Sumon *et al.*
Health Information Science and Systems (2025) 13:44
https://doi.org/10.1007/s13755-025-00361-7

Health Information Science and Systems

RESEARCH

CardioTabNet: a novel hybrid transformer model for heart disease prediction using tabular medical data

Md. Shaheenur Islam Sumon[1], Md. Sakib Bin Islam[1], Md. Sohanur Rahman[2], Md. Sakib Abrar Hossain[1], Amith Khandakar[1], Anwarul Hasan[3], M. Murugappan[4,5*] and Muhammad E. H. Chowdhury[1*]

Abstract

The early detection and accurate prediction of cardiovascular diseases (CVDs) are critical to reduce global severe morbidity and mortality. Machine learning (ML) methods, operated by Transformers have proved its efficiency in interpreting complex data interactions. One prime example would be its notable success in Natural Language Processing (NLP), with its multi-headed self-attention mechanism to disentangle the complex interactions within high-dimensional spaces. However, the relationships between various features within biological systems remain ambiguous in these spaces, making it difficult to apply transformers in clinical datasets. We introduce CardioTabNet, a transformer-driven framework designed precisely for clinical cardiovascular data. It leverages the strength of the tab transformer architecture to effectively extract meaningful insights from clinical data. As a result, downstream classical models' performance significantly showed outstanding results. We utilized an open-source cardiovascular dataset with 1190 instances and 11 features. These features are categorized into numerical (age, resting blood pressure, cholesterol, maximum heart rate, old peak, weight, and fasting blood sugar) and categorical (resting Electrocardiograms, exercise angina, and ST slope) variables. Tab transformer was used to extract significant features and rank them using a Random Forest (RF) feature ranking algorithm which highlighted the important clinical predictors. We used ten classical machine-learning models trained on these transformer extracted-features. An optimized ExtraTree classifier achieved an average accuracy of 94.1% and area under curve (AUC) of 95%. Furthermore, we performed nomogram analysis to draw out cardiovascular risk assessment to demonstrate clinical interpretability. Benchmarking against state-of-the-art methodologies affirmed the superior predictive capability of our CardioTabNet framework, demonstrating its potential as a robust tool for clinical decision support in cardiovascular disease prediction and early detection. In addition, SHAP (SHapley Additive exPlanations) analysis was carried out to provide insights into feature contributions and enhance model interpretability.

Keywords: Tab transformer, Machine learning, Deep learning, Heart disease, Classification, Nomogram

Introduction

Heart diseases are among the leading causes of death and disability worldwide, affecting millions of people every year. According to the World Health Organization (WHO), 16% of all fatalities in 2019 [1] were attributed to cardiovascular diseases. Early detection of cardiac diseases can help prevent or delay severe complications, such as myocardial infarction and arrhythmias. However, the early diagnosis and prediction of heart diseases require numerous variables like age, gender, blood pressure, cholesterol levels, diabetes, smoking habits, family

*Correspondence: m.murugappan@kcst.edu.kw; mchowdhury@qu.edu.qa
[2] Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka 1000, Bangladesh
[4] Department of Electronics and Communication Engineering, Intelligent Signal Processing (ISP) Research Lab, Kuwait College of Science and Technology, Block 4, 13133 Doha, Kuwait
Full list of author information is available at the end of the article

history, and lifestyle, which is a challenging task [2–4]. Moreover, the presence of diverse risk factors and symptoms across various forms of heart disease complicates the process of determining the most suitable diagnostic test or treatment for a given patient. A comprehensive evaluation of a patient's cardiac status is crucial for the diagnosis and management of cardiovascular disorders. Through the utilization of computed tomography (CT), magnetic resonance imaging (MRI), and electrocardiograms (ECGs), a solid assessment can be achieved [5]. However, due to resource limitations, testing with these diagnostic modalities may prove challenging to administer and sometimes less accurate Therefore, alternative and innovative methods for early prediction of cardiac diseases could save millions of lives, especially in developing nations [6, 7].

Early detection and prompt intervention in cardiovascular diseases play crucial role for reducing premature death [8, 9]. Predictive models significantly contribute to identifying high-risk patients. As a result, clinicians would be able to facilitate timely clinical interventions. Conditions such as coronary heart disease, heart failure, congenital heart disease, cyanotic heart disease, and cardiomyopathy are some of the many forms of cardiovascular diseases that affect health and well-being [10]. Coronary heart disease limits the supply of vital nutrients and oxygen to the heart by obstructing coronary arteries. This often leads to potentially fatal outcomes. Heart failure, a condition in which the heart is unable to pump blood effectively, represents a more advanced stage of cardiovascular diseases, frequently triggered by factors such as coronary artery disease. Congenital heart disease is characterized by structural abnormalities in the heart, such as septal defects. It develops during fetal growth and is present from birth [11]. Cyanotic heart disease arises from restrictive defects that limit oxygen delivery or impede the blood flow [12]. Cardiomyopathy weakens the heart's ability to circulate blood, potentially leading to heart failure.

Machine Learning (ML) methodologies offer tools to extract valuable from complex cardiovascular datasets, paving the way for accurate disease predictions. ML, a subfield of artificial intelligence, enables computers to learn trends from data and perform tasks that are dependent on human intelligence [13, 14]. It has been widely explored in various domains of biomedical research and healthcare, including genomics, proteomics, drug discovery, diagnosis, prognosis, and personalized medicine [8]. ML has shown significant promise, particularly in predicting cardiovascular diseases using diverse data sources, such as genomic information, clinical records, demographic data, lifestyle factors, biomarkers, electrocardiogram signals, and imaging data [15–17]. It can aid in identifying risk factors, patients' classification, estimation of disease development likelihood, and suggest appropriate actions or treatments. Despite all these advantages, ML faces several challenges and limitations in the context of prediction of cardiovascular diseases. One significant challenge would be handling the diversity and complexity of the data sources and formats. Datasets often come across varying properties, sizes, distributions, quality, and reliability coupled with containing errors, missing values, outliers, or noise. Additionally, some datasets may present multicollinearity or high dimensionality. These issues can affect the performance and generalizability of ML models [18]. Selecting the optimal algorithm for predicting tasks requires careful considerations of multiple factors, including the nature and size of the data, the computational cost and efficiency of the algorithm, and the complexity and interpretability of the model [19].

Compared to previous research that primarily focused on conventional heart disease datasets with satisfactory outcomes but did not explore more robust datasets ore embedding techniques for categorical variables, our study evaluates the efficacy of the tab transformer architecture. Drawing on recent studies, we hypothesize that implementing tab transformers will improve predictive performance. This approach converts categorical variables into powerful contextual embeddings which address a crucial aspect often overlooked in earlier methodologies. As a result of leveraging this novel technique, CardioTabNet significantly improves prediction accuracy, distinguishing itself from previous models and providing a more comprehensive representation of data for the prediction of heart disease.

Innovative predictive models that provide reliable, interpretable, and accurate cardiovascular risk assessments are urgently needed to address these challenges. The recent advances in machine learning, particularly in handling tabular medical data, offer a promising pathway to improve heart disease prediction. The objective of our study is to bridge this gap by proposing a novel model that utilizes state-of-the-art machine learning techniques to enhance predictive accuracy and provide insights into cardiovascular diseases risk factors.

A pioneering approach to predicting cardiovascular diseases based on the IEEE Data Port Heart Disease dataset [20–24]. The motivation for this research is the growing global burden of cardiovascular diseases, which are the leading cause of death in the world. It is crucial to make accurate and timely predictions of cardiovascular diseases to reduce mortality rates and initiate early interventions. Despite the availability of numerous predictive models, there is a clear need for more robust, interpretable, and efficient systems that can use

modern machine-learning techniques and large datasets to improve prediction accuracy and clinical applicability. In summary, we have made the following contributions:

- In this study, we present CardioTabNet, a state-of-the-art model that was specifically designed to predict cardiovascular diseases. A rigorous and comprehensive statistical analysis was conducted to provide a solid foundation for the findings of this study. This model incorporates the rich set of data available from the IEEE Data Port to enhance the accuracy and reliability of cardiovascular risk assessment. Our proposed model, CardioTabNet, is validated and credible through this analysis.
- CardioTabNet utilizes tab transformer to extract essential features from a dataset. A feature ranking strategy was used to identify the top 10 features, which were then used to train 10 classical machine-learning models. In this way, a more focused and efficient predictive model can be developed.
- There has been a notable improvement in experimental results compared to previous studies in this area. CardioTabNet's effectiveness combined with the top 10 features has resulted in superior predictive performance for cardiovascular diseases.
- We incorporated logistic regression and nomogram analysis to further enhance binary classification precision. The sophisticated analysis enhances accuracy and introduces a scoring system that facilitates nuanced differentiation between positive and negative categories.
- We employed SHapley Additive exPlanations (SHAP) analysis to assess the impact of each individual feature on the model's predictions.

