Hybrid Transformer-XGBoost Model Optimized with Ant Colony Algorithm for Early Heart Disease Detection: A Risk Factor-Driven and Interpretable Method

Moch Deny Pratama[1,*], Faris Abdi El Hakim[2], Dimas Novian Aditia Syahputra[3], Dodik Arwin Dermawan[4], Asmunin[5], Salamun Rohman Nudin[6], Andi Iwan Nurhidayat[7]

[1,2,3,4,5,6,7]*Department of Informatics Management, Universitas Negeri Surabaya, Surabaya, Indonesia*

(Received: May 5, 2025; Revised: July 1, 2025; Accepted: November 1, 2025; Available online: December 9, 2025)

Abstract

Cardiovascular Diseases (CVDs) remain the leading cause of death worldwide, with significant socioeconomic consequences due to premature death and chronic disability. Although clinical screening techniques have evolved, early and accurate prediction of heart disease is still partial due to the limited capacity of conventional machine learning algorithms to model the complex nonlinear interactions among various contributing risk factors e.g., hypertension, diabetes, hyperlipidemia, and genetic predisposition. To address these challenges, this research introduces a hybrid framework that combines the Transformer architecture known for its robust self-attention mechanism and high representational capabilities with Ant Colony Optimization (ACO), a nature-inspired metaheuristic algorithm modeled on the foraging behavior of ants, to enable adaptive and efficient hyperparameter optimization. This dataset is relatively balanced, with 55.34% of patients diagnosed with heart disease and 44.66% without heart disease. To ensure reliable evaluation and minimize the risk of overfitting, we implemented a nested cross-validation protocol, maintaining a consistent class distribution across folds. The proposed model processes structured data by encoding categorical variables into embeddings and normalizing features, resulting in a unified tabular representation suitable for transformer-based analysis. ACO improves model efficiency by optimizing key parameters e.g., embedding configuration, learning rate, and depth, reducing manual intervention and computational overhead. The proposed Hybrid Transformer-ACO model focuses on interpretable clinical features to provide actionable risk stratification. Model evaluation was performed using classification metrics e.g., accuracy, precision, recall, F1 score, and time complexity to measure predictive performance and computational efficiency during the training and inference phases. These evaluation criteria provide evidence of the model's diagnostic reliability, and potential feasibility for application. The model achieved average accuracy of 99.67% (±0.12), sensitivity of 99.59% (±0.18), specificity of 99.76% (±0.10), and an F1 score of 99.63% (±0.14). Time complexity analysis demonstrated efficient training and testing, while the model interpretability supports transparency and trust.

Keywords: Cardiovascular Disease, Heart Disease Detection, Transformer, Ant Colony Optimization, Machine Learning

1. Introduction

Cardiovascular Disease (CVD) continues to be a major public health problem worldwide, causing approximately 17.9 million deaths annually, representing approximately 31% of total global deaths [1]. The majority of these deaths are due to acute events e.g., myocardial infarction and stroke, with one-third occurring in individuals under the age of 70 years [2]. This premature loss of life has profound implications for public health systems and disproportionately affects the productive age group, with direct consequences for socioeconomic development and national productivity [3]. In addition, complications arising from CVD, e.g., Heart Disease, substantially increase the burden of disease both in terms of increased health care costs and reduced quality of life, ultimately impacting families and society at large [4]. In this context, appropriate and accessible early detection strategies are essential to reduce CVD-related morbidity and mortality and to support more efficient and evidence-based health care delivery [5]. Key risk factors for CVD including hypertension, diabetes mellitus, hyperlipidemia, and a family or personal history of cardiovascular conditions have long been established as key predictors of future cardiac events [6]. However, the effective integration of these risk variables into a reliable and interpretable prediction system remains a major challenge, especially given the increasing

***Corresponding author: Moch Deny Pratama (mochpratama@unesa.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i1.969

complexity and abundance of modern healthcare data [7]. Conventional Machine Learning (ML) methods, while useful in structured prediction tasks, often fail to capture the nuanced nonlinear interrelationships inherent in medical datasets [8]. These shortcomings have led to a shift in research focus towards deep learning and attention-based models that offer improved scalability and predictive performance.

In this way, originally designed for natural language processing tasks, Transformer-based models are increasingly being adopted in the healthcare domain due to their capacity to capture long-range dependencies and complex feature interactions in heterogeneous and high-dimensional clinical datasets [9]. Unlike recurrent or convolutional neural networks, transformers employ a self-attention mechanism that allows the model to assign differential importance to input features, thereby capturing critical risk interactions that simpler architectures might overlook [10]. Furthermore, the interpretability of transformer models, particularly through visualization of attention, promises to increase transparency and clinician confidence in high-stakes decision-making [11]. Several studies have demonstrated the efficacy of deep learning and hybrid techniques for CVD risk prediction using imaging or wearable data [12]. However, most of these approaches require high-resolution medical imaging, specialized hardware, or large-scale labeled datasets, which limit their generalizability and practical implementation in primary care or resource-constrained settings [13]. In contrast, transformer models applied to structured tabular data e.g., demographic and clinical parameters are under-explored, despite their abundance and easy accessibility in Electronic Health Records (EHRs) [14]. Furthermore, while these models demonstrate robust performance, the need for computationally efficient optimization methods to manage their complexity remains a critical issue [15]. To address this gap, recent research has begun to integrate metaheuristic optimization techniques e.g., Ant Colony Optimization (ACO) with deep learning architectures. ACO, inspired by the foraging behavior of ant colonies in finding optimal paths, offers a biologically inspired solution for parameter tuning in high-dimensional spaces [16]. When applied to transformer-based models, ACO can improve model performance by optimizing dimension, learning rates, and other key parameters without extensive manual tuning or brute-force search strategies [17]. Despite its potential, the application of ACO-transformer hybrids in healthcare specifically in CVD prediction using tabular clinical data has received limited attention in the literature [18].

Given the limitations of current predictive models and the underutilization of structured clinical data, there is an urgent need for a scalable and interpretable framework that leverages widely available health records while maintaining high predictive accuracy [19]. Transformer models, with their ability to learn contextual relationships across multivariate inputs . However, the effectiveness of these architectures relies heavily on careful tuning of hyperparameters such as embedding configuration, learning rate, network depth, and embedding dimensionality [21]. Manual hyperparameter selection is often labor-intensive and suboptimal, especially in the healthcare domain where model generalization, computational feasibility, and clinical reliability are critical [22]. To address this, ACO a biology-inspired metaheuristic algorithm, is used to dynamically explore the hyperparameter search space and identify high-performing configurations through iterative pheromone-based reinforcement learning [23]. Therefore, the integration of modeling addresses two important challenges e.g., improving the performance of predictive models and ensuring computational feasibility for real-world applications [24]. This research introduces introduces a novel hybrid predictive framework that integrates the representational power of Transformer models with the adaptive tuning capabilities of ACO, specifically applied to the domain of early detection of heart disease. Cardiovascular risk factors e.g., age, gender, hypertension, diabetes, cholesterol, and exercise capacity are widely recognized as key determinants of heart disease. In the dataset used in this study, these variables exhibited a diverse distribution; e.g., 21.02% female and 78.98% of patients were male, 23.31% had diabetes, and 54.03% had asymptomatic chest pain, which constituted the largest subgroup. Resting electrocardiograms were mostly normal 60.13%, while 40.41% of patients reported exercise-induced angina. Further correlation analysis highlighted significant relationships, such as a negative correlation between age and maximum heart rate $r = -0.38$, reflecting a decline in cardiac responsiveness with age, and a positive correlation between age and ST depression $r = 0.26$. To mitigate the potential influence of multicollinearity among predictors, we relied on the regularization mechanism inherent in XGBoost and the attention mechanism of the Transformer architecture, both of which naturally reduce the weight of redundant or highly correlated features. These characteristics ensure that the predictive contribution of each variable remains clinically interpretable and statistically reliable. The proposed workflow combines comprehensive preprocessing, embedding-based feature transformation for categorical variables, and iterative optimization cycles guided by ACO. The training and evaluation pipeline includes baseline and optimized

hybrid models, where classification performance is assessed through standard metrics such as Accuracy, Precision, Recall, and F1-Score. Additionally, Time Complexity is introduced as a critical computational metric, measuring the efficiency of the training and testing phases to ensure operational feasibility for clinical applications. Confusion matrices provide detailed insights into misclassification patterns and model robustness. The model's reliance on structured tabular data and its architecture optimized for efficiency make it well-suited for implementation, including in resource-constrained healthcare settings that lack access to advanced imaging or sensor technologies.

