessing involves tokenization, where the input text T is split into tokens:

Natural Resources for Human Health
eISSN: 2583-1194

$$T = "Patient\ has\ elevated\ BP\ and\ chest\ pain" \rightarrow ["Patient","has",....] \quad (4)$$

Each token is then normalized through lowercasing and lemmatization (e.g., "running" → "run") and stripped of noise such as punctuation and stopwords. The tokens are cleaned and then organized into input sequences that work with Large Language Models (LLMs), including GPT-based or BioBERT versions. These sequences are represented as high-dimensional vectors.

$$E = [e_1, e_2, ..., e_n] \in \mathbb{R}^{n \times d} \quad (5)$$

Where e_i is the embedding of the i-th token and d is the embedding dimension. This comprehensive preprocessing pipeline ensures that both numerical and textual modalities are clean, consistent, and semantically enriched, enabling the graph-based learning model to effectively integrate and leverage all available clinical information.

3.3 Causal Graph Construction using LiNGAM

To model the causal dependencies among structured clinical variables, the Linear Non-Gaussian Acyclic Model (LiNGAM) [15] is applied. LiNGAM assumes a linear structural equation model in which each variable is represented as a linear combination of its direct causes, along with an independent non-Gaussian noise component. Fig.2 displays the LiNGAM architecture diagram for the estimate of heart disease.

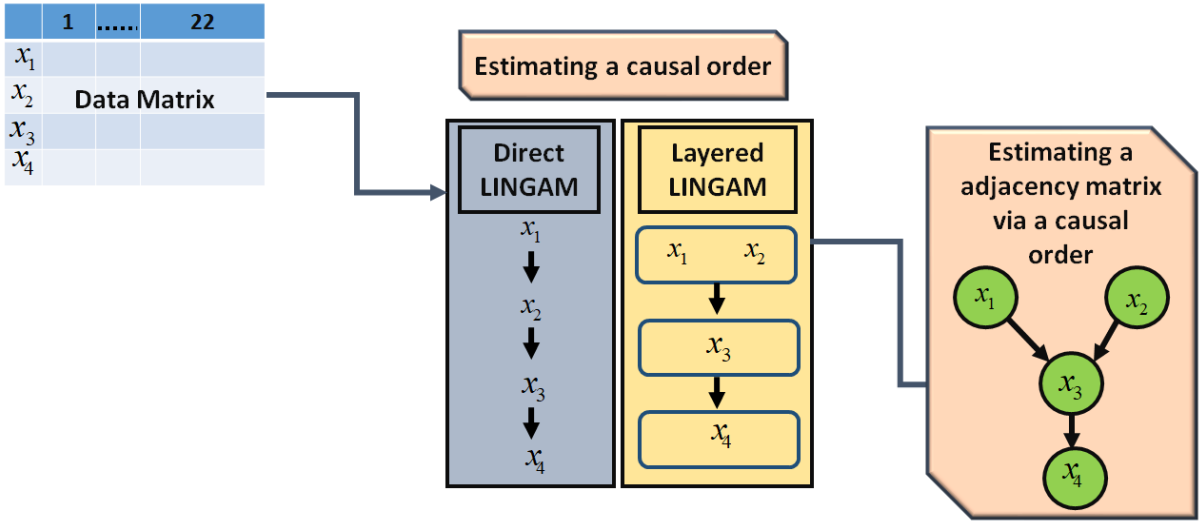

Figure 2: Architecture diagram for LiNGAM based on heart disease prediction.

The model is defined as follows given an observed data vector $X = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n$, where each x_i denotes a clinical characteristic like cholesterol or blood pressure:

$$X = BX + e \quad (6)$$

Here, $B \in \mathbb{R}^{n \times n}$ is a strictly lower triangular matrix encoding the direct causal effects between variables, and $e \in \mathbb{R}^n$ is a vector of statistically independent non-Gaussian noise terms. This equation can be rearranged as follows:

$$X = (I - B)^{-1} e \quad (7)$$

The LiNGAM algorithm estimates the matrix B, from which a Directed Acyclic Graph (DAG) is constructed. In this graph, the nodes correspond to the clinical variables, and a directed edge from x_j to x_i is drawn if and only if $B_{ij} \neq 0$. Formally, the set of edges is defined as:

$$E = \left\{ \left(x_j, x_i \right) | \, B_{ij} \neq 0 \right\} \tag{8}$$

The causal structure included in the clinical dataset is captured by the resultant causal graph $G = \left(V, E \right)$, where $V = \left\{ x_1, x_2, ..., x_n \right\}$. For example, the model may reveal that smoking influences blood pressure, which in turn affects the likelihood of developing heart disease, represented as:

$$smoking \rightarrow blood \;\; pressure \rightarrow heart \;\; disease \tag{9}$$

This causal DAG provides the structural foundation for the GAT, ensuring that subsequent learning stages are both data-driven and clinically interpretable.

3.4 Text Feature Extraction using LLM

To incorporate contextual information from clinical narratives, domain-specific Large Language Models (LLMs), such as BioBERT or GPT-based architectures, are utilized for semantic and causal feature extraction [16]. The construction diagram of the LLM is shown in Fig.3.

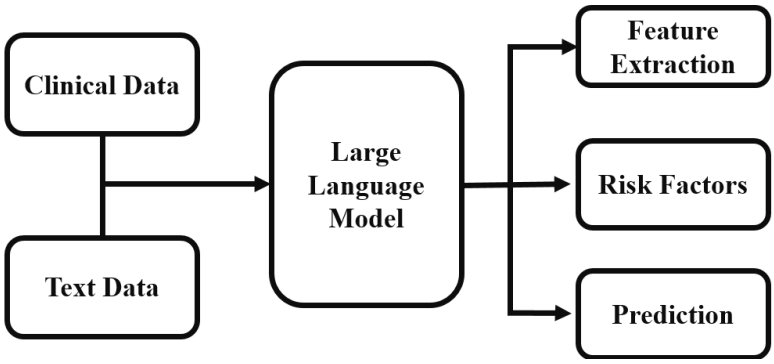

Figure 3: Architecture diagram for LLM based on heart disease prediction.

Given an unstructured clinical text sequence $T = \left[t_1, t_2, ..., t_n \right]$, where each t_i denotes a token (e.g., "chest", "pain", "elevated"), the LLM processes the sequence and generates corresponding high-dimensional contextual embeddings:

$$E = \left[e_1, e_2, ..., e_n \right], \;\; e_i \in \mathbb{R}^d \tag{10}$$

Each embedding e_i captures the semantic meaning of the token t_i in relation to its surrounding context, where d is the embedding dimension. To obtain a unified representation of the entire text, the embeddings are aggregated using a pooling strategy or extracted directly from a special token, such as [CLS] in BERT-based models.

$$e_{text} = Pooling\left(E \right) \in \mathbb{R}^d \tag{11}$$

This embedding e_{text} encodes not only the semantic context but also implicit causal cues embedded in the clinical notes (e.g., "smoking leads to elevated blood pressure"). To enrich the structured clinical variables in the graph, the textual representation is aligned with relevant nodes. For a structured variable x_i, let $e_{x_i}^{struct}$ represent its structured feature embedding, and $e_{x_i}^{text} \subseteq e_{text}$ represent the semantically aligned component from the text. These are combined to form a fused embedding:

$$e_{x_i}^{fused} = Fuse\left(e_{x_i}^{struct}, e_{x_i}^{text} \right) \tag{12}$$

Natural Resources for Human Health
eISSN: 2583-1194

The fusion can be implemented via concatenation, attention-based gating, or projection. This enriched node representation $e_{x_i}^{fused}$ ensures that each clinical variable in the graph benefits from both quantitative measurements and qualitative narrative insights, laying a strong foundation for interpretable and accurate graph-based heart disease prediction.