By following this structured format, our research establishes a robust foundation for advancing the field of prediction of cardiovascular diseases, displaying the efficacy of CardioTabNet and contributing valuable insights for future studies. The rest of the paper is organized as follows: "Related works" Sect. reviews related studies, "Materials and methods" Sect. explains the proposed methodology, "Results" Sect. presents and discusses the results, and "Conclusion" Sect. concludes the study with a summary and suggestions for future research.

Related works

This section provides an overview of several recent studies that utilize machine learning techniques and the IEEE Data Port dataset to predict cardiac disease. Five distinct datasets from various sources are merged to form the IEEE Data Port dataset; these include the Hungarian, Cleveland, Long Beach, Virginia, Switzerland, and Statlog datasets. The dataset includes 1190 instances and 11 features, which include both numeric and nominal attributes. The target variable denotes the presence or absence of cardiac disease in patients as a binary class.

Various machine learning models have been applied to the prediction of heart disease in recent research, with notable success. In Tiwari et al.'s study [21], classifiers such as ExtraTrees (ET), Random Forest (RF), and XGBoost were applied to the IEEE Data Port and achieved high accuracy (92.34%) and F1-scores of 92.74%, showing robust recall (93.49%) and sensitivity (91.07%). Similar to that study, another study [25] evaluated multiple ML models using the IEEE Data Port, including Logistic Regression (LR), K Nearest Neighbor (KNN), Decision Tree (DT), and RF, resulting in an accuracy score of 91.60%. A recall of 94.30%, a sensitivity of 88.39%, and an F1-score of 92.43% were reported in the study. Comprehensive investigation highlighted the effectiveness of ensemble methods as a means of predicting heart disease, reinforcing the notion that combining different models can lead to more accurate outcomes. Nagarajan et al. [26] contributed further to this field by employing LR, KNN, and RF models, which resulted in an accuracy of 90.67%, sensitivity of 92.68%, and specificity of 88.39%. As a result of their work, they were able to demonstrate that combinations of models can achieve balanced and comprehensive performance in classification tasks. A different approach integrated SHAP [27] (SHapley Additive Explanations) for feature importance analysis and XGBoost for classification. By utilizing the IEEE Data Port, this study sought to increase interpretability while achieving a respectable accuracy of 90.08%, with a sensitivity of 91.46% and specificity of 88.39%.

Several datasets, including the IEEE Data Port and the Mendeley Data Center cardiovascular data, were used by Doppala et al. [28] to illustrate the versatility of ML. Across different datasets, they found that Naive Bayes (NB), RF, and XGBoost consistently performed well, achieving impressive accuracy levels of 96.75% for the Mendeley dataset and 93.39% for the IEEE Data Port. There were also traditional models [29] that showed competitive performance. As an example, using Decision Trees (DT), RF, and Support Vector Machine (SVM) with the IEEE Data Port dataset resulted in an accuracy of 85.12%. The results show that classical models can still provide reliable results even though newer techniques offer enhancements. According to Paul et al. [24], it is crucial to account for dataset variability. They used artificial neural networks with scaled conjugate gradient backpropagation to analyze the Cleveland Hungarian Statlog dataset and the Cleveland processed heart dataset. Consequently, the accuracy of the model varied significantly, ranging from 63.38 to 88.48%, emphasizing the

importance of the choice of dataset. Advanced techniques were further explored by Baccouche et al. [30], who combined Bidirectional Long Short Term Memory (BiLSTM) or Bidirectional Gated Recurrent Unit (BiGRU) models with Convolutional Neural Network (CNN) to form an ensemble classifier. Based on their study of the MIT-BIH Arrhythmia Database, they achieved excellent accuracy and F1 scores ranging from 91 to 96%, demonstrating the potential of deep learning in medical applications. Similarly, Dubey et al. [31] used the IEEE CHD dataset, applying Naive Bayes, Bayes Nets, and Multilayer Perceptrons, and achieved 93.67% accuracy. In light of this, Bayesian methods are reliable methods for predicting heart disease provided they are applied to medical data. Mohan et al. [32] achieved 88.7% accuracy using Decision Trees and Neural Networks. As a result of these findings, models must be carefully selected following the specific dataset and research objectives.

Moreover, studies using Kaggle and Cleveland Clinic Foundation datasets indicated that K Nearest Neighbor often outperformed other models. In contrast, weighted Naive Bayes and XGBoost with Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and SMOTE-ENN (Edited Nearest Neighbor) [33] excelled in precision, demonstrating the strengths of probabilistic methods in classification tasks. Using a MultiLayer Perceptron (MLP), an accuracy of 87.28% and a high Area Under Curve (AUC) score of 0.95 were demonstrated, highlighting the benefits of neural networks in the application domain. An overview of the prior research that utilized this dataset to forecast heart disease is presented in Table 1. According to Table 1, we developed our research questions after reviewing the most recent and prominent research in this field.

Materials and methods

CardioTabNet, a framework that uses transformer technology to extract superior feature spaces from clinical cardiovascular data, is intended to fundamentally alter the detection of cardiovascular disease through the use of transformer technology. The overall methodology is visually depicted in Fig. 1. In the first phase, the dataset is subjected to in-depth statistical analysis to determine relevant characteristics. Subsequently, a meticulous data normalization process is implemented to ensure a consistent and standardized structure. Feature extraction is executed through the tab transformer model, followed by feature ranking using the RandomForest algorithm. Subsequently, a rigorous evaluation of ten ML models is conducted, leading to the identification of the top-performing model. Heart disease outcomes are effectively predicted using the CardioTabNet model, which combines tab transformer architecture with advanced data processing and machine learning techniques. The model's ability to identify key features and optimize hyperparameters using Optuna (a source open source hyperparameter optimization) highlights its potential for clinical applications in heart disease risk assessment and personalized treatment planning.

Data description

The dataset utilized in this study was sourced from the IEEE Data Port. IEEE Data Port is an online platform that provides access to various datasets related to engineering and technology fields. The dataset that we use in this paper is a comprehensive dataset for heart disease prediction that combines five well-known datasets [41–43] from different sources: Hungarian dataset, Cleveland dataset, Long Beach VA dataset, Switzerland dataset, and Statlog dataset. The dataset consists of 1190 instances with 11 features related to heart disease diagnosis. Table 2 describes the IEEE Data Port Heart Disease Prediction Dataset. This dataset combines five well-known datasets from diverse sources to support heart disease prediction research. Age, gender, blood pressure, cholesterol, and exercise-induced symptoms are included. The dataset includes valuable characteristics including chest pain kinds, ECG readings, and exercise-induced angina for proper diagnosis. The goal variable—heart disease status—improves predictive modeling using the dataset. The large and multidimensional dataset from these multiple data sources allows researchers to study detailed patterns and construct superior ML models for heart disease prediction.

Data preprocessing

The initial step in the dataset preprocessing involves the application of a z-score-based filtering technique to enhance data quality by mitigating the impact of outliers [21]. Following the outlier removal process, the dataset is prepared for the training of the tab transformer model. In this subsequent phase, a systematic segregation of features is performed, distinguishing between categorical features, numerical features, and a weight column. This segmentation is outlined in Table 3. We used SMOTE to handle class imbalance (see "SMOTE augmentation method" Sect. for details) [21, 44].