2. Literature Review

This chapter provides a comprehensive review of the existing literature relevant to CVD prediction, deep learning architectures particularly Transformer models and hyperparameter optimization techniques such as ACO. The discussion begins with an overview of the clinical significance and risk factors for cardiovascular disease, followed by an examination of conventional machine learning models and their limitations in modeling complex nonlinear interactions in medical data. Further, recent advances in transformer-based models for structured data are explored, emphasizing their mechanisms and representational power. The chapter also highlights the role of metaheuristic algorithms, with a focus on ACO, in addressing the challenges of hyperparameter tuning in high-dimensional search spaces. Through this literature synthesis, gaps in existing methodologies are identified, justifying the need for an interpretable, scalable, and optimized hybrid framework for early CVD prediction using tabular clinical data.

2.1. Cardiovascular Risk Detection

Several studies have addressed the challenges of early detection of CVD by leveraging conventional machine learning models e.g., Support Vector Machines (SVM), Logistic Regression, and Decision Trees. The Research conducted by [25] applied hybrid methods for CVD classification using structured clinical data, yielding moderate accuracy but limited interpretability with an accuracy of 90.94%. Research conducted by [26] explored Random Forest and Naïve Bayes on table datasets, but their models often suffer from feature interaction limitations and overfitting. Recently, deep learning architectures have been explored to address the complexity of healthcare data. To address these limitations, recent research has explored the application of Transformer-based architectures, as an alternative. These models have shown success in capturing long-range dependencies and complex non-linear relationships in tabular and multimodal medical datasets [27]. Research by [28] applied tabular transformers to clinical data, which showed improved performance over conventional ML models and iterative models. However, computational complexity and hyperparameter sensitivity remain major challenges, especially when applied in real-world clinical settings.

2.2. Heart Disease Detection

Heart disease detection using artificial intelligence and machine learning has become a prominent research focus due to the high global prevalence and mortality rate of cardiovascular conditions [29]. Many studies have used traditional classification algorithms e.g., SVM, k-Nearest Neighbors (k-NN), Random Forest (RF), and Logistic Regression (LR), to estimate heart disease risk using structured datasets, including publicly available heart disease datasets [30]. The Research conducted by [31] using the stacked model method k-NN, random forest, and SVM proved to be most effective with accuracy of 75.10%. The research [32] was conducted using machine learning with an accuracy of 83%. While these models offer interpretability and low computational overhead, their performance is often suboptimal when dealing with non-linear and high-dimensional relationships between multiple risk factors. The research conducted by [33] using the transfer learning transformer obtained an accuracy of 99.44%. The research conducted by [34] obtained an accuracy of 85% using the Feature Extraction and Artificial Neural Network (ANN) technique to identify scent patterns. The research conducted by [35] applied an hybrid learning networks, reporting significant improvements in accuracy, obtained an accuracy of 85.7%. The research conducted by [36] results show an accuracy level of 88.7% used the Hybrid Model of random forest and decision tree. The research conducted by [37] on the topic of heart failure cleveland data obtained 100% accuracy. The research conducted by [38] using ML and oversampling and standardscaling obtain 90% accuracy with a computational time cost of 65,935 in the training stage and 3,558 in the testing phase. The research conducted by [39] obtained 99.2% accuracy using SMOTE sampling and hyperparameter optimization. Consequently, there is growing interest in developing models that leverage readily available tabular clinical data while maintaining high accuracy and interpretability. Hybrid Transformer-based models especially when optimized through metaheuristic algorithms, e.g., ACO provide a new opportunity to bridge this gap, enabling accurate,

scalable, and interpretable heart disease detection systems that can be deployed across a wide range of healthcare infrastructures.

2.3. Optimization of Deep Learning Models Using Metaheuristic Algorithms

Optimizing many parameters in deep learning models, especially Transformer, is computationally intensive and often requires heuristic or brute-force approaches [28]. Metaheuristic algorithms e.g., Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and ACO have emerged as alternatives to efficiently tune complex models [40]. ACO, inspired by the foraging behavior of ants, has been successfully applied to neural network training, feature selection, and hyperparameter tuning [41]. The integration of ACO with Transformer-based models for healthcare data analysis remains largely unexplored. Research conducted by [42] has shown that integrating ACOs can significantly improve model generalization and convergence times, but the research has mostly focused on text classification and big data analysis.

2.4. State of the Art

The current state of the art in CVD risk prediction and heart disease detection increasingly favors deep learning models, especially those capable of handling complex feature interactions and high-dimensional data also identified in Section 2.1 and Section 2.2. Conventional approaches e.g., Logistic Regression, SVM, and hybrid techniques while effective for simple classification tasks often suffer from the nonlinearity and heterogeneity present in real-world health records [31]. Furthermore, many high-performing models rely on data derived from medical imaging or wearable sensors, which, while informative, require sophisticated infrastructure, high computational resources, and domain-specific data that may not be universally accessible [34]. In contrast, this research proposes a cardiovascular disease espsecially heart disease detection framework based on a hybrid Transformer model optimized using the ACO algorithm. Unlike previous works that rely heavily on high-resolution images or hardware-based sensor data, the proposed approach leverages structured tabular data e.g., clinical metrics and demographic features that are cost-effective and widely available in electronic health records [43]. The self-attention mechanism of the Transformer architecture facilitates modeling of complex nonlinear feature relationships more effectively compared to traditional machine learning approaches [14]. Furthermore, the integration of ACO for hyperparameter optimization overcomes the major computational bottleneck of deep learning models. This novel combination ensures improved model scalability and efficiency without incurring the high computational costs typically associated with imaging-based deep learning models. Thus, the proposed methodology not only improves predictive performance but also contributes to the development of interpretable and deployable AI tools suitable for resource-constrained healthcare settings.

3. Methodology

This chapter outlines the methodological framework adopted to develop, optimize, and evaluate the proposed hybrid model for early heart disease prediction. The research methodology is systematically structured to cover each critical phase of the workflow, starting from data acquisition and preprocessing, followed by feature engineering, model construction, hyperparameter optimization, training, and finally, performance evaluation. The integration of the Transformer architecture with ACO is explained in detail, along with the rationale for the selection of evaluation metrics and comparative baselines. Each component of the methodology is designed to ensure reproducibility, scalability, and clinical relevance in real-world healthcare settings.

The workflow as shown in figure 1, represents a comprehensive methodology for building and evaluating a Hybrid Transformer Model optimized using ACO for early heart disease prediction. The process begins with Resource Data Acquisition, where raw heart disease data is collected. This stage is followed by a Data Preprocessing phase involving several subtasks, including data loading, label separation, feature type identification, visualization, and initial feature engineering. These preprocessing steps produce a clean and structured dataset, followed by feature engineering to transform numeric and categorical variables into a format compatible with Transformer-based architectures. Categorical features are converted into embeddings using the token embedding technique, and the embeddings of all relevant columns are stacked into a unified representation, forming the final feature embedding format. This transformation allows the model to leverage the transformer's inherent attention mechanism to capture complex inter-feature relationships. Simultaneously, the Optimization Algorithm module defines the hyperparameter search space

ISSN 2723-6471

and initializes the ACO-specific pheromone matrix. The ACO process simulates an ant-based sampling path through the hyperparameter space to iteratively find the configuration that maximizes model accuracy. For each configuration, cross-validation is performed to evaluate performance, and if improvement is found, the pheromone trail is updated. The optimization process is repeated until the optimal configuration is determined based on validation accuracy, after which model training is run using the selected parameters. This includes training both baseline and hybrid with refined parameter settings. During training, time complexity monitoring is implemented to track computational costs at the training and inference stages, which is critical for feasibility in real-world deployments. The output of this process is a model with the best learned parameters and probabilistic output predictions. Next, model pipeline construction is performed, integrating the transformer model with standard preprocessors e.g., StandardScaler, and applying the configurations in a unified training-inference pipeline detailed anti-leakage pipeline; scalers, and encoders, are fitted on the training split and applied to validation/test splits using scikit-learn pipeline logic. This pipeline ensures model reproducibility and consistency during deployment. The model performance evaluation module calculates classification metrics e.g., accuracy, sensitivity, specificity, and f1 score, along with time complexity, to evaluate computational efficiency. Collectively, these metrics provide a holistic assessment of the model diagnostic reliability and scalability. Finally, Comparative Model Analysis combines the performance results of multiple models e.g., Transformer-ACO with an XGBoost, LightGBM, and Random Forest, providing a consolidated summary and interpretation of the metrics. This step enables evidence-based decision-making regarding the best model configuration for clinical applications. Overall, this flowchart outlines a robust and methodologically rigorous workflow that integrates deep learning, metaheuristic optimization, and structured clinical data, facilitating interpretable and efficient cardiovascular risk prediction suitable for real-world healthcare environments.