3.5 Graph Construction and Fusion

To unify structured clinical data with semantic information extracted from text, we construct a heterogeneous GAT based on a causal graph inferred using LiNGAM. The base graph is definite as $G=(V,E)$, where each node $v_i \in V$ corresponds to a clinical variable such as blood pressure or cholesterol, and each directed edge $(v_j \to v_i) \in E$ represents a causal relationship identified by the LiNGAM model. The edges are derived from the non-zero entries of the causal coefficient matrix B, such that:

$$E=\{(v_j,v_i) \mid B_{ij} \neq 0\} \tag{13}$$

Each node is initialized with a fused feature vector that combines structured attributes $e_{x_i}^{struct} \in \mathbb{R}^{d_1}$ and semantic embeddings from the LLM $e_{x_i}^{text} \in \mathbb{R}^{d_2}$. These are concatenated and then projected into a shared space.

$$h_i = \mathrm{Re}LU\left(W\left[e_{x_i}^{struct} \,\|\, e_{x_i}^{text}\right]+b\right) \tag{14}$$

Where $W \in \mathbb{R}^{d\times(d_1+d_2)}$ and $b \in \mathbb{R}^{d}$ are learnable parameters, and $\|$ denotes vector concatenation. Each edge (i,j) is also assigned a feature vector $e_{ij} \in \mathbb{R}^{k}$, incorporating both the causal strength from LiNGAM B_{ij} and optional semantic relation weights α_{ij}^{text} extracted from the LLM:

$$e_{ij}=\left[B_{ij} \,\|\, \alpha_{ij}^{text}\right] \tag{15}$$

The GAT then computes attention coefficients for each edge, dynamically weighting the contributions of neighboring nodes during message passing. The following formula is used to regulate the kindness coefficient among node i and neighbor j:

$$\alpha_{ij}=\frac{\exp\left(Leaky\,\mathrm{Re}LU\left(a^T\left[Wh_i \,\|\, Wh_j \,\|\, e_{ij}\right]\right)\right)}{\sum_{k\in N(i)}\exp\left(Leaky\,\mathrm{Re}LU\left(a^T\left[Wh_i \,\|\, Wh_k \,\|\, e_{ik}\right]\right)\right)} \tag{16}$$

Finally, the updated illustration of node i is obtained through an attention-weighted aggregation of its neighbors:

$$h_i^{'}=\sigma\left(\sum_{j\in N(i)}\alpha_{ij}Wh_j\right) \tag{17}$$

This process produces a multimodal, causally grounded, and semantically enriched graph representation that improves both prediction accuracy and interpretability for downstream heart disease risk classification.

3.6 Graph Attention Network (GAT) Learning

To refine node representations in the multimodal causal graph, a GAT is employed to assign dynamic attention weights to neighboring nodes based on their relative importance [17]. Each node v_i with initial feature vector $h_i \in \mathbb{R}^{F}$ is first linearly transformed into a new feature space using a learnable weight matrix $W \in \mathbb{R}^{F^{'}\times F}$:

$$h_i^{'} = Wh_i \tag{18}$$

For each pair of linked nodes (i, j) a shared attention mechanism computes a raw attention score e_{ij} that reflects the influence of neighbor j on node i, combining both structural and semantic relevance:

$$e_{ij} = Leaky\,\mathrm{Re}\,LU\left(a^T\left[h_i^{'} \| h_j^{'}\right]\right) \tag{19}$$

Here, $a \in \mathbb{R}^{2F^{'}}$ is a learnable attention vector, and $\|$ denotes vector concatenation. These raw scores are normalized using the softmax function across all neighbors $N(i)$ of node i:

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in N(i)}\exp\left(e_{ik}\right)} \tag{20}$$

The regularized attention weights α_{ij} are then used to compute the updated illustration h_i^{new} of node i through a weighted aggregation of its neighbors' transformed features:

$$h_i^{new} = \sigma\left(\sum_{j\in N(i)}\alpha_{ij}h_j^{'}\right) \tag{21}$$

An activation function that is not linear, such as ReLU, is represented here by σ. The attention-based aggregation approach enhances the expressiveness and interpretability of node embeddings for tasks such as heart disease prediction by enabling the perfect to selectively emphasize the most semantically and causally relevant neighbors.

3.7 Final Prediction

Once the node embeddings h_i^{new} have been updated through GAT layers, a graph-level representation h_G is computed by aggregating these node embeddings using a simple readout function such as mean pooling:

$$h_G = \frac{1}{|V|}\sum_{i=1}^{|V|}h_i^{new} \tag{22}$$

This global graph embedding h_G captures both causal and semantic information and is passed into a classification layer (e.g., softmax or MLP) to predict the patient's heart disease risk level:

$$\hat{y} = soft\max\left(W_c h_G + b_c\right) \tag{23}$$

Where $W_c \in \mathbb{R}^{C\times d}$ and $b_c \in \mathbb{R}^{C}$ are learnable limitations, and C is the number of prediction classes (e.g., No Heart Disease, Mild, Severe). For interpretability, attention weights α_{ij} learned by the GAT indicate the relative influence of neighboring nodes, enabling feature importance analysis. The total importance of a node v_i is quantified as:

$$\mathrm{Im}\,por\tan ce\left(v_i\right) = \sum_{j\in N(i)}\alpha_{ij} \tag{24}$$

In parallel, narrative explanations are generated from the LLM's text embeddings to justify decisions in a human-readable form, thereby enhancing clinical trust and enabling expert validation.

Natural Resources for Human Health
eISSN: 2583-1194

3.8 Algorithm for proposed method

Input:

X_structured ← Structured clinical data (e.g., BP, cholesterol)

T_unstructured ← Unstructured clinical text (e.g., physician notes)

Datasets ← {MIMIC-III, UCI Heart Disease}

Output:

y_hat ← Predicted heart disease risk scores

Explanations ← Attention-based and LLM-generated interpretations

Begin

1. Preprocessing Structured Data

For each feature x_j in X_structured:

If missing(x_j): x_j ← Mean(x_j)

x_j ← ZScoreNormalize(x_j)

X_encoded ← OneHotEncodeCategorical(X_structured)

2. Preprocessing Unstructured Text Data

For each text record t_i in T_unstructured:

tokens_i ← Tokenize(t_i)

tokens_i ← CleanText(tokens_i) lowercasing, lemmatization, remove stopwords

z_i ← LLM_Embed(tokens_i) e.g., BioBERT or GPT

3. Causal Graph Construction using LiNGAM

B ← LiNGAM(X_encoded)

G ← ConstructGraph(B)

For each node v_i in G:

h_struct[i] ← X_encoded[i]

For each edge e_ij:

If B_ij ≠ 0: AddDirectedEdge(G, v_j → v_i)

4. Feature Fusion

For each node v_i in G:

h_text[i] ← z_i // from LLM

h_fused[i] ← Concat(h_struct[i], h_text[i])

SetNodeFeature(G, v_i, h_fused[i])

5. Graph Attention Network (GAT)

Initialize GAT weights W, attention vector a

For each node v_i in G:

For each neighbor v_j ∈ Neighbors(v_i):

e_ij ← LeakyReLU(a^T · [W·h_fused[i] || W·h_fused[j]])

α_ij ← Softmax(e_ij over j ∈ Neighbors(i))

h_updated[i] ← σ(Σ_j α_ij · W·h_fused[j])

6. Graph Pooling and Prediction

h_graph ← MeanPooling(h_updated)

y_hat ← Softmax(W_out · h_graph + b_out)

7. Interpretability

For each node v_i:

importance[v_i] ← Σ_j α_ij

For each patient:

narrative[i] ← LLM_GenerateExplanation(z_i, attention_weights[i])