Statistical analysis

The dataset's characteristics were statistically interpreted using Stata/MP version 15.00. The examination encompassed statistical measures such as the mean, median, standard deviation (STD), 25th and 75th quartile values, as well as the mean and maximum values of a certain characteristic. Additionally, p-values were adopted to determine the relationship with the output [44].Three

Table 1 A compilation of previous research endeavors focused on the prediction of cardiac disease

Author name	Dataset	Methods	Results and observation
Tiwari et al. [21]	IEEE Data Port	ExtraTrees Classifier, Random Forest, XGBoost, Gradient Boosting Machine (GBM) and Logistic Regression	Accuracy: 92.34%, Recall: 93.49%, Sensitivity: 91.07%, F1-Score: 92.74%
Rajdhan et al. [25]	IEEE Data Port	Logistic Regression, KNN, Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, XGBoost, LightGBM, CatBoost	Accuracy: 91.60%, Recall: 94.30%, Sensitivity: 88.39%, F1-Score: 92.43%
Nagarajan et al. [26]	IEEE Data Port	Logistic Regression, KNN, Decision Tree, Random Forest, SVM, Naive Bayes, XGBoost	Accuracy: 90.67%, Sensitivity: 92.68%, Specificity: 88.39%, F1-score: 91.18%
Tjoa et al. [27]	IEEE Data Port	SHAP (SHapley Additive exPlanations) framework for feature importance analysis and XGBoost for classification	Accuracy: 90.08%, Sensitivity: 91.46%, Specificity: 88.39%, F1-score: 90.43%
Doppala et al. [28]	IEEE Data Port	Naive Bayes, Random Forest, XGBoost, SVM	Mendeley Data Center cardiovascular disease dataset: 96.75% accuracy, IEEE Data Port: 93.39% accuracy, and 88.24% accuracy on the Cleveland dataset
Dinesh et al. [29]	IEEE Data Port	Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Logistic Regression	Accuracy: 85.12%
Paul et al. [24]	Cleveland Hungarian Statlog heart dataset, Cleveland processed heart dataset	Scaled conjugate gradient backpropagation in artificial neural networks	Minimum accuracy is 63.3803% for the Cleveland processed heart dataset and 88.4754% for the Cleveland Hungarian Statlog heart dataset
Baccouche et al. [30]	MIT-BIH Arrhythmia Database	Ensemble classifier with BiLSTM or BiGRU model with CNN model	Accuracy and F1-score between 91 and 96%
Dubey et al. [31]	IEEE CHD	Naive Bayes, Bayes Net, and Multilayer Perceptron	Accuracy: 93.67%
Mohan et al. [32]	UCI Dataset	Decision Trees, Neural Networks (NN), Support Vector Machines (SVM), and K-closest Neighbors (KNN)	Accuracy: 88.7%
Sharma et al. [34]	UCI Dataset	Random Forest, SVM, Naive Bayes, and Decision Tree ML techniques	The Naïve Bayes with a 90% accuracy rate, and Random Forest with an accuracy of 87%
Moreno-Sanchez et al. [35]	Heart Failure Survival Dataset	SCI-XAI automated data processing pipeline	Balanced Accuracy of 0.74 (std 0.03)
Sarra et al. [36]	Cleveland and Statlog (heart) datasets	The chi-squared-SVM technique	The suggested model improved accuracy from 85.29 to 89.7%
Reddy et al. [37]	Cleveland, Switzerland, Hungarian, V.A. Medical, and Statlog project heart disease datasets	SVM-linear, Naive Bayes, and Neural Network	Best Accuracy at 84.81%
An Dinh et al.[38]	National Health and Nutrition Examination Survey (NHANES) Dataset	Information gain of tree-based models identifies patient data characteristics	WEM performs best with an AU-ROC score of 83.1% without laboratory data and 83.9% with lab data
Shah et al. [39]	Cleveland database of UCI repository of heart disease patients	Naive Bayes, decision tree, K-nearest neighbor, and Random Forest algorithm	K-nearest neighbor scored best in accuracy
Nagavelli et al. [33]	UCI repository and Kaggle datasets	Weighted Naive Bayes, XGBoost, duality optimization, and XGBoost with DBSCAN and SMOTE-ENN	XGBoost with DBSCAN and SMOTE-ENN had the greatest precision, accuracy, f1-measure, and recall
Bhatt et al. [40]	Cleveland Clinic Foundation	MultiLayer Perceptron (MLP)	MLP has cross-validation accuracy of 87.28%, recall, precision, F1 score, and AUC scores of 84.85, 88.70, 86.71, and 0.95
Ours	IEEE Data Port	CardioTabNet	Accuracy: 94.08%, Precision: 92.84%, Recall: 97.37%, F1 Score: 94.47%, Specificity: 91.92%

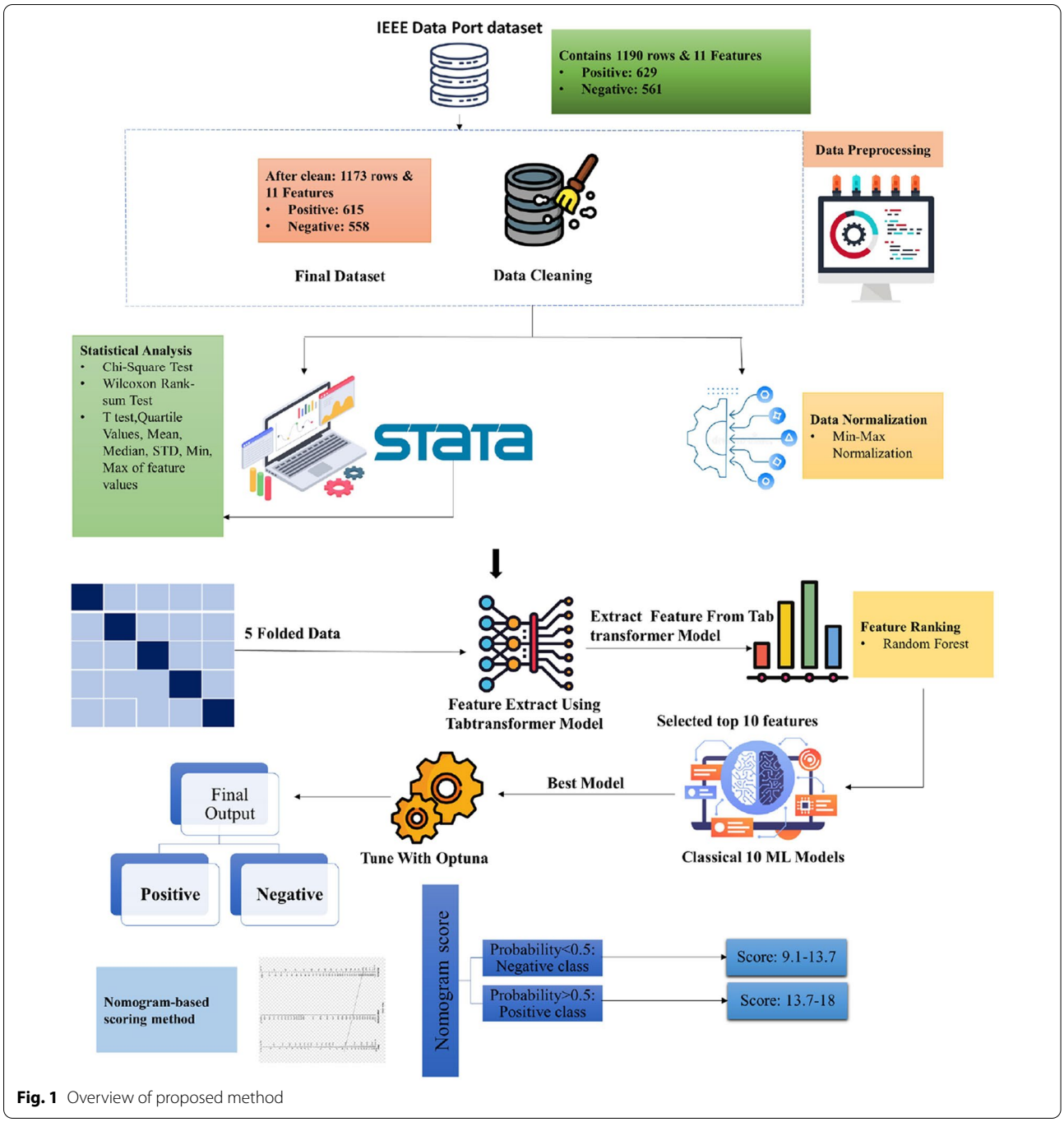

Fig. 1 Overview of proposed method

distinct statistical tests were conducted to derive the p-value. The Chi-square test was employed to determine the statistical correlation between the target characteristic and discrete-valued features, such as binary values. Before analysis, an assessment was made to see if the values of continuous characteristics followed a normal distribution. The t-test was employed to get the p-value if the data had a normal distribution. We used the Wilcoxon Rank-sum test when the feature values did not follow a normal distribution to determine the p-value.

Data splitting

The dataset consists of 1173 instances, each with 11 attributes after preprocessing. Implementing a fivefold cross-validation methodology improves the resilience, dependability, and applicability of our prediction models.