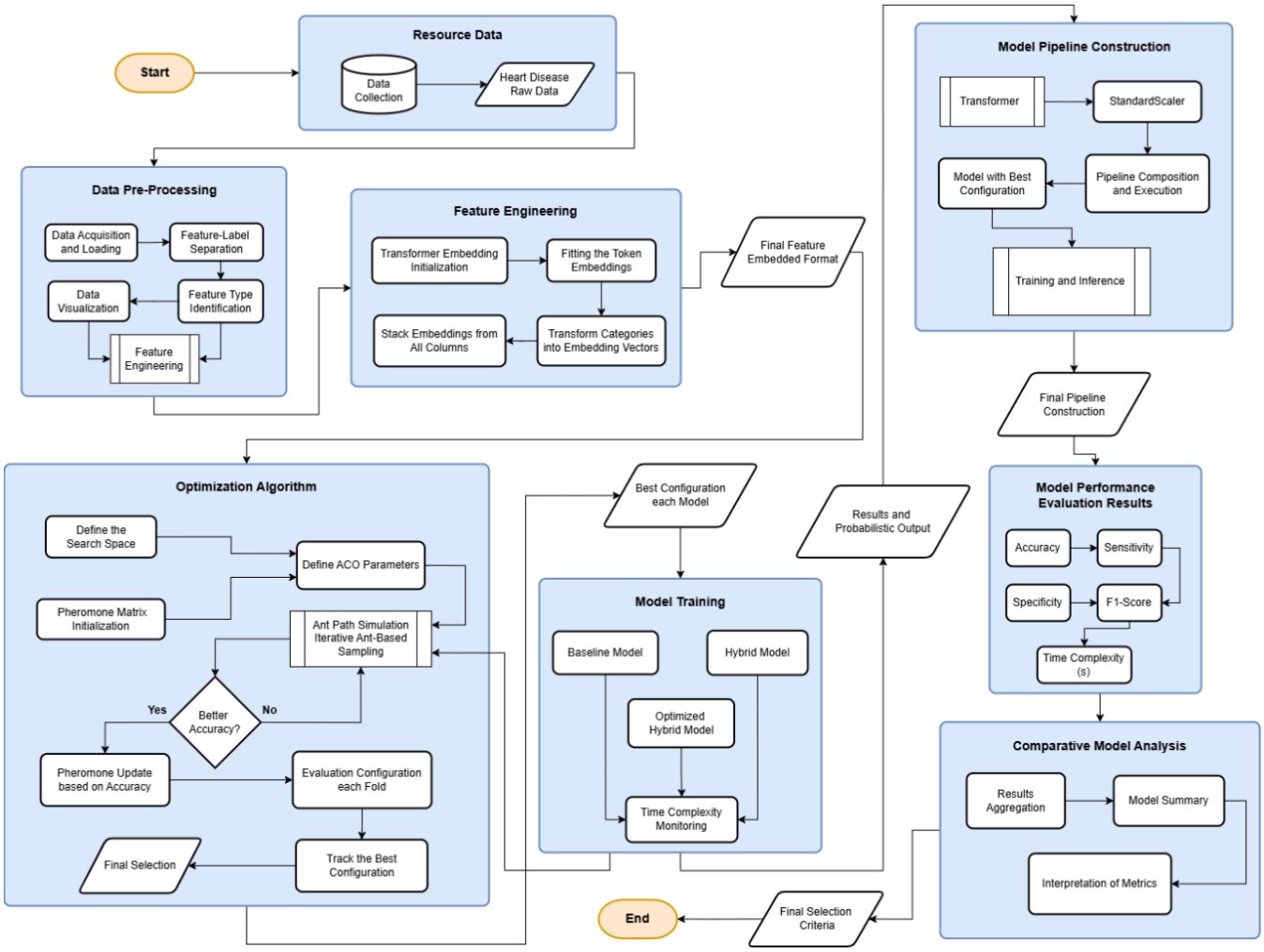

Figure 1. Research Flow Diagram

An AI-based early detection system for CVD is proposed to address the challenges of analyzing complex multidimensional health data involving risk factors such as hypertension, diabetes, and hyperlipidemia. The system utilizes a Transformer-based model capable of capturing complex patterns in heterogeneous tabular data, outperforming traditional ML approaches. The workflow includes structured data collection, preprocessing e.g., denoising, feature selection, normalization, and data splitting for robust evaluation. The core innovation integrates ACO to dynamically tune Transformer hyperparameters, reducing manual configuration, and improving efficiency. A comparative evaluation was performed against a baseline Transformer model and a traditional machine learning model using classification metrics such as accuracy, precision, recall, F1 score, and time complexity. This interpretable and scalable approach improves early detection of cardiovascular disease and facilitates clinical decision-making in resource-constrained settings.

3.1. Dataset and Pre-processing

This research uses the publicly available Heart Failure Prediction dataset from Kaggle, consisting of 918 patient records with 11 clinical features and one binary target variable. This dataset is relatively balanced, with 55.34% of patients diagnosed with heart disease and 44.66% without heart disease. To ensure reliable evaluation and minimize the risk of overfitting, we implemented a nested cross-validation 5-fold protocol, maintaining a consistent class distribution across folds; as shown in table 1. Numeric features include variables e.g., age, resting blood pressure, cholesterol level, maximum heart rate, and ST depression (Oldpeak), while categorical features include gender, type of chest pain, resting ECG, and ST slope. Binary features e.g., fasting blood sugar and exercise angina are also present.

Table 1. Dataset of Heart Disease Data

Age	Sex	Chest Pain Type	Resting BP	Cholesterol	…	Oldpeak	ST Slope	Heart Disease
62	M	ASY	158	210	…	3	Down	1
55	M	NAP	136	245	…	1.2	Flat	1
75	M	ASY	136	225	…	3	Flat	1
40	M	NAP	106	240	…	0	Up	0
67	M	ASY	120	0	…	1.5	Down	1
58	M	ASY	110	198	…	0	Flat	1
60	M	ASY	136	195	…	0.3	Up	0
…	…	…	…	…	…	…	…	…
35	M	NAP	123	161	…	-0.1	Up	0

To ensure data quality and model readiness, a structured preprocessing workflow was implemented. This process begins with handling missing or implausible values (e.g., zero values for cholesterol or fasting blood sugar). Categorical features were encoded appropriately, was applied to nominal variables e.g., ChestPainType, RestingECG, ST_Slope, while binary variables e.g., Sex, ExerciseAngina were encoded with labels. To support uniform feature scaling and optimize convergence in Transformer-based models, continuous variables were normalized using MinMaxScaler or standardized using StandardScaler. The outer loop uses 5-fold nested validation for unbiased model performance estimation, while the inner loop uses 5-fold cross-validation for hyperparameter tuning. This design ensures a clear separation between model selection and performance evaluation, thereby reducing optimism bias in reported metrics. Each feature was then treated as a token and embedded, forming a multidimensional tensor suitable for Transformer input. This preprocessing ensures clean, consistent, and semantically structured data for the Hybrid Transformer-ACO model. Furthermore, three clinically significant features RestingBP (hypertension), FastingBS (diabetes), and Cholesterol (lipid profile) are highlighted as key cardiovascular risk indicators. Elevated values in RestingBP ≥140 mmHg, FastingBS ≥126 mg/dL is defined as a binary indicator where 1 indicates fasting blood sugar >120 mg/dL and 0 otherwise, and Cholesterol ≥240 mg/dL indicate the presence of hypertension, diabetes, and hyperlipidemia, respectively, making them crucial components in the model's risk stratification framework and early heart disease prediction.