8. Evaluation

Evaluate y_hat using Accuracy, Precision, Recall, F1, MCC, NPV

Report execution time and compare with baseline models

End

3.9 Layer diagram for proposed method

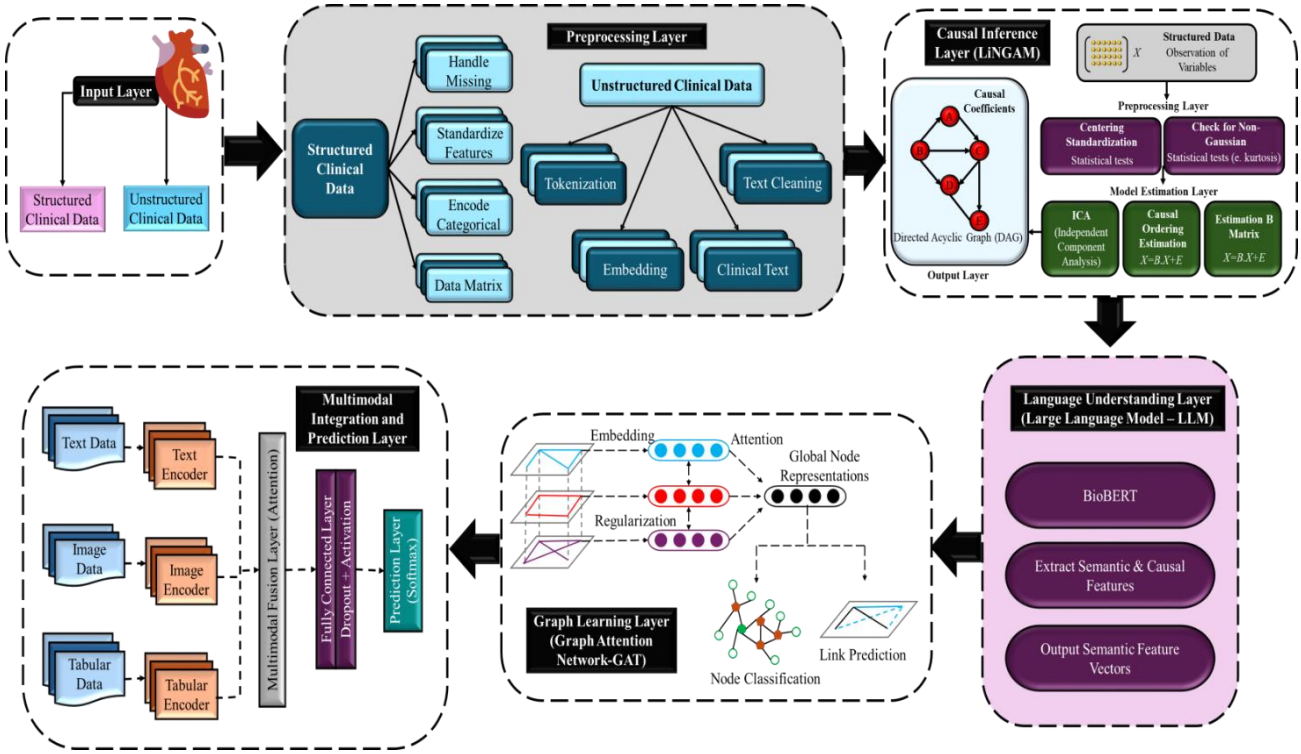

Figure 4: Layer diagram for proposed method

The Fig.4 shows a complete hybrid structure of explainable heart disease prediction, which combines both structured and unstructured data. Usage of sophisticated AI-based approaches. The first step in the input layer is to collect clinical information, both structured (like lab results and vital signs) and unstructured (like clinical notes). At Pre-processing Layer, the data that is structured is processed such as dealing with missing value, standardization of features, and replacing categorical data with encodings to build a data matrix. At the same time, the unstructured data is washed, token-ized, embedded and converted into the clinical text. The next step is when the Causal Inference Layer (LiNGAM) is used in the structured data to infer the resulting determination of the causal relationships through Directed Acyclic Graph (DAG). This is done by centering, standardization, testing non-Gaussianity, Independent Component Analysis (ICA) and estimation of the causal B matrix. Simultaneously, Language Understanding Layer (LLM) extracts semantic and causal attributes of unstructured clinical text, and provides semantic feature vectors as output based on BioBERT. Such outputs are then fed to Graph Learning Layer in which a Graph Attention Network (GAT) learns node embeddings through attention to meaningful clinical features. They are learned embeddings, and they are combined in the Multimodal Fusion Layer with separately encoded text data, image data and tabular data (encoded using related encoders) using attention mechanisms. And, lastly, the combined features are fed in a fully connected layer comprising dropout and activation functions and the final output comes in the form of a softmax-based Prediction Layer that provides the heart disease prediction. This end-to end system promises both accuracy in diagnosis and interpretability in the form of causal and semantic knowledge.

4. RESULTS

The Causal-GAT-LLM-LiNGAM framework proposed to forecast heart disease is fully evaluated here. The model's presentation on the MIMIC-III and UCI Heart Disease benchmark datasets is estimated using a variety of metrics, including as execution time, accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Negative Predictive Value (NPV). The efficiency, robustness, and interpretability of the given approach are the factors that are emphasized by the presented comparison with the existing models. Its experimental outcomes prove that it is better in terms of accuracy and speed of prediction.

4.1 Experimental setup

A PC with an Intel Core i7 CPU, 64-bit operating system, and machine learning perfect were advanced and trained using 16 GB of RAM to predict cardiac disease. Python and popular machine learning frameworks like scikit-learn were used in the implementation. On this computing environment, sensitivity, specificity and accuracy of different supervised learning algorithms to classify heart diseases were compared. In the research using the existing models are PS-GWO [18], CNN [19], LR [20] and MLP [21].

4.2 Performance metrics

Accuracy: The number of accurately predicted positive and negative cases is divided by the overall number of instances to regulate the accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{25}$$

Precision: Precision is defined as the fraction of accurately expected positive cases among all cases expected as positive. It focuses on the accuracy of optimistic forecasts.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{26}$$

Recall: Recall is the amount of definite positive cases that the model accurately detects.

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{27}$$

F1-Score: The F1-score is a composed evaluation metric that works especially well with imbalanced datasets. It is computed as the harmonic mean of accuracy and recall. The trade-off among these two measures is accurately reflected.

$$F1-Score=2\times\frac{\Pr ecision\times \operatorname{Re} call}{\Pr ecision+\operatorname{Re} call}=\frac{2TP}{2TP+FP+FN} \quad (28)$$

MCC: An impartial metric for assessing binary classifier performance, particularly in cases of class imbalance, is the MCC. True positives, true negatives, false positives, and false negatives are the four parts of the confusion matrix that are explained.

$$MCC=\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (29)$$

NPV: The NPV is the proportion of correctly anticipated negative cases overall. It shows that test results that are negative are reliable.

$$NPV=\frac{TN}{TN+FN} \quad (30)$$

Execution Time: Execution time refers to the total duration taken by an algorithm or model to complete its task, such as training, testing, or inference.

$$Execution\ Time=T_{end}-T_{start} \quad (31)$$

Table 2: Description of MIMIC-III and UCI Heart Disease Datasets.

Dataset	Classes	Attributes	Instances
1. MIMIC-III	Class-0 (No Heart Disease)	29	5,000
	Class-1 (Heart Disease)	29	5,000
2. UCI Heart Disease	Class-0 (No Heart Disease)	13	164
	Class-1(Heart Disease)	13	139

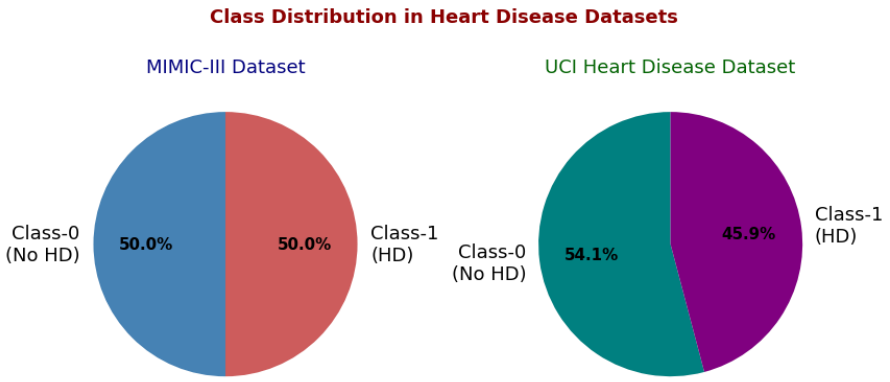

Figure 5: Description of MIMIC-III and UCI Heart Disease Datasets

1. MIMIC-III Dataset

Tan.2 and Fig.5 shows The UCI Heart Disease Datasets and MIMIC-III are described. The MIMIC-III dataset, widely used in clinical and medical research, consists of comprehensive health records of intensive care unit patients. In this study, the dataset has been divided into two classes:

- Class 0 represents patients without heart disease, comprising 5,000 instances and 29 clinical characteristics, including age, heart rate, blood pressure, cholesterol, and test findings.
- Class-1 includes patients diagnosed with heart disease and consists of 5,000 instances with the same set of 29 attributes. Effective training and assessment of machine learning perfect for the estimate of heart disease are made possible by this balanced distribution.