Table 2 Characteristics of the IEEE data port heart disease prediction dataset

Feature	Description
Age	Age of the patient (in years)
Sex	Gender of the patient (0 = Female; 1 = Male)
Chest pain type	Type of chest pain (1 = Typical Angina; 2 = Atypical Angina; 3 = Non-Anginal Pain; 4 = Asymptomatic)
Fasting blood sugar	Fasting blood sugar level (> 120 mg/dl; 0 = False; 1 = True)
Resting ECG results	Resting electrocardiogram results (0 = Normal; 1 = Having ST-T wave abnormality; 2 = Showing probable or definite left ventricular hypertrophy)
Cholesterol	Cholesterol level of the patient (measured in some unit, e.g., mg/dl)
Max heart rate	Maximum heart rate achieved during exercise (in bpm)
Exercise induced angina	Presence of exercise-induced angina (0 = No; 1 = Yes)
ST depression	ST depression induced by exercise relative to rest (in mm)
Peak exercise ST segment slope	Slope of the peak exercise ST segment (1 = Upsloping; 2 = Flat; 3 = Downsloping)
Presence or absence of heart disease	Target variable (0 = Absence of Heart Disease; 1 = Presence of Heart Disease)

Table 3 Summary of numerical and categorical features with weight column

Feature type	Feature names
Numerical	Age
	RestingBP (resting blood pressure)
	Cholesterol
	MaxHR (maximum heart rate)
	Oldpeak
	Weight column name
	FastingBS (Fasting Blood Sugar)
Categorical	Sex
	ChestPainType
	RestingECG
	ExerciseAngina
	ST_Slope

The dataset is divided into an 80% training set and a 20% test set to enable thorough evaluation.

Data normalization

Normalizing the input data is necessary to increase the ML models' training efficiency on our data. By ensuring that each feature contributes appropriately, this normalization enhances performance as a whole. We used the Standard Scaler technique to encourage robust training and generalization [45].

Feature ranking

Feature ranking assumes a pivotal role in ML [46], particularly in the context of datasets characterized by a considerable number of features. This step serves as a crucial precautionary measure, instrumental in addressing overfitting—an occurrence where a model excessively tailors itself to the nuances of the training data, thereby compromising its accuracy when applied to novel datasets. In our specific methodology, we employed feature ranking after feature extraction using the tab transformer Model built upon self-attention-based Transformers. Within the scope of this study, we utilized ML-based feature ranking techniques, with a specific emphasis on Random Forest.

SMOTE augmentation method

The Synthetic Minority Oversampling Technique (SMOTE) [47] generates synthetic samples for minority classes to address class imbalance in datasets. The method creates new samples along the line segments between an instance and its k-nearest neighbors, effectively balancing the class distribution. Therefore, the majority class is less likely to be biased, enhancing the model's performance and reliability. In addition to its ability to generate robust and unbiased predictions, SMOTE is a crucial step in preparing imbalanced datasets for machine learning. By incorporating artificial data points, SMOTE ensures that the minority class is more accurately represented. This study applied SMOTE to address the imbalance between classes in the training data. In the training dataset, there were 503 positive samples and 449 negative samples, resulting in 952 instances in total. With the application of SMOTE, the class distribution was balanced, with 503 positive samples and 503 synthetic negative samples in the training dataset, yielding 1,006 instances.

CardioTabNet model development

The CardioTabNet model is built upon the foundational elements of the tab transformer model [48] and classical ML models [49]. The tab transformer model, used for feature extraction, employs a tab transformer architecture based on self-attention Transformers. This architectural framework includes a column embedding layer and a series of N Transformer layers. The Transformer layers,

following the principles outlined by Vaswani et al. [50], consists of a multi-head self-attention layer and a position-wise feed-forward layer.

The tab transformer operates on feature-target pairs (x, y), encompassing categorical (x_{cat}) ad continuous (x_{cont}) features. Categorical features undergo embedding using a parametric technique known as Column embedding. This process involves multiple Transformer layers, each comprising a multi-head self-attention layer and a position-wise feed-forward layer. The self-attention mechanism facilitates contextual embedding by allowing each input embedding to attend to others. These contextual embeddings, concatenated with continuous features, are then directed into an MLP for target prediction. The unique identifier approach in column embedding is elucidated as a method tailored for embedding categorical features in tabular data. Each categorical feature entails an embedding lookup table with embeddings for each class, along with an additional one for missing values. The unique identifier distinguishes classes within a column, while the embedding encompasses a category-specific segment and a feature-value-specific component. The loss function governing model training is defined as cross-entropy for classification tasks or mean square error for regression tasks. The model parameters, encompassing those for column embedding (φ), Transformer layers (θ), and the top MLP layer (ψ), undergo learning through end-to-end training. Noteworthy is the tab transformer's unique identifier approach in column embedding, designed specifically for tabular data without positional encodings. The discussion also alludes to an ablation study comparing various embedding strategies, encompassing variations in dimensionality and the incorporation of unique identifiers.

Figure 2 illustrates a systematic procedure in which categorical characteristics are subjected to column embedding and pass through a transformer block, while numerical features undergo layer normalization. Afterward, the two sets of features are combined to produce the extracted features. After performing feature extraction using a tab transformer, a Random Forest (RF) method is utilized to rank the features and select the top 10. The next step entails the training of ten traditional ML models (Extra tree classifier, Random Forest Classifier, Gradient Boost Classifier, Cat Boost classifier, XGB Classifier, MLP classifier, Light Gradient Boosting Machine (LGBM) classifier, Linear Discriminant Analysis, and Logistic Regression). The model that performs the best is determined among these options. The selected model undergoes further refining through optimization using Optuna [51], to improve its prediction capabilities. This rigorous procedure concludes with the anticipation of the ultimate result.

Let (x, y) Represent a feature-target pair, where x consists of categorical features x_{cat} and continuous features $x_{\text{cont.}}$ The continuous features x_{cont} belong to Rc With c representing the number of continuous features. The categorical features (x_{cat}) are denoted by $\{x_1, x_2, \ldots, x_m\}$ where each x_i Is a categorical feature for $i = 1, \ldots, m$.

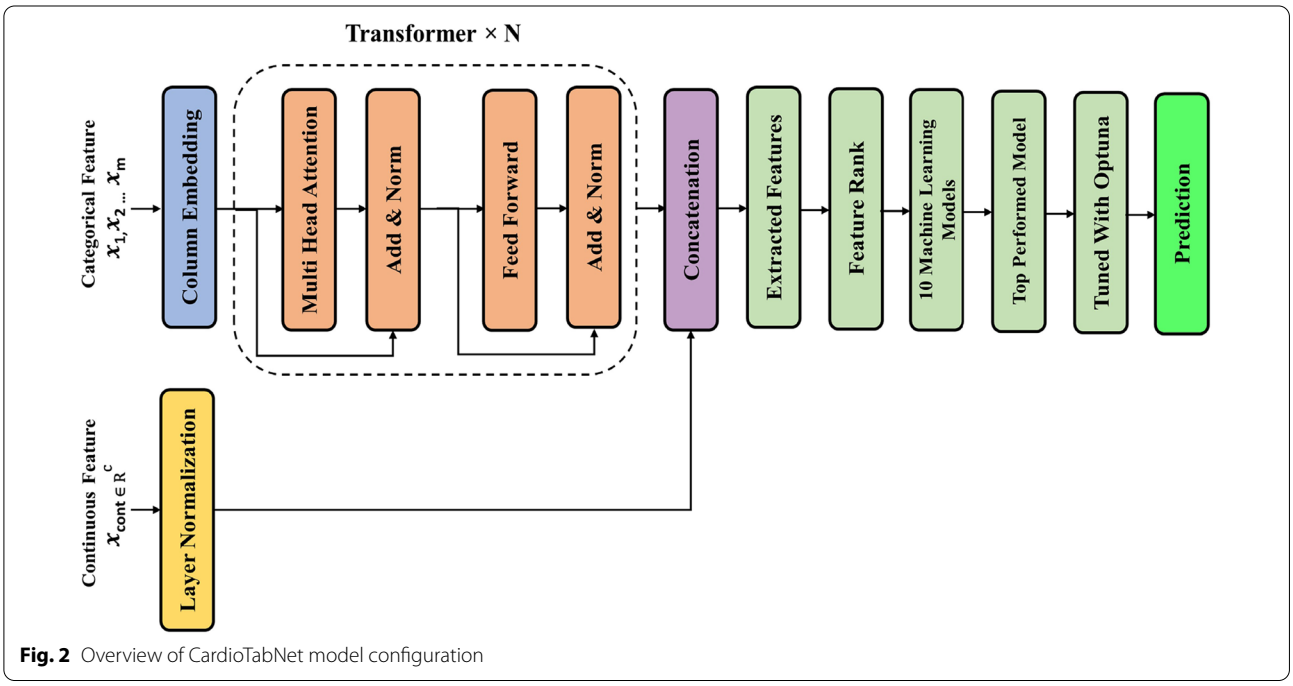

Fig. 2 Overview of CardioTabNet model configuration

Each categorical feature x_i Is embedded into a parametric vector of dimension d using column embedding. Let the embedding for x_i be denoted by $e\phi i(xi) \in R^{\,\mathrm{d}}$, with ϕ_i representing the embedding parameters forx_i. We denote the set of all categorical embeddings as:

$$E_\phi(x_{\text{cat}}) = \{e_{\phi_1}(x_1), \ldots, e_{\phi_m}(x_m)\} \tag{1}$$

The embeddings $E_\phi(x_{\text{cat}})$ are fed into the first Transformer layer. Each Transformer layer processes the input embeddings and outputs contextual embeddings by aggregating context from other embeddings through successive layers. A function denotes this transformation $f\theta$, which takes the parametric embeddings $\{e_{\phi_1}(x_1), \ldots, e_{\phi_m}(x_m)\}$ and outputs contextual embeddings $\{h_1, \ldots, h_m\}$, where each $h_i \in R^d fori = 1, \ldots, m$.