3.2. Hybrid Transformer-Based Model Architecture

The proposed model adopts a hybrid architecture hybrid-based that integrates a Transformer-based feature extractor with classifiers e.g., XGBoost, LightGBM, and Random Forest, optimized through ACO. The architecture is specifically designed to process tabular data, TabTransformer model is configured with four encoder blocks, each containing eight attention heads. Categorical features are embedded into 32-dimensional vectors and projected into a 64-dimensional space for multi-head attention, while numeric features are standardized and passed through a linear projection layer. Each feedforward layer uses 128 hidden units with a dropout rate of 0.1 to reduce the risk of overfitting. Training is performed for up to 100 epochs with a batch size of 32, using early stopping with a patience of 5 epochs. Meanwhile, numeric features are standardized using StandardScaler and concatenated with the output of the attention mechanism, forming a unified feature representation. The concatenated feature vector is then fed into a downstream classifier, which can be XGBoost, LightGBM, or Random Forest, depending on its configuration. The interpretability of the proposed framework is assessed using the intrinsic mechanisms of the XGBoost and Transformer models. XGBoost provides feature importance is used for the gradient boosting component, is reported using the gain metric, which reflects the relative contribution of each feature to the predictive performance and importance based on gain strengthens these findings, with Age and Chest Pain Type dominating the separate contributions. while the Transformer architecture inherently generates attention weights that highlight feature-level contributions. These two complementary perspectives allow us to identify clinically meaningful variables without the need for additional external interpretability frameworks.

Model hyperparameters including the number of estimators, tree depth, learning rate, and subsampling ratio are dynamically optimized using ACO, which simulates the foraging behavior of an ant colony to efficiently explore the high-dimensional parameter space. The ACO algorithm iteratively refines the model configuration through pheromone-guided probabilistic selection and evaporation mechanisms. For each candidate solution (ant), the model performance is validated using nested cross-validation and the average accuracy is used as the fitness score. The solution with the highest score is selected as the optimal hyperparameter set. During training, the hybrid model is evaluated based on its convergence speed, predictive performance, and computational efficiency. The final model evaluation is performed on unseen data using Accuracy, F1-Score, Sensitivity, Specificity, and Time Complexity with training and inference durations measured to ensure its feasibility for real-time healthcare applications. This hybrid architecture enhances representation capacity and interpretability through attention mechanisms, while addressing hyperparameter tuning challenges through biology-inspired optimization, offering a robust and scalable solution for early detection of heart disease using structured clinical data.

$$x_{cat} = [x_1, x_2, \dots, x_n]; ei \in R^d; E \in R^{c \cdot d}; z_i = E(x_i); Z_{cat} = [z_1, z_2, \dots, z_n] \quad (1)$$

Transformer embedding as shown in Eq. (1), the categorical input features are represented as $x_{cat} = [x_1, x_2, \dots, x_n]$, where each x_i denotes the categorical value of the i-th feature. Each categorical token is mapped into a dense vector representation $ei \in R^d$, where d is the embedding dimension. The embedding matrix $E \in R^{c \cdot d}$ contains learnable parameters for all c unique categorical values across the dataset. Through an embedding lookup operation, each token is converted into its corresponding vector using $z_i = E(x_i)$, resulting in an embedding vector z_i for each categorical feature. The complete set of categorical embeddings is then organized into a sequence $Z_{cat} = [z_1, z_2, \dots, z_n]$, which serves as the input to the Transformer encoder. This sequential representation enables the model to learn contextual interactions categorical features, thereby enhancing its representational capacity and performance on tabular data [44].

$$\hat{y}i = \sum_{k=1}^{k} f_k(x_i); \;\; f_k \in F \quad (2)$$

XGBoost objective function as shown in Eq. (2), the estimated output for the i-th data instance is expressed as $\hat{y}i = \sum_{k=1}^{k} f_k(x_i)$, where each $f_k \in F$ represents an individual regression tree selected from the function space F. $\hat{y}i$ denotes the final predicted value, x_i is the input feature vector for the i-th observation, and K indicates the total number of boosting rounds or trees used in the ensemble. Each function f_k models a weak learner that contributes incrementally to the final prediction through additive learning. The function space F is defined as a set of decision trees parameterized

by their structure and leaf weights. This formulation allows to construct a strong model by sequentially adding trees that correct the residual errors of previous models, thereby enhancing accuracy through gradient boosting [45].

3.3. Hyperparameter Search Space

To improve training efficiency and generalization in Transformer-based hybrid models, ACO is used as a metaheuristic for hyperparameter tuning. Modeled after the pheromone-driven foraging behavior of ants, ACO allows for iterative exploration of a discretized hyperparameter space, guided by historical performance results. The defined search space includes common parameters (e.g., embedding dimension, max_depth, n_estimators) and advanced parameters (e.g., learning_rate, subsample, colsample_bytree, min_child_samples), which are well-suited for hybrid gradient boosting algorithms such as XGBoost and LightGBM. In each iteration, candidate configurations are sampled, and model training is performed using a structured pipeline that includes TabTransformer for categorical embedding, StandardScaler for numerical normalization, and the selected hybrid learner. Performance is evaluated through nested cross-validation, with the average accuracy across folds serving as the fitness function for pheromone updates. Pheromone evaporation mitigates premature convergence, while reinforcement supports high-performance configurations. This process enables ACO to efficiently navigate complex search spaces, offering a scalable and automated alternative to grid search. The configuration achieving the highest validation accuracy is selected for implementation, resulting in a hybrid model optimized for ACO that combines predictive accuracy with the computational efficiency essential for real-time clinical decision support systems.

$$\tau_{ij} = (1-\rho)\cdot\tau_{ij} + \Delta_{\tau_{ij}}, \quad \eta_{ij} = \frac{1}{CV\ \varepsilon\ ij} \tag{3}$$

Pheromone update mechanism and heuristic matrix η as shown in Eq. (3), where τ_{ij} denotes the current pheromone level on component (i,j) where $\rho \in (0,1]$ is denoted the pheromone evaporation rate, and $\Delta_{\tau_{ij}}$. η denotes the heuristic matrix which computed as the inverse of the mean cross-validation error (ε) for each candidate config. The quantity of pheromone deposited on a selected component by an ant is determined by the quality of the locally constructed solution, thereby reinforcing components associated with more favorable outcomes. Following the completion of solution construction by all ants, a global pheromone update is executed, wherein only the best-performing solutions contribute to reinforcing the pheromone trails, guiding the search toward high-quality regions in the solution space. This iterative process balances exploration and exploitation, enabling ACO to approximate near-optimal solutions efficiently [46].

The predefined hyperparameter search space used for optimization by the ACO algorithm as shown in table 2, was selected based on established practices in previous literature and preliminary experimental testing. This search space includes architectural and learning-related parameters critical to the performance of the Hybrid Transformer model. The embedding dimensionality is limited to two discrete values (32 and 64), represents a commonly adopted size that balances representational capacity and computational cost. For the gradient boosting classifier, key parameters such as max_depth (ranging from 3 to 9), n_estimators (100 to 300 in 50-step increments), and learning_rate (ranging from 0.01 to 0.2) is consistent with configurations frequently applied in prediction studies, ensuring adequate model expressivity without excessive overfitting.

Table 2. Hyperparameter Optimization Search Space

Parameter	Search Space
embedding dimension	32, 64
max_depth	3 – 9
n_estimators (XGBoost); num_leaves (LightGBM)	100, 150, 200, 250, 300 (step size: 50)
learning_rate	0.01, 0.03, 0.05, 0.075, 0.1, 0.2
subsample (XGBoost); bagging_fraction (LightGBM)	0.6, 0.8, 1.0
colsample_bytree (XGBoost); feature_fraction (LightGBM)	0.6, 0.8, 1.0
min_child_samples (LightGBM); min_child_weight (XGBoost)	5, 10, 15, 20, 30

n_iterations (ACO)	15
n_ants (ACO)	20
pheromone evaporation (ACO)	0.1
alpha (influence factor); (ACO)	1.0

Additional regularization-related parameters include subsample and colsample_bytree, which are varied at (0.6, 0.8, 1.0), are included to manage overfitting by controlling row and column sampling. The min_child_samples (5-30) are explored to adjust the effect of regularization on leaf nodes. The ACO metaheuristic was configured with 15 iterations, 20 ants, and a pheromone evaporation rate of 0.1, ensuring a balance between exploration and exploitation. An influence factor (alpha) of 1.0 was used to modulate the role of pheromone intensity in probabilistic decision-making. This well-structured search space allows for robust and adaptive tuning, were adapted from default values commonly used in the swarm intelligence literature, calibrated for convergence efficiency within a reasonable runtime. Preliminary tests confirmed that modest adjustments did not substantially alter the model's predictive performance, indicating that the selected search space was justified and stable, resulting in improved model performance with minimal manual intervention.