2. UCI Heart Disease Dataset

The UCI Heart Disease dataset is regularly used as a standard in machine learning studies aimed at predicting heart disease. It is categorized as follows:

- Class 0 indicates individuals that are not diagnosed with heart disease. Age, sex, type of chest discomfort, resting blood pressure, cholesterol, fasting blood sugar, and maximum heart rate are among the 13 essential characteristics of the 164 patients.
- Class-1 corresponds to individuals diagnosed with heart disease, comprising 139 instances and the same 13 attributes. Although relatively small, this dataset is highly useful for initial model training and comparative analysis due to its clean and structured format.

Table 3: Heart Disease Classification

Strategy	Classes	Accuracy	Precision	Recall	F1-Score
PS-GWO	Class-0	91.34	88.23	93.12	74.35
	Class-1	89.82	85.46	90.47	72.73
CNN	Class-0	92.59	81.53	94.39	71.46
	Class-1	90.36	85.26	96.25	78.10
LR	Class-0	89.53	85.85	94.36	75.35
	Class-1	88.61	81.63	90.53	74.56
MLP	Class-0	88.85	80.34	93.71	75.75
	Class-1	87.36	85.21	95.56	79.43
Proposed	**Class-0**	**96.09**	**94.12**	**98.46**	**96.24**
	Class-1	**93.33**	**95.83**	**90.79**	**93.24**

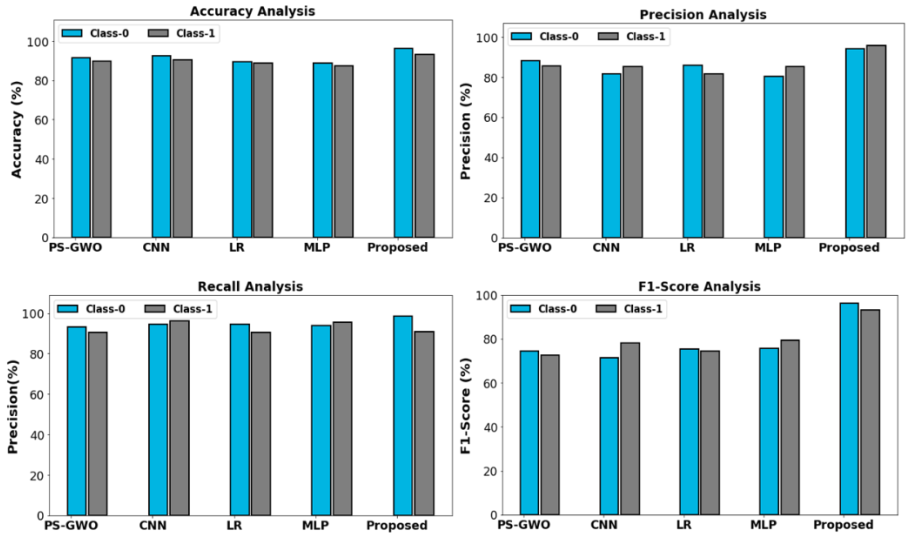

Figure 6: Heart Disease Classification

Table 3 and Fig.6 provide performance analysis of a number of existing models applied to the classification of heart diseases with PS-GWO, CNN, LR, MLP, and proposed model. In the case of PS-GWO, Class-0 yields an accuracy of 91.34 with Class-1 a little lower with 89.82. Class-0 precision is 88.23 per cent and Class-0 recall is 93.12 per cent, which is good in recognizing Class-0. In the case of CNN, Class-0 has 92.59 percent accuracy, and its precision (81.53) is comparatively poor to other models. Class-1 recalls (96.25) are higher and the precision (85.26) is lower. The accuracy of Class-0 by the LR model is 89.53 with only slightly less accuracy of 88.61 in Class-1. Class-1 (81.63%), however, has better precision whereas Class-0 (94.36%) has high recall. In the **MLP** model, Class-0 accuracy is 88.85%, with Class-1 achieving a better recall (95.56%) and precision (85.21%). The **Proposed** model outperforms all others, with Class-0 achieving 96.09% accuracy, 94.12% precision, 98.46% recall, and 96.24% F1-score. For Class-1, it achieves 93.33% accuracy, 95.83% precision, 90.79% recall, and 93.24% F1-score. This comparison clearly demonstrates the proposed model as the most effective in classifying heart disease, achieving superior performance across all metrics.

Table 4: Performance metrics Analysis for MIMIC-III and UCI Heart Disease Datasets

Strategy	**MIMIC-III**		**UCI Heart Disease**	
	MCC	**NPV**	**MCC**	**NPV**
PS-GWO	83.35	86.98	65.54	73.64
CNN	87.74	81.55	71.13	67.61
LR	85.48	77.21	70.36	69.23
MLP	79.54	84.30	68.47	70.57
Proposed	**88.46**	**89.54**	**73.52**	**75.10**

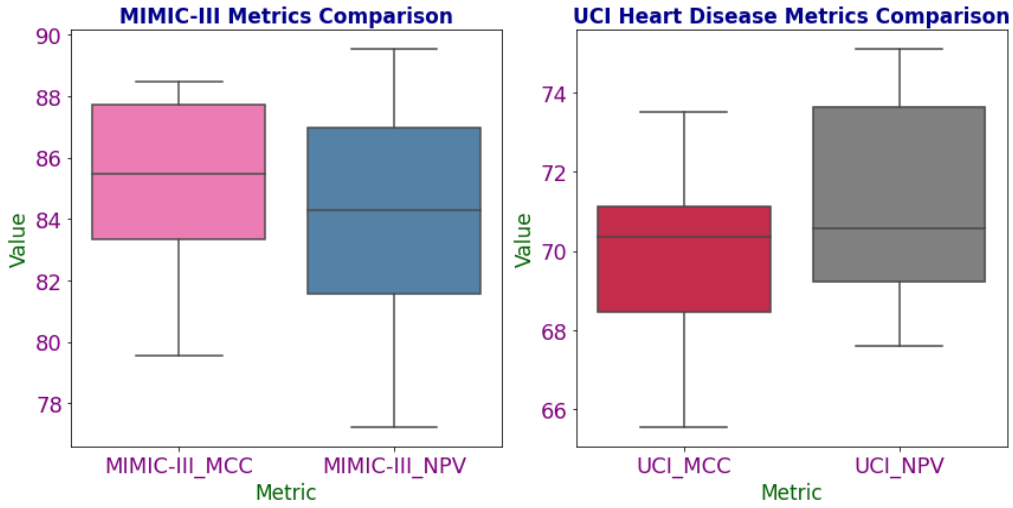

Figure 7: Performance metrics Analysis for MIMIC-III and UCI Heart Disease Datasets.

Table 4 and Fig.7 current the presentation metrics analysis of various strategies on the MIMIC-III and UCI Heart Disease datasets using MCC and NPV scores. On the MIMIC-III dataset, the suggested model achieved the highest MCC (88.46) and NPV (89.54), outperforming CNN (MCC: 87.74, NPV: 81.55) and LR (MCC: 85.48, NPV: 77.21). Likewise, the proposed model attained the highest MCC (73.52) and NPV (75.10) on the UCI dataset, followed by CNN (MCC: 71.13) and LR (NPV: 69.23). These outcomes demonstrate the effectiveness of the proposed strategy in predicting outcomes on both datasets.