These contextual embeddings are concatenated with the continuous features x_{cont} to form a vector of dimension $(d \times m + c)$. This vector is then passed to a multilayer perceptron (MLP), represented by g_ψ, to predict the target y. The loss function $L(x, y)$, which could be cross-entropy for classification or mean squared error for regression, is minimized to learn all the parameters of tab transformer, including $\phi(columnembeddings)$, $\theta(Transformerlayers)$, and $\psi(MLPlayer)$. The loss function can be written as:

$$L(x, y) = H\left(g_\psi\left(f_\theta\left(E_\phi(x_{\text{cat}}), x_{\text{cont}}\right), y\right)\right) \tag{2}$$

The Transformer consists of multi-head self-attention layers followed by position-wise feed-forward layers, with each layer having element-wise addition and normalization. The self-attention layer uses three parametric matrices: Key (K), Query (Q),and Value (V). The embeddings are projected into these matrices to generate key, query, and value vectors. Let:

$$K \in R^{m\times k}, \quad Q \in R^{m\times k}, \quad V \in R^{m\times v} \tag{3}$$

where m is the number of embeddings inputted to the Transformer, and k and v are the dimensions of the key and value vectors, respectively. The attention mechanism calculates the relevance of other embeddings for each embedding using:

$$\text{Attention}\ (K, Q, V) = A \cdot V \tag{4}$$

where,

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{k}}\right) \tag{5}$$

The attention matrix $A \in R^{m\times m}$ calculates how much each embedding attends to others, resulting in contextual embeddings.

For each categorical feature iii, we maintain an embedding lookup table$e_{\phi_i}(\cdot)$. If the feature has d_i classes, the embedding table $e_{\phi_i} has(d_i + 1)$ embeddings, where the extra embedding is used for missing values. The embedding for a value $x_i = j$ is defined as:

$$e_{\phi_i}(j) = \left[c_{\phi_i}, w_{\phi_{ij}}\right] \tag{6}$$

where,

$$c_{\phi_i} \in R^{\ll}, \quad w_{\phi_{ij}} \in R^{d-\ll} \tag{7}$$

here, ℓ is a hyper-parameter that helps distinguish between different classes across columns, providing uniqueness.

As shown in Fig. 3, pseudocode for tab transformer-based feature selection and model training algorithm presents a structured approach for training ML models for tabular data. Preprocessing is the first step in the algorithm, which ensures that the data is appropriately tokenized and cleansed. Input data is processed using the tab transformer model to identify critical patterns and derive valuable features. After the extracted features are prioritized according to their significance, the top 20 are selected for further analysis. This is followed by the initialization and training of a ML model using the selected attributes. Finally, the algorithm estimates the accuracy of the trained model using test data and provides both the model's accuracy and

Input: D_{train}: Set of training data, D_{test}: Set of testing data

Output: F_{train}, F_{test}: Top N selected features, M_{ML}: Trained machine learning model

1. Preprocessing
   For each d $\in D_{train}$ and d $\in D_{test}$:
   1. *Clean d*
   2. Normalize and prepare data

   End For
2. Feature Extraction
   Initialize F_{train} andF_{test}
   1. Extract features for D_{train} and D_{test} using TabTransformer:
      $F_{train} \leftarrow$ TabTransformer. $extract_{features(D_{train})}$
      $F_{test} \leftarrow$ TabTransformer. $extract_{features(D_{test})}$
3. Feature Ranking and Selection
   1. Rank features based on importance:
      Ranked_features $\leftarrow$ FeatureRanking(F_{train}, Y_{train})
   2. Select top N features:
      $F_{train\_20} \leftarrow Select\ Top\ N\ Features\ (Ranked\_features, F_{train}, N = 20)$
      $F_{test\_20} \leftarrow Select\ Top\ N\ Features\ (Ranked\_features, F_{test}, N = 20)$
4. Model Initialization
   Initialize the machine learning model:
   $M_{ML} \leftarrow InitializeMLModel()$
5. Model Training:
   $Train\ M_{ML}\ on\ F_{train\_20}, Y_{train}$
6. Model Evaluation:
   1. $Predict\ Y_{test} \leftarrow M_{ML}.predict(F_{test\_20})$
   2. ComputeAccuracy: Accuracy $= \frac{1}{|Y_{test}|}\sum_{i=1}^{|Y_{test}|} \mathbb{1}\left(Y_{test}^{(i)} = Y_{pred}^{(i)}\right)$
7. Output
   $Return\ M_{ML}, Accuracy$

Fig. 3 Pseudocode for the tab transformer-based feature selection and model training algorithm

performance metrics. As a result of this workflow, it is possible to improve predictive performance by selecting features efficiently and training models.

Nomogram-based scoring system

Nomograms [52] are visual representations that condense statistical prediction models into a singular numerical estimation of the likelihood of an occurrence, specifically customized for an individual patient's characteristics. This leads to developing a grading system that can substantially aid healthcare practitioners in quickly differentiating between good and negative categories. Graphical interfaces that are easy for users to use are available to generate these estimates. These interfaces make it easier for healthcare professionals to use nomograms during clinical encounters, providing valuable information for clinical decision-making.

Our investigation culminates in presenting a nomogram, a tool clinicians favor for its frequent utilization during investigative processes. A nomogram was constructed in our inquiry using the most optimal features identified by the random forest feature ranking technique. In addition, calibration took place aligning it with the ground truth labels and the forecast likelihood of the occurrences. Decision Curve Analysis (DCA) was used to present a threshold for each attribute graphically. The aforementioned actions were performed with Stata 15.

Machine learning models

In our study, we implemented feature extraction using the tab transformer model and subsequently trained ten classical ML models (Extra tree classifier, Random Forest Classifier, Gradient Boost Classifier, Cat Boost classifier, XGB Classifier, MLP classifier, Light Gradient Boosting Machine (LGBM) classifier, Linear Discriminant Analysis (LDA), and Logistic Regression) [49]. The aim was to leverage the rich representations learned by tab transformer to enhance the predictive capabilities of these traditional models.

Evaluation metrics

The efficacy of the model cannot be evaluated based solely on its accuracy. To enhance the dependability of the findings, a comprehensive range of evaluation criteria was implemented, acknowledging that exclusive reliance on precision was inadequate [53]. A variety of metrics can be identified by utilizing the subsequent formulations, which span from Eq. 8 to Eq. 12:

$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision(P) = \frac{TP}{TP + FP} \tag{9}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{10}$$

$$Specificity(S) = \frac{TN}{TN + FP} \tag{11}$$

$$F1 - Score(F1) = 2 * \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

where True Positive, True Negative, False Positive, and False Negative are denoted, respectively, by the letters TP, TN, FP, and FN. The model performance can be obtained by evaluating the classification performance using the ROC (Receiver Operating Characteristic), AUC (area under the curve), and confusion matrix.

Experimental setup

The research utilized Python 3.10 and the scikit-learn package to implement all models. Scikit-learn [54] is a prevalent Python package for ML, constructed upon NumPy, SciPy, and matplotlib. It offers a straightforward interface for executing both supervised and unsupervised learning algorithms, including classification, regression, clustering, and dimensionality reduction. The training of these models adhered to specific hardware specifications, including an Nvidia GForce 1050ti GPU, an AMD Ryzen 7 5800X 8-Core Processor, and 32 GB of high RAM. We employed a straightforward pipeline for this experiment, integrated with all classical ML models.