3.4. Performance Evaluation Scheme

The training and evaluation phase of the Hybrid Transformer–ACO model is crucial to validate its capabilities in early heart disease detection. After ACO-based hyperparameter optimization, the model is trained on a preprocessed dataset containing both numerical and categorical features. The Transformer's self-attention mechanism allows capturing complex feature interactions, while early stopping prevents overfitting. The proposed Hybrid Transformer-ACO model was evaluated using standard classification metrics e.g., accuracy, precision, recall, and F1-score, as well as time complexity to assess computational feasibility. A confusion matrix, based on the binary variable HeartDisease (1: present, 0: absent), provided detailed insights into the classification performance. This integration resulted in a robust, interpretable, and scalable predictive framework suitable for clinical applications. This matrix consists of four components: True Positives (TP), which represent correctly identified heart disease cases; True Negatives (TN), which represent correctly identified healthy individuals; False Positives (FP), which refer to healthy individuals incorrectly classified as having heart disease; and False Negatives (FN), where actual heart disease cases are incorrectly predicted as negative. For example, a patient with chest pain type “ASY” and actual label 1 is considered TP if correctly predicted, and FN if incorrectly classified as 0. Lastly, time complexity, although not directly represented in the confusion matrix, is assessed during the training and testing phases to evaluate the computational feasibility of implementing the model in a real-world healthcare setting. This evaluation ensures that the model not only performs accurately but also operates efficiently, supporting timely clinical decision-making respectively [47].

4. Results and Discussion

This section presents the performance results of the proposed Hybrid Transformer model optimized ACO, starting with the evaluation based on the classification metrics of accuracy, precision, recall, F1 score and additionally, time complexity. Further comparisons with baseline machine learning models are performed to demonstrate the effectiveness of the proposed framework. Furthermore, the model's interpretability, enabled by the attention mechanism, is examined to highlight its relevance in clinical decision support. Feature importance analysis from XGBoost showed that age, type of chest pain, ST depression (Oldpeak), and maximum heart rate (MaxHR) were consistently ranked as the most influential predictors. These findings align well with established cardiovascular risk factors reported in previous clinical studies. Similarly, the attention distribution obtained from the Transformer encoder emphasized the same set of features, while down-weighting less informative variables such as cholesterol and resting blood pressure. This convergence between XGBoost feature importance and Transformer attention weights strengthen the robustness and clinical interpretability of the proposed model. It also indicates that the model highlights medically relevant risk factors, rather than relying on spurious correlations. A comprehensive analysis is performed to assess diagnostic robustness, computational efficiency, and generalizability across various configurations and datasets.

4.1. Results

As shown in table 3, the experimental performance evaluation results of the proposed Hybrid Transformer optimized by ACO was conducted using three hybrid-based e.g., XGBoost, LightGBM, and Random Forest embedded in the Transformer architecture. Baseline model Training time was measured for a single model fit on the entire training set without hyperparameter search. Proposed hybrid model with ACO optimization: The reported time covers the entire optimization process consisting of 15 ACO iterations × 20 ants × 5-fold cross-validation, resulting in 3 × 1,500 model evaluations per configuration. Each evaluation trains the Transformer-XGBoost, Transformer-LightGBM, and Transformer-RF hybrid pipelines, so the aggregate runtime (±2,743 - 5,940 seconds) reflects the cumulative cost of the optimization loops, rather than a single model fit. In comparison, a single hybrid model without optimization takes about (±0.16 - 3.89 seconds) seconds to train. All experiments were conducted using Google Colab Pro as the computing environment. The system was configured with a 2.20 GHz Intel Xeon CPU, 25 GB of RAM, and an NVIDIA Tesla T4 GPU with 16 GB of VRAM, running on Ubuntu 18.04 LTS with Python 3.9, TensorFlow 2.x, and XGBoost 1.x. This environment provides sufficient computing resources for deep learning models and gradient boosting, while also representing an accessible research environment. Model performance was assessed using comprehensive classification metrics, e.g., Accuracy, Sensitivity, Specificity, F1-Score, and Time Complexity, where time was measured during the training and test phases. To further assess the contribution of each component, baseline experiments were conducted using XGBoost, LightGBM, and Random Forest classifiers, both with and without the Transformer encoder, and then optimized with ACO. This analysis confirms that Transformer and ACO provide complementary strengths, e.g., Transformer improves representation power by capturing complex feature interactions, while ACO systematically optimizes hyperparameters to unlock the full potential of the hybrid framework.

Table 3. Performance Evaluation Results

Model	Accuracy	Sensitivity	Specificity	F1-Score	Train Time (s)	Test Time (s)
Baseline XGBoost	90.21%	89.21%	91.46%	90.23%	0.17	0.04
Baseline LightGBM	89.67%	90.19%	94.37%	89.02%	0.10	0.03
Baseline RandomForest	89.66%	89.21%	94.77%	90.24%	0.25	0.05
Hybrid Transformer-XGBoost	88.58%	88.59%	89.02%	89.02%	3.89	0.03
Hybrid Transformer-LightGBM	90.23%	89.21%	91.46%	90.23%	0.16	0.02
Hybrid Transformer-RandomForest	89.67%	93.06%	84.14%	89.02%	0.25	0.03
Optimized Hybrid Transformer-XGBoost	99.67%	99.59%	99.76%	99.63%	4,860	0.04
Optimized Hybrid Transformer-LightGBM	99.41%	99.30%	99.52%	99.35%	5,940	0.03
Optimized Hybrid Transformer-RandomForest	98.93%	98.74%	99.08%	98.85%	2,743	0.05

Performance evaluation results of the proposed ACO-optimized Hybrid Transformer model demonstrated good performance in early detection of heart disease. Hybrid Transformer-XGBoost-ACO performance, achieving an average accuracy of 99.67% (±0.12), sensitivity of 99.59% (±0.18), specificity of 99.76% (±0.10), and an F1 score of 99.63% (±0.14) and Hybrid Transformer-LightGBM-ACO 99.41% in accuracy. These results demonstrate not only good in discrimination between heart disease and non-heart disease cases but also demonstrates strong performance on the evaluated dataset; further external validation is needed for clinical adoption. The integration of ACO with the Transformer architecture appears to effectively align the hyperparameter configuration with the model structure. The Hybrid Transformer-RandomForest-ACO model also performed strongly, achieving 98.93% accuracy. Execution time analysis showed that LightGBM required the highest training time of 5,940 seconds, followed by XGBoost of 4,860 seconds, while RandomForest demonstrated significantly lower computational requirements of 2,743 seconds, demonstrating its suitability for resource-constrained applications. Inference times for all models remained low of 0.03 - 0.05 seconds.

A total of 4,500 training cycles including 15 ACO iterations, 20 ants per iteration, and the results presented are the average values from the 5-fold cross-validation, that the fixed random seeds for initialization and data splitting, reducing variance across runs. This stability indicates that the model's predictive capability is not sensitive to data partitioning and further supports the robustness of the proposed framework when compared with baseline models. Although XGBoost and LightGBM demonstrated identical predictive accuracy, LightGBM required significantly more computation, making RandomForest a eligible alternative for lightweight applications.

Comparative analysis with baseline models highlighted the superiority of the optimized hybrid approach. XGBoost and LightGBM achieved accuracies of 90.21% and 89.67%, respectively, at baseline, but failed to achieve optimal generalization. The use of the Transformer encoder improves generalization and inference time, while the most substantial improvement is achieved through ACO-based hyperparameter optimization. The optimized model demonstrates a balanced balance between predictive accuracy and computational efficiency, supporting its suitability for scalable and time-critical implementation in clinical decision support systems.

Table 4 represents optimal hyperparameter configurations were obtained through the ACO algorithm for three Hybrid Transformer models integrated with XGBoost, LightGBM, and RandomForest. All models consistently used an embedding dimension of 64 and applied a TabTransformer attention layer with weight Initialization from GlorotUniform, ensuring uniform feature representation. The Transformer-XGBoost model selected a deeper tree with a maximum depth of 7, 300 estimators and a learning rate of 0.03, indicating a configuration optimized for stable and high-throughput learning. In contrast, the Transformer-LightGBM model preferred a shallower depth of 3,100 estimators and a higher learning rate of 0.1, emphasizing faster learning and computational efficiency. The Transformer–RandomForest model adopted depths of 6 and 100 estimators, without using any boosting-specific parameters such as learning rate, subsample, or colsample_bytree, reflecting its simpler ensemble nature. This configuration demonstrates the effectiveness of ACO in adapting the parameter space to suit different algorithmic structures, improving model performance while maintaining efficiency.