Table 5: Execution Time for proposed model with other models

Strategy	Execution Time
PS-GWO	13.210
CNN	12.192
LR	12.019
MLP	13.132
Proposed	10.966

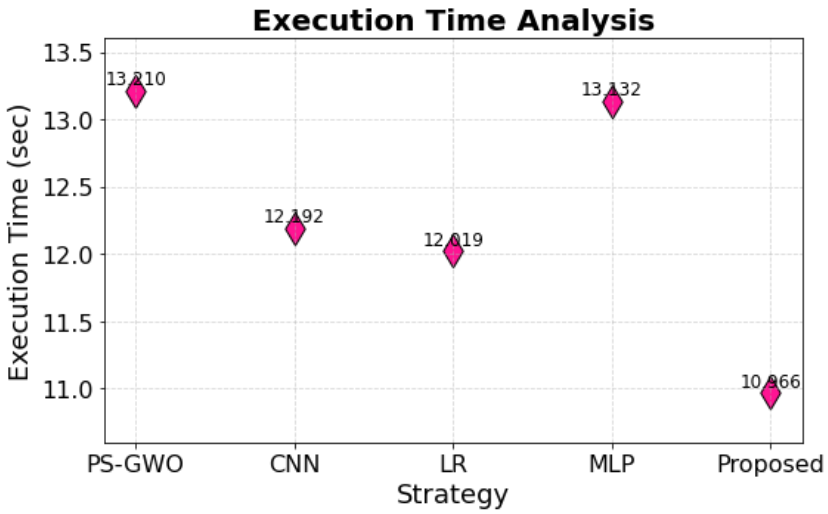

Figure 8: Execution Time for proposed model with other models

Table 5 and Fig.8 contain the execution time of different models of performance comparison. The given model had the smallest execution time of 10.966 seconds, which means that it was more computationally efficient. Conversely, PS-GWO and MLP had recorded more times of 13.210 and 13.132 seconds respectively, with CNN and LR recording average times of 12.192 and 12.019 seconds, respectively. These conclusions demonstrate the profit of the suggested model in decreasing the processing time.

Table 6: Correct & Incorrect Prediction by Sex and Age

Category	Prediction Label	Sex	Age	Final Prediction
Correct Predictions	0 - Correct	Female	42	Females predicted correctly
	1 - Correct	Male	20	Males predicted correctly
Incorrect Predictions	0 - Incorrect	Female	24	Females predicted incorrectly
	1 - Incorrect	Male	46	Males predicted incorrectly
Correct Predictions	0 - Correct	Female	32	Females predicted correctly
	1 - Correct	Male	64	Males predicted correctly
Incorrect Predictions	0 - Incorrect	Female	21	Females predicted incorrectly
	1 - Incorrect	Male	14	Males predicted incorrectly

Natural Resources for Human Health
eISSN: 2583-1194

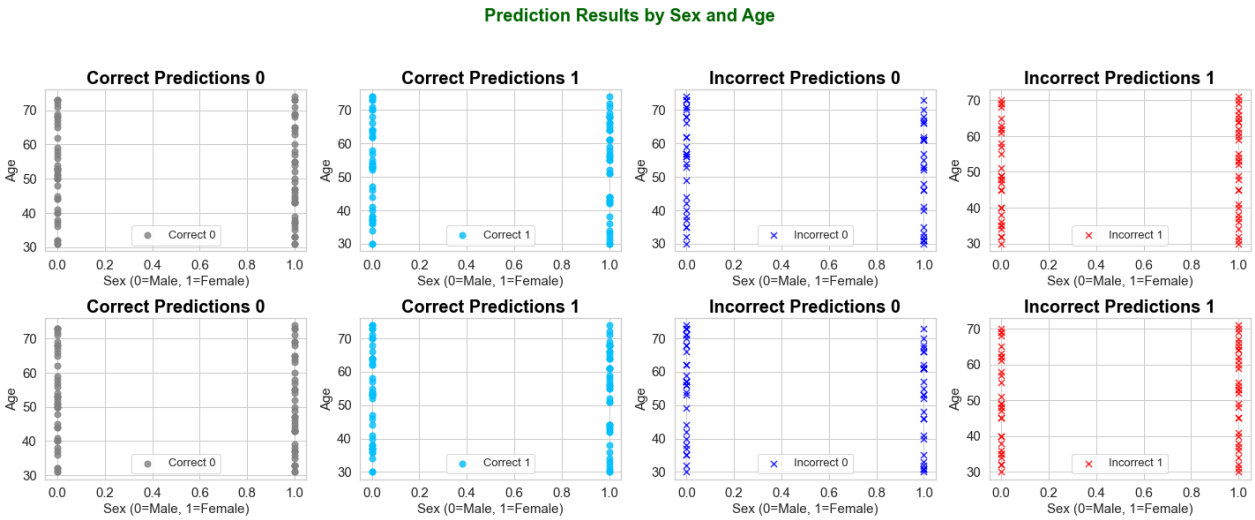

Figure 9: Correct & Incorrect Prediction by Sex and Age

Table 6, Fig.9 show the distribution of correct and incorrect predictions according to sex and age. The accurate prediction in the correct prediction group was females at the age of 42 and 32 and men at the age of 20 and 64. This means that there is a consistent performance among the young and older generations. On the other hand, the model also displayed false predictions: there were the wrong predictions in women of the ages 24 and 21, and men of the ages 46 and 14. These mistakes show that there are certain age groups on which model correction may be necessary, particularly among younger men and middle-aged women. In general, the model demonstrates a good level of predictive consistency between the sexes, although there are key areas in which one can make specific improvements in the age predictions.

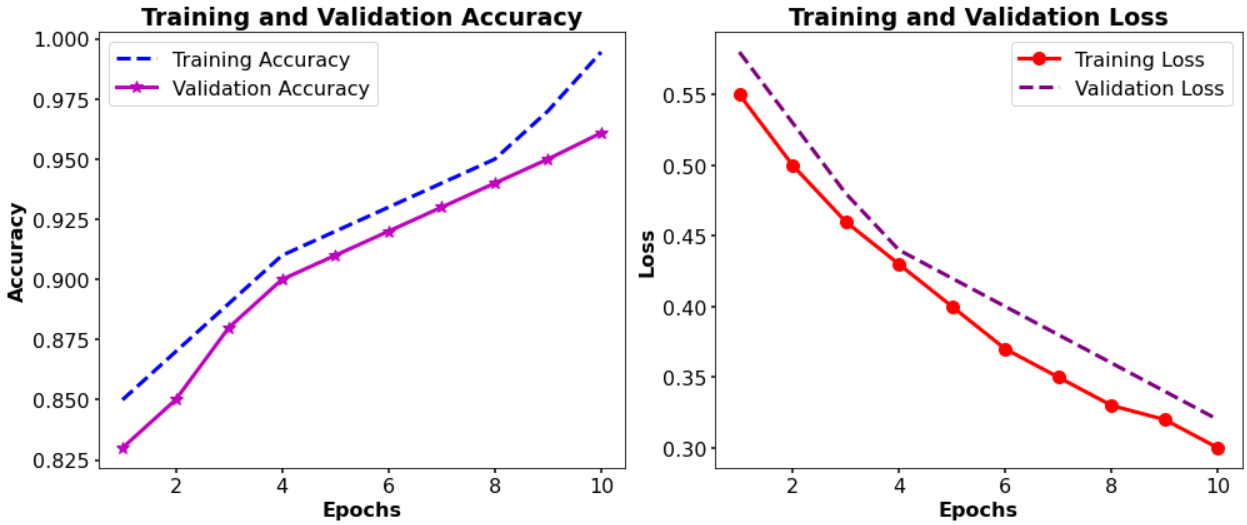

Figure.10 Training and Validation Accuracy & Loss

Fig.10 displays the perfect presentation through ten epochs of training and confirmation. The accuracy curve on the left shows good learning without overfitting, as both training and validation accuracies gradually rise to above 99.45% by the final epoch. The plot on the right shows the loss curves, where both training and validation losses consistently decrease, with the training loss dropping slightly faster. The close alignment of these curves suggests that the perfect is simplifying well to unseen information.