Table 4 shows the key hyperparameters in the tab transformer-based model, such as learning rate, weight decay, dropout rate, and batch size, as well as architecture-specific parameters like transformer blocks,

Table 4 Hyperparameter details for the tab transformer-based model

Hyperparameter	Value
Learning_Rate	0.001
Weight_Decay	0.0001
Dropout_Rate	0.2
Batch_Size	265
Num_Epochs	500
Optimizer	Adam
Num_Transformer_Blocks	3
Num_Heads	4
Embedding_Dims	8
MLP_Hidden_Units_Factors	[2, 1]
Num_MLP_Blocks	2

attention heads, and MLP hidden units. All these hyperparameters are selected based on heuristic approach. To optimize the performance of the model, these hyperparameters are crucial.

Results

This section presents the top feature importances obtained from the TabTransformer models, followed by the results of statistical analysis and traditional machine learning models. Additionally, it includes correlation analysis to identify key features, as well as nomogram and SHAP (SHapley Additive exPlanations) analyses for interpretability.

Feature ranking

The CardioTabNet model was employed to extract features, leveraging the capabilities of the tab transformer. Subsequently, a Random Forest Model was utilized to rank the extracted features. In comparison to alternative models such as XGBoost and ExtraTree classifier, the Random Forest model demonstrated superior performance. The selection process focused on identifying the top 10 features, which were then utilized for the final prediction. Figure 4 visually presents the top 20 features extracted through the tab transformer, providing a comprehensive illustration of the key contributors to the predictive modeling process.

Table 5 presents the extracted features, which were identified as the most correlated with the original features, as shown in Fig. 6. Further evaluation of the relative importance of these features was performed using a Random Forest-based feature selection model. Notably, chest_pain_type_atypical angina and cholesterol frequently appeared among the top-ranked extracted features, indicating their high predictive power. A consistent presence of cholesterol-related features (e.g., features 72, 74, 75, 76, and 77) indicates the importance of cholesterol measurements in the model's decision-making. In addition, as shown in the table, both sex_male and rest_ecg_ left ventricular hypertrophy were included among the top features, indicating their relevance in classifying individuals. Based on these findings, the extracted features have a high degree of interpretability, and their alignment with clinically relevant variables is validated.

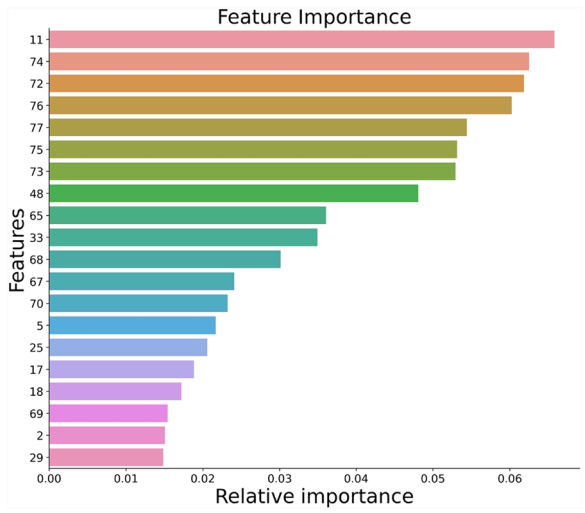

Fig. 4 Top-20 extracted features ranked using the Random Forest feature selection algorithm

Table 5 Names of the extracted Top features

Feature no	Extracted feature	Most related original feature
F1	11	chest_pain_type_atypical angina
F2	74	Cholesterol
F3	72	Cholesterol
F4	76	Cholesterol
F5	77	Cholesterol
F6	75	Cholesterol
F7	73	Cholesterol
F8	48	st_slope_flat
F9	65	st_slope_upsloping
F10	33	st_slope_upsloping
F11	68	st_slope_upsloping
F12	67	st_slope_upsloping
F13	70	st_slope_upsloping
F14	5	sex_male
F15	25	chest_pain_type_typical angina
F16	17	chest_pain_type_non-anginal pain
F17	18	st_slope_upsloping
F18	69	sex_male
F19	2	sex_male
F20	29	st_slope_upsloping

Feature name identification using correlation analysis

A correlation-based analysis was performed to associate each extracted feature with its most similar original feature. Because the extracted features are unlabeled and derived from the Transformer model, determining their correspondence to meaningful variables is essential for interoperability. As discussed in Interpretable Machine Learning by Christoph Molnar [55], correlation analysis between extracted and original features may serve as an initial, model-independent approximation of feature relevance. By comparing the values of an extracted feature to those of its original, it determines how closely a feature aligns with itself. In the presence of a strong positive or negative correlation, the extracted feature captures the

same information as the original feature. By identifying the original feature with the highest correlation, we can infer the likely representation of each extracted feature.

The correlation heatmap in Fig. 5 displays the relationship between 78 extracted features (Y-axis) and the original input features (X-axis). In each cell, the Pearson correlation coefficient is displayed between an extracted feature and an original feature, with red indicating positive correlations and blue indicating negative correlations. The strength of correlation is indicated by the intensity of the color. The extracted features were matched to the original feature with the highest absolute correlation to give them meaningful context. Semantic interpretation of the extracted features, supporting the analysis of feature importance and promoting greater transparency of the model, was enabled by this approach. Therefore, the Y-axis represents extracted feature indices aligned with their most closely related original features.

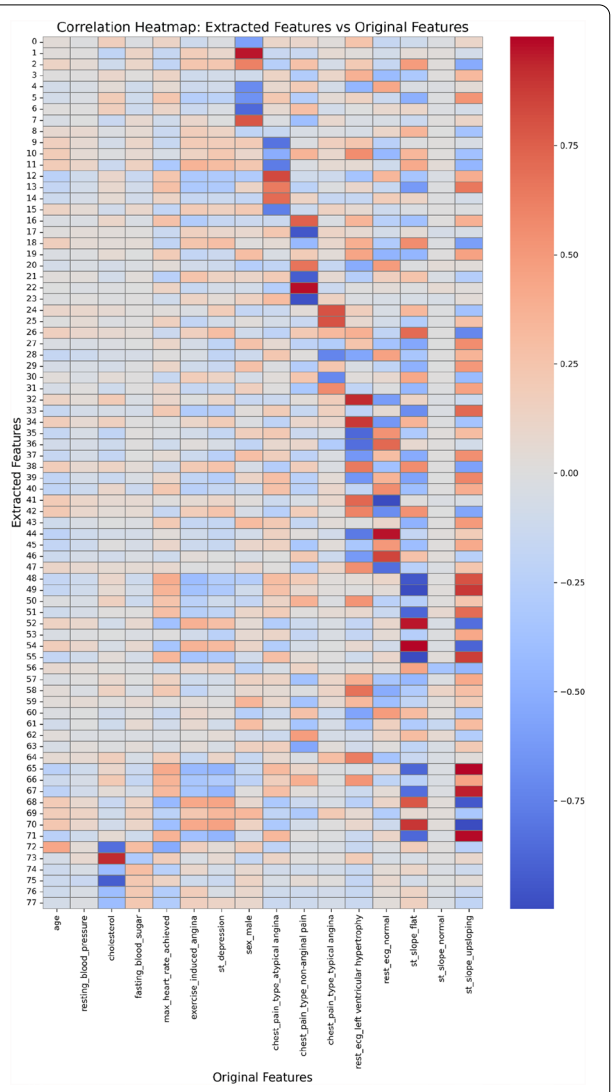

Fig. 5 Correlation heatmap between extracted features and original input features

Statistical analysis

A statistical analysis of the dataset's characteristics is shown in Table 6**.** We are using the *p*-value to represent the difference between two observations, which in our case is the target variable and its corresponding feature column. Based on the lower *p*-value, the differences between the two columns are not the result of random events, but rather of a strong association between them. As a result, a lower value of 0.05 or less than *p* value would provide stronger significance. As a result of the following table, we can gain some insight into the statistical association between features and output labels. All other biomarkers carry a value less than 0.05, except rest_ecg_left ventricular hypertrophy and st_slope_normal. There are 89.11% of positive cases in men, representing a significant portion of the population affected by the disease. There was a higher percentage of older patients in the disease-positive group in the study, which included individuals aged 28 to 77.

Classification using CardioTabNet model

In Table 7, the results of the first feature extraction are presented, along with their accuracy for each fold. As part of the initial analysis, features were extracted using the tab transformer model from a dataset divided into five folds. By applying the Random Forest algorithm, a comprehensive feature ranking process was conducted, which enabled the identification and selection of the top 10 features. Based on the selected top 10 features, an ensemble of 10 ML models was trained.