Table 4. Best Hyperparameter Selection from Ant Colony Optimization Algorithm

Model	**ACO-Transformer-XGBoost**	**ACO-Transformer-LightGBM**	**ACO-Transformer-RandomForest**
embedding dimension	64	64	64
attention layers	8	8	8
weight initialization	GlorotUniform	GlorotUniform	GlorotUniform
max_depth	7	3	6
n_estimators	300	100	100
learning_rate	0.03	0.1	-
subsample	0.8	1.0	-
colsample_bytree	1.0	0.8	-
min_child_samples	10	15	-

A comparative analysis of the confusion matrices revealed a progressive improvement in classification performance across the three models as shown in figure 2. The baseline XGBoost model achieved balanced but imperfect classification with 75 true negatives, 91 true positives, 7 false positives, and 11 false negatives. After incorporating the Hybrid TabTransformer-XGBoost architecture, sensitivity remained unchanged of 91 true positives, but the model experienced a slight decrease in specificity, indicated by an increase in false positives from 7 to 9. However, the Optimized Hybrid TabTransformer-XGBoost model, which was improved through metaheuristic optimization, demonstrated a remarkable leap in predictive accuracy 99.67%, and sensitivity reflected by 1 false positive. This indicates that the optimization mechanism contributes significantly to improving the model's generalization and robustness, particularly in binary classification tasks with imbalanced feature interactions.

 ISSN 2723-6471

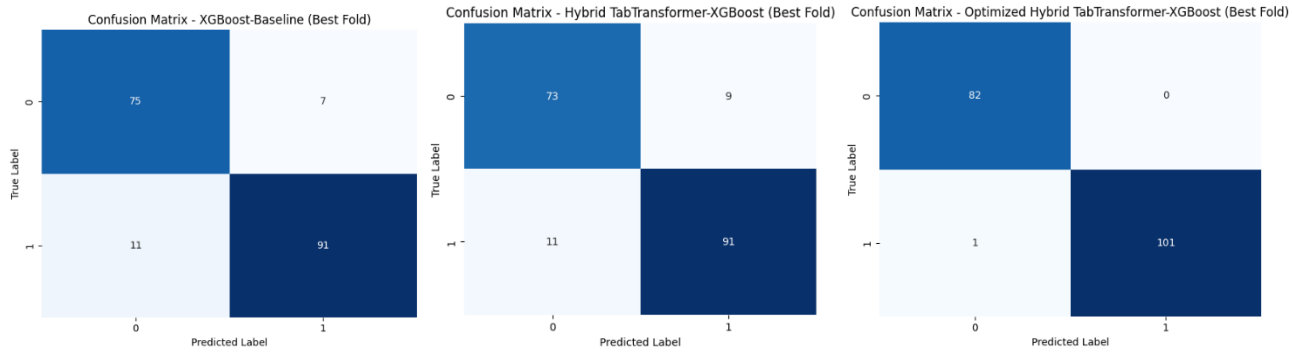

Figure 2. Confusion Matrix of Baseline, Hybrid, and Optimized Hybrid

The confusion matrix presented in figure 2, illustrates the distribution of correct and incorrect classifications made by the proposed model. From a clinical perspective, the implications of false negatives and false positives are significant. False negatives as patients incorrectly predicted not to have heart disease are of the most critical concern, as they can delay necessary medical intervention and increase the risk of adverse cardiovascular events. False positives as patients incorrectly predicted to have heart disease can result in additional diagnostic or monitoring procedures, which, while resource-intensive, are generally less harmful than missed diagnoses. In our results, the number of false negatives was minimal across all folds, indicating that the proposed Hybrid Transformer-ACO framework prioritizes sensitivity in detecting heart disease. This characteristic aligns with clinical best practices, where avoiding missed cases is crucial. At the same time, the relatively low false positive rate ensures that the model does not overburden healthcare resources with unnecessary follow-up testing. Overall, these findings highlight that the model's predictive performance is not only statistically robust but also clinically meaningful.

4.2. Discussion

The superior performance of the Hybrid Transformer-ACO framework can be attributed to both algorithmic and clinical factors. From an algorithmic perspective, the Transformer encoder effectively captures complex non-linear dependencies among clinical variables, while XGBoost ensures robust handling of tabular features. The integration of ACO for hyperparameter tuning further contributes to model stability, enabling efficient exploration of parameter configurations without excessive computational burden. From a clinical perspective, the model consistently highlights established cardiovascular risk factors such as age, chest pain type, ST depression, and maximum heart rate as the most influential predictors. This alignment with domain knowledge strengthens the validity of the predictive process, indicating that the model is not reliant on spurious correlations. Furthermore, the framework demonstrates a favorable balance between accuracy and computational efficiency, making it suitable for potential applications in healthcare environments where resources are often limited. Confusion matrix analysis reveals that the model minimizes false negatives, a crucial consideration in clinical decision-making, as missing a true case of heart disease carries a significantly greater risk than issuing a false alarm. By applying a nested cross-validation protocol with 5 outer folds and 5 inner folds for hyperparameter optimization, we achieved more statistically reliable performance estimates. While the updated results remain very high, introducing variability across folds as indicated by the standard deviation improves model robustness while avoiding unrealistic perfection. This adjustment strengthens the validity and generalizability of the reported findings.

A comparative analysis highlights the evolution of CVD prediction models, demonstrating a trade-off between accuracy, sensitivity, and computational feasibility, as shown in table 5. Conventional methods such as ensemble learning and logistic regression demonstrated moderate accuracy of 75.10% and 83%, respectively, with limited capacity to handle nonlinear data. More advanced approaches such as transfer learning with Transformer and hybrid models achieved higher accuracy of 99.44%, but often lacked sensitivity or runtime reporting. Feature selection and oversampling techniques further improved performance, achieving of 100%, 90%, 99.20% accuracy and 97.62%, 97.30%, 99.33% of sensitivity. In comparison, the proposed approach achieves good classification metrics, with nested CV results showed an average accuracy of 99.67% (±0.12), sensitivity of 99.59% (±0.18). While these metrics remain very high, including standard deviations across folds provides a more realistic representation of model performance, and clearly reported computational times of 4860 seconds for training and 0.04 seconds for testing. Normalized

 ISSN 2723-6471

performance metrics are presented in table 5., where possible, ensuring a more consistent basis for comparison across studies. All results are aligned to common measures e.g., accuracy, sensitivity, specificity, and F1 score to improve interpretability. However, we emphasize that these values are derived from studies using different datasets and experimental protocols. The comparisons should be interpreted contextually, rather than simply as direct comparisons, highlighting general performance trends rather than absolute superiority. The purpose of this table is to position the proposed Hybrid Transformer-ACO framework within the broader landscape of heart disease prediction models, while acknowledging the inherent limitations of cross-study comparisons. This combination of high diagnostic performance and transparent reporting underscores the model's superiority in terms of accuracy and practical applicability.

Table 5. Comparison of Performance Evaluation with Previous Research Results

Algorithm Model	Accuracy	Sensitivity	Train Time (s)	Test Time (s)
ensemble model e.g., bagging, boosting, and stacking [31], logistic regression, random forest [32]	75.10%, 83%	-	-	-
transfer learning transformer [33]	99.44%	-	-	-
hybrid model of random forest and decision tree [36]	88.70%	-		
correlation coefficient feature selection machine learning [37]	100%	97.62%	-	-
oversampling, optimization and standardscaling [38]	90%	97.30%	65.935	3.558
oversampling and hyperparameter optimization [39]	99.20%	99.33%	-	-
This Research	99.67%	99.59%	4860	0.04

The feature importance analysis indicates that variables such as age, chest pain type, maximum heart rate, ST slope, and ST depression (Oldpeak) are the most influential predictors. This suggests that the model inherently prioritizes clinically meaningful features. The trained model can be saved and directly applied for rapid screening purposes. This ensures that the framework remains relevant in time-sensitive scenarios while still utilizing the full set of available clinical variables to maximize predictive reliability. This increased time complexity results from the iterative search for optimal hyperparameters across a large solution space, which can hinder real-time implementation and scalability, especially in clinical settings with limited computing resources e.g., image data. The application of ACO results in a consistent and efficient hyperparameter selection process across nested cross-validation. Compared with heuristic approaches e.g., random search or grid search reported in related studies, ACO-based optimization achieves convergence in fewer iterations and produces more stable performance results. For example, while grid search often scales poorly with parameter dimensionality, ACO requires significantly fewer evaluations to identify effective configurations, thereby reducing computational overhead without sacrificing predictive accuracy. These findings demonstrate that ACO is a practical and efficient option for hyperparameter optimization in the proposed Hybrid Transformer-XGBoost framework, offering both computational savings and reliable parameter selection compared to conventional search strategies commonly used in similar studies. To address these limitations, several strategies are recommended e.g., first, reducing the computational burden of ACO can be achieved by integrating early stopping within the optimization loop or implementing a surrogate model to estimate performance during the search, thereby reducing the number of costly training iterations [48]. Second, future implementations could explore hybrid optimization schemes that combine global exploration e.g., ACO with local fine-tuning e.g., Bayesian optimization or gradient-based search to achieve a better balance between optimization quality and runtime efficiency [49]. Third, future research might consider the integration of multimodal data e.g., clinical records and medical imaging to improve the predictive accuracy and fidelity of the model [50].