Table 7: Comparative Analysis for proposed model with other models

Reference	Method	Dataset	Accuracy
Saboor, et al. [22]	Machine learning Algorithms	UCI	91.80%
Taylan, et al. [23]	Machine learning Algorithms	UCI	94.7%
Mehmood, et al. [24]	CNN	UCI	86.67%
Hera, et al. [25]	Multitier Ensemble models	Cleveland, Switzerland, long beach Va Satlog	93.76%
Proposed	**Causal-GAT-LLM-LiNGAM**	**MIMIC-III and UCI Heart Disease Datasets**	**99.45%**

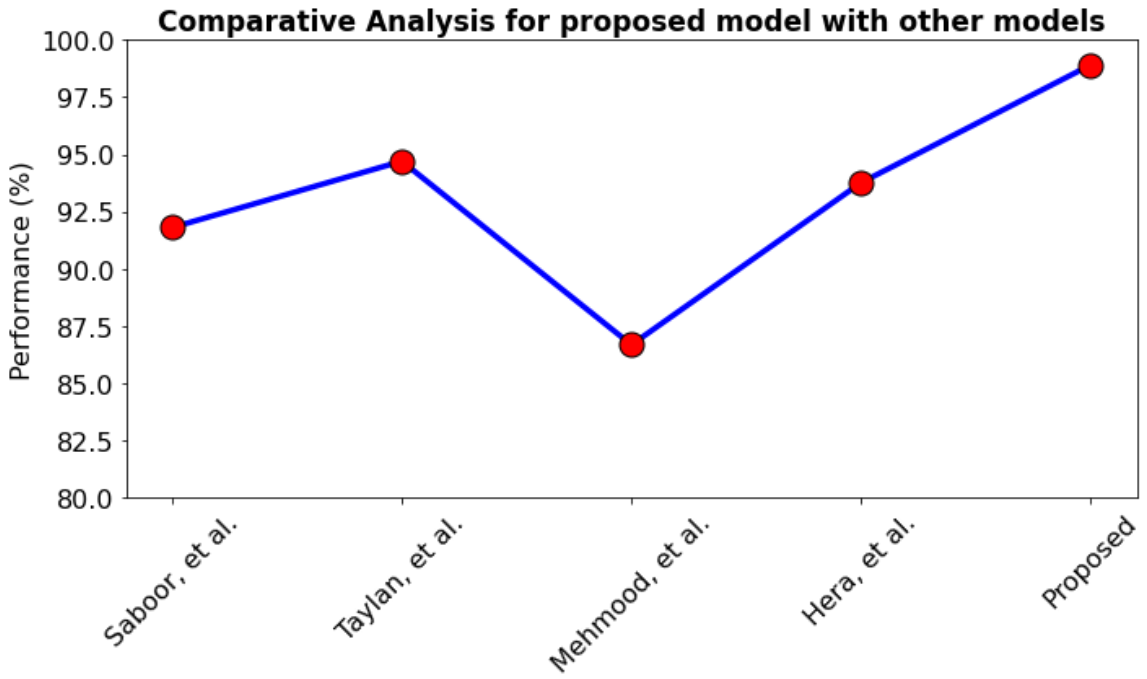

Figure 11: Comparative Analysis for proposed model with other models

Table 7 and Fig.11 compare the proposed Causal-GAT-LLM-LiNGAM model with the most advanced methods currently available for predicting heart disease. Saboor et al. [5] used methods including AdaBoost (AB), Logistic Regression (LR), CART, SVM, LDA, RF, and XGBoost to achieve 91.80% accuracy on the UCI dataset. Taylan et al. [6] applied SVR, ANFIS, and M5 Tree on an extended UCI dataset with 1,028 instances, achieving 94.7% accuracy. Mehmood et al. [7] employed a CNN-based model and obtained an accuracy of 86.67%. Hera et al. [8] reported that a loading collective comprising LR, RF, SGD, and an collective of GDC and ADA achieved 93.76% accuracy on multiple datasets, including Cleveland and Switzerland. In comparison, the causal inference and language modeling-based proposed Causal-GAT-LLM-LiNGAM framework outperformed existing studies significantly, achieving a high accuracy of 99.45% on both the MIMIC-III and UCI heart disease dataset with high predictive performance and stability across multiple datasets.

Natural Resources for Human Health
eISSN: 2583-1194

4.3 Statistical Analysis

Table 8: Summary Statistics distribution Analysis

Summary Statistics	Age	Chest pain type	BP	Max HR	Thallium
Mean	54.433333	3.1740741	131.34444	149.67778	4.6962963
Std Dev	9.1090665	0.95009	17.861608	23.165717	1.940659
Std Err Mean	0.5543601	0.0578206	1.0870229	1.4098206	0.1181047
Upper 95% Mean	55.52477	3.2879126	133.4846	152.45346	4.9288235
Lower 95% Mean	53.341897	3.0602355	129.20429	146.90209	4.4637691
N	270	270	270	270	270
N Missing	0	0	0	0	0

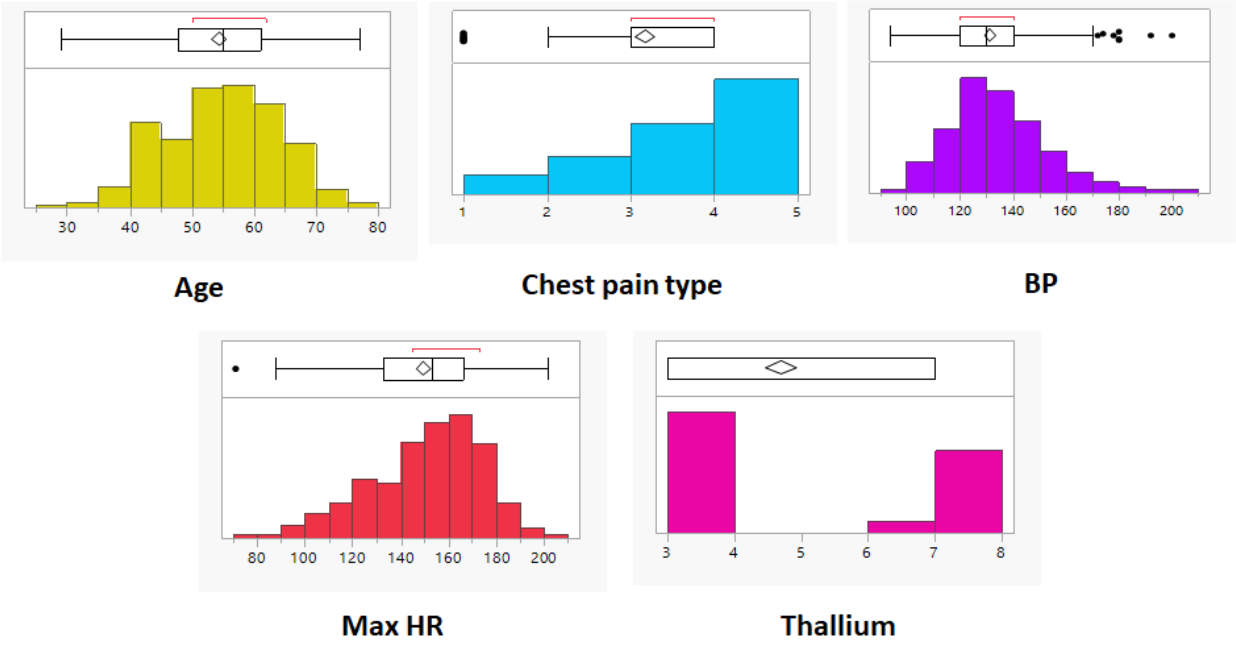

Figure 12: Summary Statistics distribution Analysis

The descriptive statistical approach is done in Table 8 and Fig.12 to analyze some significant clinical variables, which are as follows: Age, kind of chest pain, blood pressure (BP), maximum heart rate (Max HR), and thallium levels in a sample of 270 patients without any missing values. The age of the participants has a mean that is about 54.43 years and a standard deviation (SD) of 9.11 implying that the data is moderately concentrated along the mean. Type of chest pain has a smaller mean, which is 3.17 with SD of 0.95 and implies not as much divergence among patients. Mean BP is approximately 131.34 mmHg, with the highest heart rate averaging 149.68 bpm, where the ranges have a greater degree of variance (SD = 17.86 as well as 23.17 correspondingly), which represents physiological variability in the population. Thallium is usually utilized in stress testing and shows an average of 4.70 and SD 1.94. The values of standard error of the mean (SEM) that give the measures of the means between variables are relatively small meaning that there is precise measures of the mean. This exactness is further endorsed by the 95% confidence limits that provide a close range, of the upper and lower limits. These values all indicate the overall trends and range of significant cardiovascular measures proving goodness of the data available to perform the further investigation.