As a result of feature extraction using the tab transformer, Table 8 presents the top 10 ML models. ExtraTree Classifier achieved the highest accuracy of 92.82%, with notable precision, recall, and F1 score values of 91.04%, 95.18%, and 93.02%, respectively. Additionally, the Random Forest Classifier achieved robust precision, recall, and F1 score metrics, achieving an accuracy of 92.18%. CatBoost_untuned and Gradient Boosting Classifier both produced competitive results, showing their effectiveness in classifying data. Both the XGB Classifier and AdaBoost Classifier performed well in terms of precision and recall, though they displayed slightly lower accuracy. As a result of the MLP Classifier and LGBM models, consistent results were obtained, supporting their reliability. The results of Linear Discriminant Analysis and Logistic Regression were notable for their relatively low sensitivity (recall) when compared with other classifiers despite achieving good accuracy.

Table 6 Statistical analysis on IEEE port heart disease dataset

S. no	Feature name	Positive class	Negative class	Total	Test	Test statistic	p-value
1	Gender				Chi-square test	114.7073	< 0.001
	• Male (%)	548 (89.11%)	349 (62.54%)	897 (76.47%)			
	• Female (%)	67 (10.89%)	209 (37.45%)	276 (23.53%)			
2	Age				T-test	− 9.4233	< 0.001
	• Mean ± STD	56.07 ± 8.63	51.09 ± 9.46	53.7 ± 9.37			
	• Median	57.00	51.00	54.00			
	• Q1, Q3	51, 62	44, 57	47, 60			
	• Min, Max	31, 77	28, 76	28, 77			
3	resting_blood_pressure				Rank-sum test	− 3.884	< 0.001
	• Mean ± STD	133.67 ± 17.73	129.74 ± 16.31	131.8 ± 17.17			
	• Median	131.00	130.00	130			
	• Q1, Q3	120, 144	120, 140	120, 140			
	• Min, Max	92, 185	80, 180	80, 185			
4	Cholesterol				T-test	7.097	< 0.001
	• Mean ± STD	190.4 ± 116.86	230.45 ± 67.31	209.45 ± 98.53			
	• Median	226.00	231.00	229.00			
	• Q1, Q3	135, 271	201, 267	188, 269			
	• Min, Max	0, 491	0, 468	0, 491			
5	fasting_blood_sugar				Chi-square test	54.11	< 0.001
	• True (%)	182 (29.59%)	67 (12.01%)	249 (21.23%)			
	• False (%)	433 (70.41%)	491 (87.99%)	924 (78.77%)			
6	max_heart_rate_achieved				T-test	15.78	< 0.001
	• Mean ± STD	129.77 ± 23.20	150.94 ± 22.68	139.84 ± 25.27			
	• Median	128.00	153.50	141.00			
	• Q1, Q3	113, 147	137.25, 169.00	121, 160			
	• Min, Max	67, 182	69, 202	67, 202			
7	exercise_induced_angina				Chi-square test	274.26	< 0.05
	• Yes (%)	376 (61.14%)	78 (13.98%)	454 (38.70%)			
	• No (%)	239 (38.86%)	480 (86.02%)	719 (61.30%)			
8	st_depression				Rank-sum test	− 14.152	< 0.001
	• Mean ± STD	1.34 ± 1.18	0.46 ± 0.73	0.92 ± 1.083			
	• Median	1.20	0.00	0.60			
	• Q1, Q3	0.1, 2	0, 0.8	0, 1.6			
	• Min, Max	− 2.60, 6.2	− 1.1, 4.2	− 2.60, 6.2			
9	chest_pain_type_atypical angina				Chi-square test	157.21	< 0.001
	• Yes (%)	29 (4.72%)	184 (32.97%)	213 (18.16%)			
	• No (%)	586 (95.28%)	374 (67.03%)	960 (81.84%)			
10	chest_pain_type_non-anginal pain				Chi-square test	65.25	< 0.001
	• Yes (%)	87 (14.15%)	191 (34.23%)	278 (23.70%)			
	• No (%)	528 (85.85%)	367 (65.77%)	895 (76.30%)			
11	chest_pain_type_typical angina				Chi-square test	5.94	< 0.05
	• Yes (%)	25 (4.07%)	41 (7.35%)	66 (5.63%)			
	• No (%)	590 (95.93%)	517 (92.65%)	1107 (94.37%)			
12	rest_ecg_left ventricular hypertrophy				Chi-square test	0.9152	0.34
	• Yes (%)	174(28.29%)	144(25.81%)	318(27.11%)			
	• No (%)	441(71.71%)	414(74.19%)	855(72.89%)			
13	rest_ecg_normal				Chi-square test	13.36	< 0.001
	• Yes (%)	323(52.52%)	352(63.08%)	675(57.54%)			
	• No (%)	292(47.48%)	206(36.92%)	498(42.46%)			

Table 6 (continued)

S. no	Feature name	Positive class	Negative class	Total	Test	Test statistic	p-value
14	st_slope_flat				Chi-square test	312.36	< 0.001
	• Yes (%)	451 (73.33%)	121 (21.68%)	572 (48.76%)			
	• No (%)	164 (26.67%)	437 (78.31%)	601 (51.24%)			
15	st_slope_normal				Chi-square test	0.91	0.34
	• Yes (%)	1 (0.16%)	0 (0%)	1 (0.09%)			
	• No (%)	614 (99.84%)	558 (100%)	1172 (99.91%)			
16	st_slope_upsloping				Chi-square test	384.52	< 0.001
	• Yes (%)	107 (17.40%)	415 (74.37%)	522 (44.50%)			
	• No (%)	508 (82.60%)	143 (25.63%)	651 (55.50%)			
17	Target (%)	615 (52.43%)	558 (47.57%)	1173			

Q1 = First quarter, Q3 = Third quarter

Table 7 Result of feature extraction with tab transformer

Fold	Accuracy (%)
1	89.74
2	90.12
3	88.95
4	89.30
5	90.04

(a) Hyperparameter Tuning with Optuna

Optuna, a Bayesian optimization library, efficiently tunes hyperparameters for ML models [51]. A meticulous hyperparameter tuning process was conducted utilizing Optuna to optimize the parameters of the ExtraTree Classifier, resulting in the identification of the best-performing model configuration. The optimal hyperparameters, determined through the hyperparameter tuning process with Optuna, were as follows: 'n_estimators' set to 176, 'max_depth' set to 19, and 'bootstrap' set to True. Table 9 showcases the five-fold cross-validated results for the ExtraTree Classifier after hyperparameter tuning with Optuna, which was identified as the best-performing model initially. Across all folds, the tuned ExtraTree model consistently demonstrated high accuracy, precision, recall, and F1 score values, affirming its robust performance in the classification task. Notably, the model achieved an average accuracy of 94.089%, with a balanced and elevated performance across all evaluated metrics. These results underscore the efficacy of the hyperparameter tuning process in enhancing the model's overall predictive capability and generalizability.

The Receiver Operating Characteristic (ROC) curve, presented in Fig. 5a, illustrates the performance of the tuned ExtraTree classifier model across five-fold data in the context of heart disease classification. The Area Under the Curve (AUC) serves as a comprehensive metric, representing the model's overall performance by averaging the True Positive Rate (TPR) and False Positive Rate (FPR) across various threshold settings. The AUC

Table 8 Result for the Top-10 ML after feature extraction with tab transformer

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Specificity (%)	AUC (%)
ExtraTree classifier	92.82	91.04	95.18	93.01	90.58	92.88
Random Forest Classifier	92.17	89.77	95.41	92.46	88.90	92.16
Gradient Boosting Classifier	91.92	90.30	94.14	92.13	89.76	91.95
CatBoost	90.00	88.05	92.88	90.32	87.17	90.03
XGB Classifier	86.02	84.45	89.30	86.57	83.02	86.16
AdaBoost Classifier	84.10	83.15	86.77	84.67	81.79	84.28
MLP Classifier	84.10	81.75	88.31	84.83	79.76	84.03
LGBM	83.84	81.93	88.08	84.62	79.92	84.00
Linear discriminant analysis	82.43	81.95	84.02	82.89	81.02	82.52
Logistic regression	82.43	81.96	83.98	82.90	80.98	82.48

Table 9 Results on extra-tree classifier after tuning with the Optuna

Fold	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Specificity (%)	AUC (%)
1	ExtraTrees (Tuned with Optuna)	94.23	92.77	96.25	94.47	92.10	94.95
2		97.43	97.64	97.65	97.64	97.18	98.41
3		90.38	84.14	97.18	90.19	84.70	90.95
4		92.90	90.69	96.3	93.41	89.18	95.11
5		95.48	93.90	97.47	95.65	93.42	97.63
Average		94.08	92.84	97.37	94.47	91.92	95.01

score of 0.95 for the tuned ExtraTree classifier Model signifies its excellence in distinguishing between positive and negative cases, further affirming its efficacy in the heart disease detection task.