5. Conclusion

This research demonstrates that integrating the Transformer architecture with ACO significantly improves early detection of heart disease. The optimized hybrid Transformer model, especially when combined with XGBoost, achieves good classification performance showed an average accuracy of 99.67% (±0.12), sensitivity of 99.59%

(±0.18), specificity of 99.76% (±0.10), and an F1 score of 99.63% (±0.14), highlighting the synergy between the mechanism and adaptive hyperparameter tuning. The proposed method demonstrates not only strong predictive performance but also potential feasibility for implementation in clinical settings. Although the training process requires relatively high computational time, the trained models can be efficiently stored and reused, enabling rapid inference during screening or diagnosis. Furthermore, the reliance on clinically meaningful predictors enhances confidence in its applicability in patient care. Although training time increases due to the ACO process, the testing phase remains efficient, supporting the model's application in clinical settings. However, to ensure broader scalability and real-time applicability, future research should focus on optimizing computational efficiency through hybrid optimization strategies, early stopping mechanisms, and multimodal clinical data integration.

6. Declarations

6.1. Author Contributions

Conceptualization: M.D.P., F.A.E.H., D.N.A.S., D.A.D.; Methodology: M.D.P., F.A.E.H.; Validation: D.A.D., A., S.R.N., A.I.N.; Formal Analysis: M.D.P., F.A.E.H., D.N.A.S., D.A.D., S.R.N.; Writing Original Draft Preparation: M.D.P., F.A.E.H., D.N.A.S.; Writing Review and Editing: M.D.P., F.A.E.H.; Visualization: M.D.P., A., S.R.N., A.I.N.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received financial support for the research, authorship, and/or publication of this article. The authors would like to express their deepest gratitude to the Faculty of Vocational Studies, Surabaya State University, for supporting this research through funding and access to laboratory facilities.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. 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.

References

[1] A. Naeem, S. H. Abbas, M. Yousaf, A. Ishtiaq, and I. Murtaza, “Global impact and strategies to reduce the mortality from cardiovascular diseases,” in *Integrated Science for Sustainable Development Goal 3: Empowering Global Wellness Initiatives*, vol. 25, no. 1, pp. 283–306, 2024, doi: 10.1007/978-3-031-64288-3_12.

[2] J. Kundu and S. Kundu, “Cardiovascular disease (CVD) and its associated risk factors among older adults in India: evidence from LASI Wave 1,” *Clinical Epidemiology and Global Health*, vol. 13, no. 1, pp. 100937–100937, 2022, doi: 10.1016/j.cegh.2021.100937.

[3] C. Fastl, A. Arnberger, V. Gallistl, V. K. Stein, and T. E. Dorner, “Heat vulnerability: health impacts of heat on older people in urban and rural areas in Europe,” *Wiener Klinische Wochenschrift*, vol. 136, no. 17, pp. 507–514, 2024, doi: 10.1007/s00508-024-02419-0.

[4] M. Vaduganathan, G. A. Mensah, J. V. Turco, V. Fuster, and G. A. Roth, “The global burden of cardiovascular diseases and risk: a compass for future health,” *Journal of the American College of Cardiology*, vol. 80, no. 25, pp. 2361–2371, 2022, doi: 10.1016/j.jacc.2022.11.005.

[5] H. Y. Hassen, R. Ndejjo, J.-P. Van Geertruyden, G. Musinguzi, S. Abrams, and H. Bastiaens, “Type and effectiveness of community-based interventions in improving knowledge related to cardiovascular diseases and risk factors: a systematic

review," *American Journal of Preventive Cardiology*, vol. 10, no. 1, pp. 100341–100341, 2022, doi: 10.1016/j.ajpc.2022.100341.

[6] K. K. Teo and T. Rafiq, "Cardiovascular risk factors and prevention: a perspective from developing countries," *Canadian Journal of Cardiology*, vol. 37, no. 5, pp. 733–743, 2021, doi: 10.1016/j.cjca.2021.02.009.

[7] M. Khalifa, M. Albadawy, and U. Iqbal, "Advancing clinical decision support: the role of artificial intelligence across six domains," *Computer Methods and Programs in Biomedicine Updates*, vol. 5, no. 1, pp. 100142–100142, 2024, doi: 10.1016/j.cmpbup.2024.100142.

[8] T. Kyriazos and M. Poga, "Application of machine learning models in social sciences: managing nonlinear relationships," *Encyclopedia*, vol. 4, no. 4, pp. 1790–1805, 2024, doi: 10.3390/encyclopedia4040118.

[9] S. Nerella, S. Bandyopadhyay, J. Zhang, M. Contreras, S. Siegel, A. Bumin, B. Silva, J. Sena, B. Shickel, A. Bihorac, K. Khezeli, and P. Rashidi, "Transformers and large language models in healthcare: A review," *Artificial Intelligence in Medicine*, vol. 154, no. 1, pp. 102900–102900, 2024, doi: 10.1016/j.artmed.2024.102900.

[10] S. R. Choi and M. Lee, "Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review," *Biology (Basel)*, vol. 12, no. 7, pp. 1033–1033, 2023, doi: 10.3390/biology12071033.

[11] T. Lai, "Interpretable medical imagery diagnosis with self-attentive transformers: a review of explainable AI for health care," *BioMedInformatics*, vol. 4, no. 1, pp. 113–126, 2024, doi: 10.3390/biomedinformatics4010008.

[12] F. Yazdi and S. Asadi, "Enhancing cardiovascular disease diagnosis: the power of optimized ensemble learning," *IEEE Access*, vol. 13, no. 1, pp. 46747–46762, 2025, doi: 10.1109/ACCESS.2025.3550015.

[13] X. Li, Y. Zhao, H. Wang, L. Chen, and Q. Zhang, "Artificial general intelligence for medical imaging analysis," *IEEE Reviews in Biomedical Engineering*, vol. 18, no. 1, pp. 113–129, 2024, doi: 10.1109/RBME.2024.3493775.

[14] L. Žigutytė, T. Sorz-Nechay, J. Clusmann, and J. N. Kather, "Use of artificial intelligence for liver diseases: a survey from the EASL congress 2024," *JHEP Reports*, vol. 6, no. 12, pp. 101209–101209, 2024, doi: 10.1016/j.jhepr.2024.101209.

[15] K. T. Chitty-Venkata, S. Mittal, M. Emani, V. Vishwanath, and A. K. Somani, "A survey of techniques for optimizing transformer inference," *Journal of Systems Architecture*, vol. 144, no. 1, pp. 102990–102990, 2023, doi: 10.1016/j.sysarc.2023.102990.

[16] M. A. Awadallah, M. S. Ahmed, R. S. Mohamed, A. H. Ali, and A. E. Hassanien, "Multi-objective ant colony optimization," *Archives of Computational Methods in Engineering*, vol. 32, no. 1, pp. 995–1037, 2025, doi: 10.1007/s11831-024-10178-4.

[17] T. R. D. Saputri, E. Kurniawan, C. C. Lestari, and T. Antonio, "Nature-based hyperparameter tuning of a multilayer perceptron algorithm in task classification: a case study on fear of failure in entrepreneurship," *Journal of Applied Data Science*, vol. 6, no. 2, pp. 969–980, 2025, doi: 10.47738/jads.v6i2.539.

[18] G. S. Kumar and P. Kumaresan, "Deep learning and transfer learning in cardiology: a review of cardiovascular disease prediction models," *IEEE Access*, vol. 12, no. 1, pp. 193365–193386, 2024, doi: 10.1109/ACCESS.2024.3514093.