Natural Resources for Human Health
eISSN: 2583-1194

Table 9: Summary of fit analysis

Bivariate fit analysis	BP by Age	Cholesterol by Age	Max HR by Age	ST depression by Age
RSquare	0.074558	0.048425	0.161777	0.037727
RSquare Adj	0.071105	0.044874	0.15865	0.034136
Root Mean Square Error	17.21488	50.51324	21.24879	1.125494
Mean of Response	131.3444	249.6593	149.6778	1.05
Observations (or Sum Wgts)	270	270	270	270

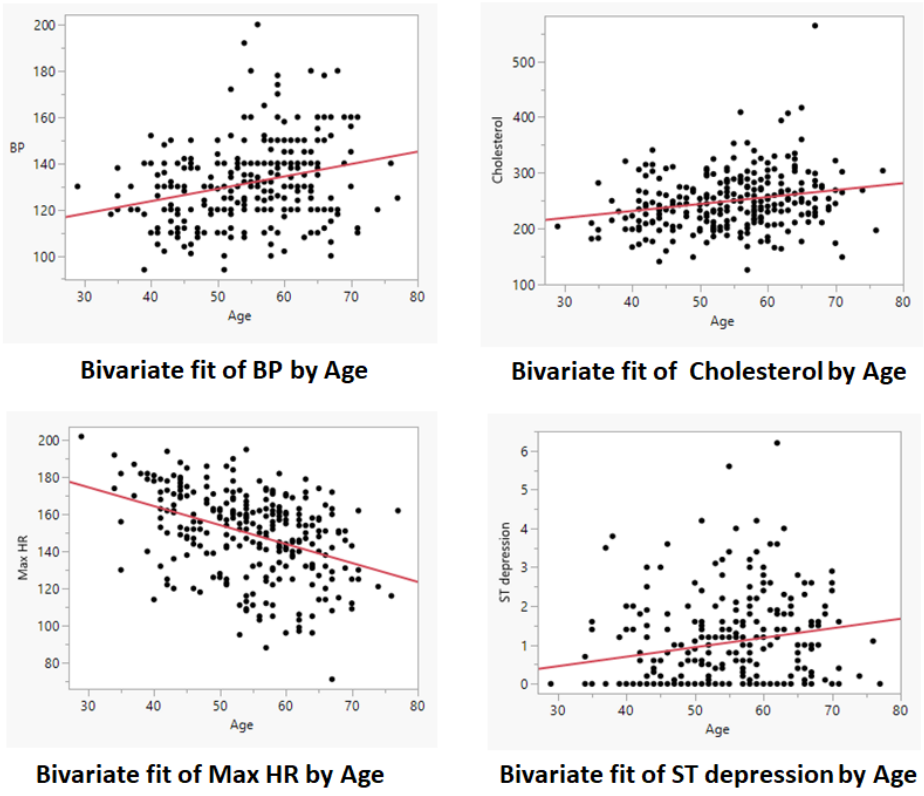

Figure 13: Summary statistics of Bivariate fit analysis

Table 9 and Fig.13 show bivariate fit analyses results of age and four clinical response variables, which include Blood Pressure (BP), Cholesterol, Maximum Heart Rate (Max HR), and ST Depression. The values of R-Square (coefficient of determination) show by how much extent age is able to explain the variance in each dependent variable. The R-Square value is the highest (0.1618) in Max HR by Age which means 16.2 percent of variability in Max HR is attributable to age. It means that there is a significant moderate correlation. On the contrary, the weakest match is achieved by BP by Age (R 2 = 0.0746), Cholesterol by Age (R 2 = 0.0484), and ST Depression by Age (R 2 = 0.0377); this indicates that age is not a strong determinant of these variables. The adjusted R-Square indexes and values (that take into consideration complexity of the model) are a little bit less, but they follow the same direction. Root Mean Square Error (RMSE) values give the arithmetic average of the mean match of the predicted values with the observed data where the greater the RMSE value the greater the error. As an illustration, Cholesterol has the maximum RMSE (50.51), which implies that the levels of cholesterol vary significantly regarding age. Clearly all the variables are founded on 270 observations and thus are statistically reliable. In general, the fit analysis demonstrates that age shows limited explanatory capacity on such clinical outcomes when the factor is examined independently.

Natural Resources for Human Health
eISSN: 2583-1194

Table 10: Fit Least Square Analysis

Source	DF	Sum of Squares	Mean Square	F Ratio
Chest Pain Type				
Model	0	0.00000	0.000000	.
Error	269	242.81852	0.902671	**Prob > F**
C. Total	269	242.81852		.
Cholesterol				
Model	0	0.00	0.00	.
Error	269	718624.65	2671.47	**Prob > F**
C. Total	269	718624.65		.
Max HR				
Model	0	0.00	0.000	.
Error	269	144358.97	536.650	**Prob > F**
C. Total	269	144358.97		.
Slope of ST				
Model	0	0.00000	0.000000	.
Error	269	101.54074	0.377475	**Prob > F**
C. Total	269	101.54074		.

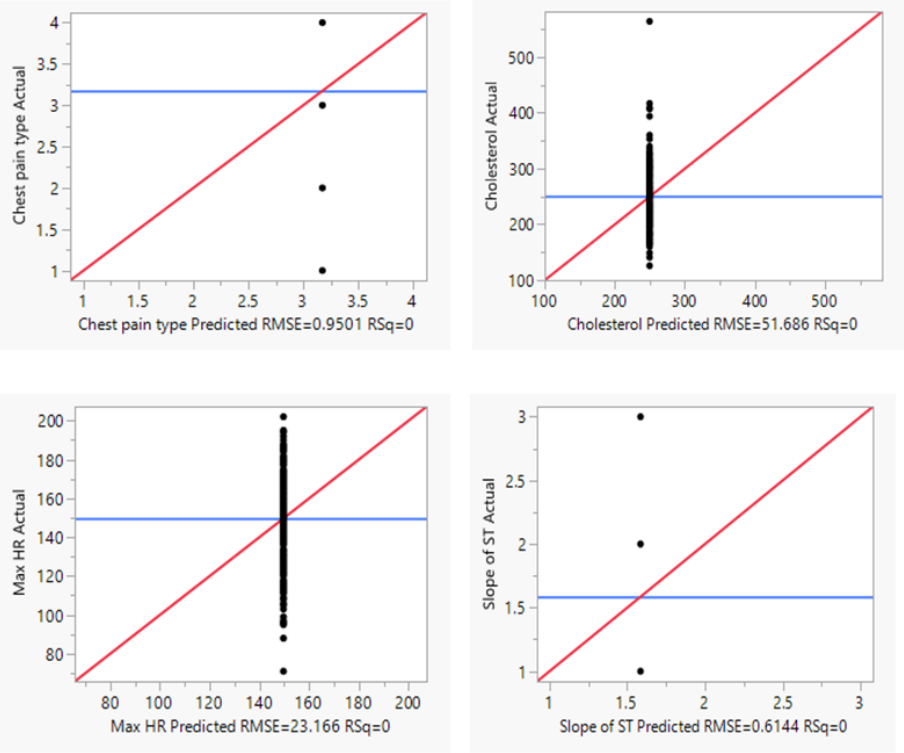

Figure 14: Fit Least Square Analysis

Table 10 and Fig.14 show the findings of a least squares regression analysis calculated with the aim of assessing the correlation between age and four clinical parameters chest pain type, cholesterol, the maximum heart rate (Max HR) and the slope of ST depression. The degree of freedom (DF) of the model has a value of 0 in all the four models, which implies that the model did not use age as a significant variable, or the effect of it was not statistically significant. This causes the model sum of squares to be zero as no part of any variability of the dependent variables is due to age. The error term contains all variability and the sum of the corrected totals of each variable is the sum of squares of errors. This is also supported by the non-determined ratio of F and the non-existence of p-value (Prob > F) which means that there is no struggle of significance of models test since they did not make any contribution to them. Further, the trends in charts provided support these figures: predictable trends are stable with no dispersion and are distributed near the mean (blue line), the same is true about actual trends which are dispersed widely. The values of root mean square error (RMSE) are quite large and the values of R- squared = 0 indicating that none of the data variance is explained by the models. All in all, this study has shown that age itself does not present an appreciable interaction or forecast of any of the four clinical along with a simple least squares regression scheme.

5. CONCLUSION

The Causal-GAT-LLM model that combines large language models, graph neural networks and causal inference to produce both accurate and interpretable results is a major milestone towards the prediction of cardiac disease. The system can infer causal relationships in structured clinical data using LiNGAM, and this is a good foundation of decision making other than correlation. The GAT component enhances the ability of the model to understand complex relationships among features to continue the spatial and relation inseparability of clinical variables. In addition, it is possible to add a Large Language Model (LLM) to the framework to enable it to process and analyze unstructured medical text data, which would enable a multimodal approach that would enhance the predictive capacity and situational awareness of patient records. This mixed approach is capable of improving the accuracy of classification in clinical applications where responsibility and reliability are required besides facilitating transparency through the provision of explainable results. Overall, Causal-GAT-LLM is a powerful, explainable, and scalable model to forecast heart diseases. The potential of its capability to combine causal discovery, deep learning, and natural language understanding is very high, and it can be applied into practice and offer more insights and data-based support to clinicians in the analysis and action of cardiovascular diseases.

CREDIT AUTHOR CONTRIBUTION STATEMENT

Amit Kumar: Conceptualization, Methodology, Investigation, Formal analysis, Writing: original draft, Writing: review and editing. **V.K. Jain:** Conceptualization, Methodology, Investigation, Formal analysis, Writing: original draft, Writing: review and editing. **Ajay Kumar:** Conceptualization, Methodology, Investigation, Formal analysis, Writing: original draft, Writing: review and editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI) in order to refine the language and improve editorial clarity in selected sections. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ETHICS APPROVAL AND CONSENT

The study was conducted in accordance with accepted ethical research standards.

REFERENCES

[1] Fekih, R.T.; Atri, P.M. Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. Comput. Biol. Med. 2020, 123, 103866. https://doi.org/10.1016/j.compbiomed.2020.103866
[2] Chao, C.; Peiliang, Z.; Min, Z.; Qu, Y.; Bo, J. Constrained transformer network for ecg signal processing and arrhythmia classification. BMC Med. Inf. Decis. Mak. 2021, 21, 184. https://doi.org/10.1186/s12911-021-01546-2
[3] Ahsan, M.M.; Siddique, Z. Machine learning-based heart disease diagnosis: A systematic literature review. Artif. Intell. Med. 2022, 128, 102289. https://doi.org/10.1016/j.artmed.2022.102289
[4] Ramesh, T.R.; Lilhore, U.K.; Poongodi, M.; Simaiya, S.; Kaur, A.; Hamdi, M. Predictive analysis of heart diseases with machine learning approaches. Malays. J. Comput. Sci. 2022, 1, 132–148. https://doi.org/10.22452/mjcs.sp2022no1.10
[5] Mohammad, F., & Al-Ahmadi, S. (2023). WT-CNN: a hybrid machine learning model for heart disease prediction. Mathematics, 11(22), 4681. https://doi.org/10.3390/math11224681
[6] Biswas, N., Ali, M. M., Rahaman, M. A., Islam, M., Mia, M. R., Azam, S., ... & Moni, M. A. (2023). Machine learning-based model to predict heart disease in early stage employing different feature selection techniques. *BioMed Research International*, *2023*(1), 6864343. https://doi.org/10.1155/2023/6864343
[7] Alshraideh, M., Alshraideh, N., Alshraideh, A., Alkayed, Y., Al Trabsheh, Y., & Alshraideh, B. (2024). Enhancing heart attack prediction with machine learning: A study at Jordan University Hospital. *Applied Computational Intelligence and Soft Computing*, *2024*(1), 5080332. https://doi.org/10.1155/2024/5080332
[8] Fatima, S., Purushothaman, R., Jeyalakshmi, M. S., Gupta, R., & Sivakumar, N. (2024). Heart Disease Prediction based on Convolutional Neural Network Feature Genetic Algorithm Solutions. *Frontiers in Health Informatics*, *13*(7).
[9] Abdullah, M. (2025). Artificial intelligence-based framework for early detection of heart disease using enhanced multilayer perceptron. *Frontiers in Artificial Intelligence*, *7*, 1539588. https://doi.org/10.3389/frai.2024.1539588
[10] Wang, Y., Rao, C., Cheng, Q., & Yang, J. (2024). Cardiovascular disease prediction model based on patient behavior patterns in the context of deep learning: a time-series data analysis perspective. *Frontiers in psychiatry*, *15*, 1418969. https://doi.org/10.3389/fpsyt.2024.1418969
[11] Kumar, S., & Thakur, B. (2024). Heart disease prediction using a stacked ensemble learning approach. *SN Computer Science*, *6*(1), 3. https://doi.org/10.1007/s42979-024-03499-5
[12] Khan, H., Javaid, N., Bashir, T., Akbar, M., Alrajeh, N., & Aslam, S. (2024). Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: EnsCVDD-Net and BlCVDD-Net. *IEEE Access*, *12*, 109230-109254. https://doi.org/10.1109/ACCESS.2024.3421241
[13] https://www.kaggle.com/datasets/asjad99/mimiciii
[14] https://www.kaggle.com/code/zeeshanyounas001/heart-disease-uci
[15] Xie, F., He, Y., Geng, Z., Chen, Z., Hou, R., & Zhang, K. (2022). Testability of instrumental variables in linear non-Gaussian acyclic causal models. Entropy, 24(4), 512. https://doi.org/10.3390/e24040512
[16] Shah, K., Xu, A. Y., Sharma, Y., Daher, M., McDonald, C., Diebo, B. G., & Daniels, A. H. (2024). Large language model prompting techniques for advancement in clinical medicine. Journal of Clinical Medicine, 13(17), 5101. https://doi.org/10.3390/jcm13175101
[17] Li, J., Hu, J., Wu, Y., & Yang, X. (2025). Pre-Routing Slack Prediction Based on Graph Attention Network. Automation, 6(2), 20. https://doi.org/10.3390/automation6020020
[18] Aliyar Vellameeran, F., & Brindha, T. (2022). A new variant of deep belief network assisted with optimal feature selection for heart disease diagnosis using IoT wearable medical devices. *Computer Methods in Biomechanics and Biomedical Engineering*, *25*(4), 387-411. https://doi.org/10.1080/10255842.2021.1955360
[19] Mehmood, A.; Iqbal, M.; Mehmood, Z.; Irtaza, A.; Nawaz, M.; Nazir, T.; Masood, M. Prediction of heart disease using deep convolutional neural networks. Arab. J. Sci. Eng. 2021, 46, 3409–3422. [Google Scholar] [CrossRef] https://doi.org/10.1007/s13369-020-05105-1
[20] Mridha, K., Kuri, A. C., Saha, T., Jadeja, N., Shukla, M., & Acharya, B. (2023, May). Toward explainable cardiovascular disease diagnosis: a machine learning approach. In International Conference on Data Analytics and Insights (pp. 409-419). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3878-0_35

[21] Nannuri S., Gantla H. R., Ananya D., Vennela A., Bhuvana B., Neha G. (2026). Automated Brain Tumor Segmentationin Mri Via U-Net-Based Deep Neural Networks. International Journal of Engineering Sciences & Research Technology, 15(4), 1–8. https://doi.org/10.64149/j.ijesrt.15.4.1-8

[22] Ali, M.M.; Kumar, B.P.; Ahmad, K.; Francis, M.B.; Julian, M.W.Q.; Moni, M.A. Heart disease prediction using supervised ML algorithms: Performance analysis and comparison. Comput. Biol. Med. 2021, 136, 104672. [Google Scholar] [CrossRef] [PubMed] https://doi.org/10.1016/j.compbiomed.2021.104672

[23] Alkhatib A. J., (2026). Integrated Computational Retinal Vascular Morphometry Using Fractal, Skeleton-Based, and Phi-Related Features for Diabetic Retinopathy and Glaucoma. International Journal of Artificial Intelligence, Machine Learning and Data Analytics, 1(1), 43-50. https://www.ijaimda.org/index.php/ijaimda/article/view/3

[24] Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A method for improving prediction of human heart disease using machine learning algorithms. *Mobile Information Systems*, *2022*(1), 1410169. https://doi.org/10.1155/2022/1410169

[25] Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukçu, E., & Leiva, V. (2023). Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology*, *12*(1), 117. https://doi.org/10.3390/biology12010117

[26] Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. *Arabian Journal for Science and Engineering*, *46*(4), 3409-3422. https://doi.org/10.1007/s13369-020-05105-1

[27] Hera, S. Y., Amjad, M., & Saba, M. K. (2022). Improving heart disease prediction using multi-tier ensemble model. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *11*(1), 41. https://doi.org/10.1007/s13721-022-00381-3