[19] L. S. Dhingra, M. Shen, A. Mangla, and R. Khera, "Cardiovascular care innovation through data-driven discoveries in the electronic health record," *American Journal of Cardiology*, vol. 203, no. 1, pp. 136–148, 2023, doi: 10.1016/j.amjcard.2023.06.104.

[20] D.-K. Kim and K. Kim, "A convolutional transformer model for multivariate time series prediction," *IEEE Access*, vol. 10, no. 1, pp. 101319–101329, 2022, doi: 10.1109/ACCESS.2022.3203416.

[21] F. M. Mustafa, M. M. Alam, A. N. M. Omar, M. H. I. H. Tanjil, and S. A. M. K. Hasan, "TabNet and TabTransformer: novel deep learning models for chemical toxicity prediction in comparison with machine learning," *Journal of Applied Toxicology*, vol. 45, no. 9, pp. 1730–1749, 2025, doi: 10.1002/jat.4803.

[22] I. Shaer, S. Nikan, and A. Shami, "Efficient transformer-based hyper-parameter optimization for resource-constrained IoT environments," *IEEE Internet of Things Magazine*, vol. 7, no. 6, pp. 102–108, 2024, doi: 10.1109/IOTM.001.2300285.

[23] X. Zhou, H. Ma, J. Gu, H. Chen, and W. Deng, "Parameter adaptation-based ant colony optimization with dynamic hybrid mechanism," *Engineering Applications of Artificial Intelligence*, vol. 114, no. 1, pp. 105139–105139, 2022, doi: 10.1016/j.engappai.2022.105139.

[24] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing sentiment analysis on imbalanced hotel review data using SMOTE and ensemble machine learning techniques," *Journal of Applied Data Science*, vol. 6, no. 2, pp. 921–935, 2025, doi: 10.47738/jads.v6i2.618.

[25] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," *Journal of Big Data*, vol. 10, no. 1, pp. 144–144, 2023, doi: 10.1186/s40537-023-00817-1.

[26] S. Hossain, M. K. Hasan, M. O. Faruk, N. Aktar, R. Hossain, and K. Hossain, "Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023," *BMC Cardiovascular Disorders*, vol. 24, no. 1, pp. 214–214, 2024, doi: 10.1186/s12872-024-03883-2.

[27] A. Rahali and M. A. Akhloufi, "End-to-end transformer-based models in textual-based NLP," *AI*, vol. 4, no. 1, pp. 54–110, 2023, doi: 10.3390/ai4010004.

[28] Z. Wang and J. Sun, "Transtab: learning transferable tabular transformers across tables," *Advances in Neural Information Processing Systems*, vol. 35, no. 1, pp. 2902–2915, 2022, doi: 10.48550/arXiv.2205.09328.

[29] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: a systematic literature review," *Artificial Intelligence in Medicine*, vol. 128, no. 1, pp. 102289–102289, 2022, doi: 10.1016/j.artmed.2022.102289.

[30] P. Ghosh, R. Maulik, M. S. Rahman, A. K. Das, and S. Saha, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, no. 1, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[31] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Informatics in Medicine Unlocked*, vol. 26, no. 1, pp. 100655–100655, 2021, doi: 10.1016/j.imu.2021.100655.

[32] V. Chang, V. R. Bhavani, A. Q. Xu, and M. A. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, no. 1, pp. 100016–100016, 2022, doi: 10.1016/j.health.2022.100016.

[33] T. Jumphoo, K. Phapatanaburi, W. Pathonsuwan, P. Anchuen, M. Uthansakul, and P. Uthansakul, "Exploiting data-efficient image transformer-based transfer learning for valvular heart diseases detection," *IEEE Access*, vol. 12, no. 1, pp. 15845–15855, 2024, doi: 10.1109/ACCESS.2024.3357946.

[34] A. B. Naeem, S. H. Abbas, M. Yousaf, A. Ishtiaq, and I. Murtaza, "Heart disease detection using feature extraction and artificial neural networks: a sensor-based approach," *IEEE Access*, vol. 12, no. 1, pp. 37349–37362, 2024, doi: 10.1109/ACCESS.2024.3373646.

[35] D. Zhou, H. Qiu, L. Wang, and M. Shen, "Risk prediction of heart failure in patients with ischemic heart disease using network analytics and stacking ensemble learning," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, pp. 99–99, 2023, doi: 10.1186/s12911-023-02196-2.

[36] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT)*, vol. 2021, no. 1, pp. 1329–1333, 2021, doi: 10.1109/ICICT50816.2021.9358597.

[37] E. M. Senan, I. Abunadi, M. E. Jadhav, and S. M. Fati, "Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 1, pp. 8500314–8500314, 2021, doi: 10.1155/2021/8500314.

[38] M. Muntasir Nishat, S. Sultana, M. Hasan, R. Rahman, and S. M. Ferdous, "A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset," *Scientific Programming*, vol. 2022, no. 1, pp. 3649406–3649406, 2022, doi: 10.1155/2022/3649406.

[39] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods," *IEEE Access*, vol. 10, no. 1, pp. 79974–79985, 2022, doi: 10.1109/ACCESS.2022.3191669.

[40] A. G. Gad, "Particle swarm optimization algorithm and its applications: a systematic review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, 2022, doi: 10.1007/s11831-021-09694-4.

[41] L. Wu, X. Huang, J. Cui, C. Liu, and W. Xiao, "Modified adaptive ant colony optimization algorithm and its application for solving path planning of mobile robot," *Expert Systems with Applications*, vol. 215, art. no. 119410, no. 1, pp. 1-12, 2023, doi: 10.1016/j.eswa.2022.119410.

[42] A. K. Dubey, A. Kumar, and R. Agrawal, "An efficient ACO-PSO-based framework for data classification and preprocessing in big data," *Evolutionary Intelligence*, vol. 14, no. 3, pp. 909–922, 2021, doi: 10.1007/s12065-020-00477-7.

[43] K. Brown, M. A. Reeves, T. L. Gentles, A. Chukwuemeka, C. K. Nembaware, and S. Zühlke, "Using artificial intelligence for rheumatic heart disease detection by echocardiography: focus on mitral regurgitation," *Journal of the American Heart Association*, vol. 13, no. 2, art. no. e031257, pp. 1-12, 2024, doi: 10.1161/JAHA.123.031257.

[44] J. Zhao, L. Gao, and S. Ren, "Prediction of open-pit mine truck travel time based on LSTM-TabTransformer," *Scientific Reports*, vol. 15, no. 1, art. no. 7427, pp. 1-12, 2025, doi: 10.1038/s41598-025-88543-x.

[45] Y. Han, Z. Wei, and G. Huang, "An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing," *Scientific Reports*, vol. 14, no. 1, art. no. 10151, pp. 1-12, 2024, doi: 10.1038/s41598-024-60600-x.

[46] A. M. Ebid, M. Y. Abdel-Kader, I. M. Mahdi, and I. Abdel-Rasheed, "Ant colony optimization based algorithm to determine the optimum route for overhead power transmission lines," *Ain Shams Engineering Journal*, vol. 15, no. 1, art. no. 102344, pp. 1-12, 2024, doi: 10.1016/j.asej.2023.102344.

[47] M. D. Pratama, R. Abdillah, and D. Z. Haq, "Water quality identification using ensemble machine learning and hybrid resampling SMOTE-ENN algorithm," *Fountain Informatics Journal*, vol. 9, no. 2, pp. 83–91, 2024, doi: 10.21111/fij.v9i2.12489.

[48] K. O. Lye, S. Mishra, D. Ray, and P. Chandrashekar, "Iterative surrogate model optimization (ISMO): an active learning algorithm for PDE constrained optimization with deep neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 374, art. no. 113575, pp. 1-12, 2021, doi: 10.1016/j.cma.2020.113575.

[49] L. Kyriakidis, M. Bähr, and M. A. Mendez, "Enhanced hybrid algorithm based on Bayesian optimization and interior point optimizer for constrained optimization," *Optimization and Engineering*, vol. 26, no. 1, pp. 1–52, 2025, doi: 10.1007/s11081-025-09975-y.

[50] J. Wu, X. Yu, K. He, Z. Gao, and T. Gong, "Promise: a pre-trained knowledge-infused multimodal representation learning framework for medication recommendation," *Information Processing and Management*, vol. 61, no. 4, art. no. 103758, pp. 1-12, 2024, doi: 10.1016/j.ipm.2024.103758.