The confusion matrix and associated metrics for the Tuned ExtraTree classifier Model are visually depicted in Fig. 6b. The illustration underscores the model's exemplary performance in the classification task, demonstrating high accuracy, precision, recall, and F1 score values. These metrics collectively affirm the model's capability to accurately classify both positive and negative cases, underscoring its proficiency in the task at hand.

Nomogram-based scoring technique

In this study, we developed a predictive outcome model through regression analysis, and we visualized it through a multivariate logistic regression nomogram. A statistical pipeline was developed to combine the ten most significant features derived from the feature extraction process previously outlined. The supplementary Table 1S provides a comprehensive breakdown of the regression analysis used in the construction of the nomogram, detailing each feature's regression coefficient. Based on these coefficients, the nomogram assigns scores to individual features that are directly proportional to their predictive power. Clinicians perform a risk assessment by summing these individual scores to obtain a 'Total Score', which is shown at the bottom of the nomogram and corresponds to the probability of heart disease. A nomogram for predicting outcome probability is shown in Fig. 7. In particular, st_slope_upsloping and chest_pain_type_atypicalangina contribute positively to a high predicted risk since they increase the total score. On the other hand, features such as rest_ecg_normal or lower cholesterol values contribute negatively, lowering the total risk score. A structured scoring system facilitates clinicians' ability to quantify and communicate a patient's individualized risk, facilitating customized clinical decisions.

In this study, internal validation is used to develop a calibration technique to analyze and measure the quality

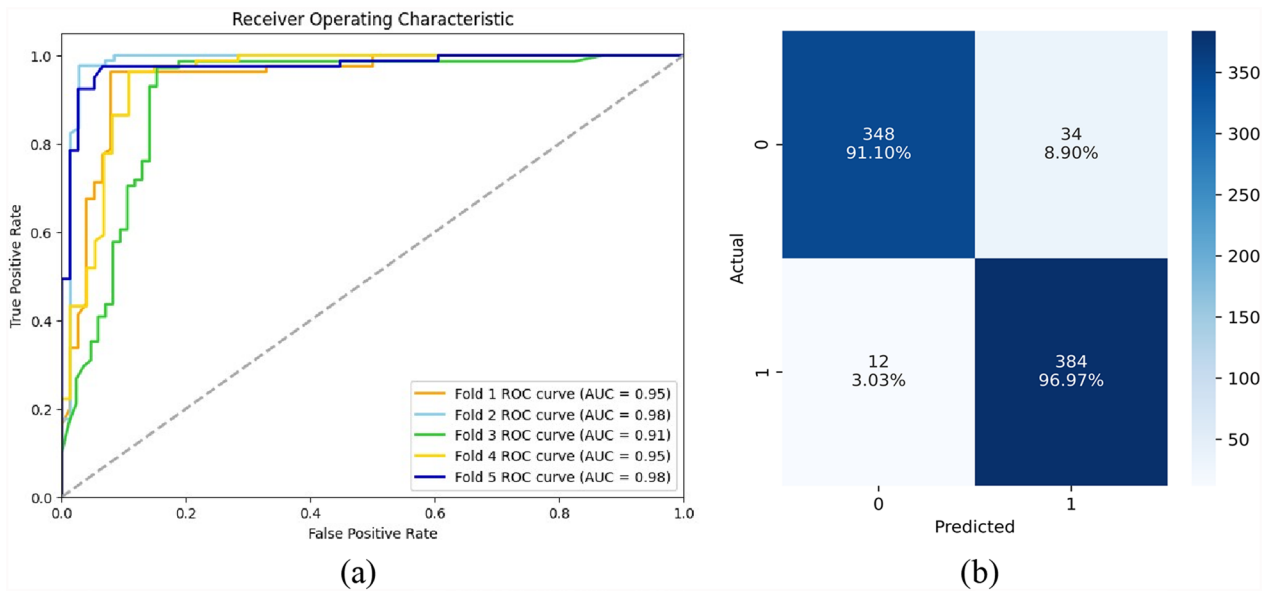

Fig. 6 **a** ROC curve of each fold for tuned ExtraTree Classifier, **b** Confusion matrix for tuned extra tree classifier

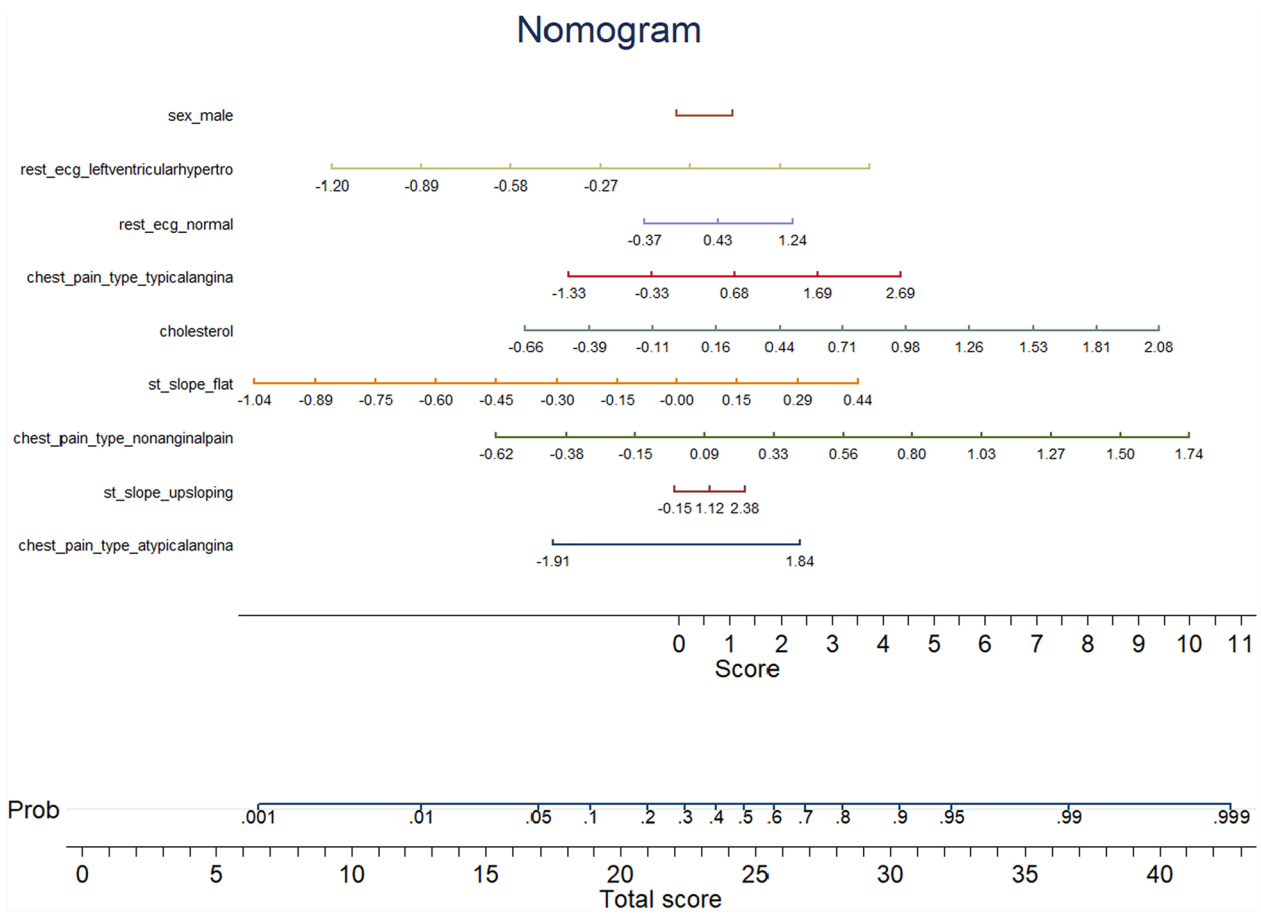

Fig. 7 Nomogram for predicting outcome probability

of logistic regression's probability estimations for predicting the actual output labels shown in Fig. 8a. There is an equal distribution between areas above and below the diagonal bisector, which accounts for a significant portion of the calibration belt. As a result, the expected probabilities were perfectly aligned with the actual

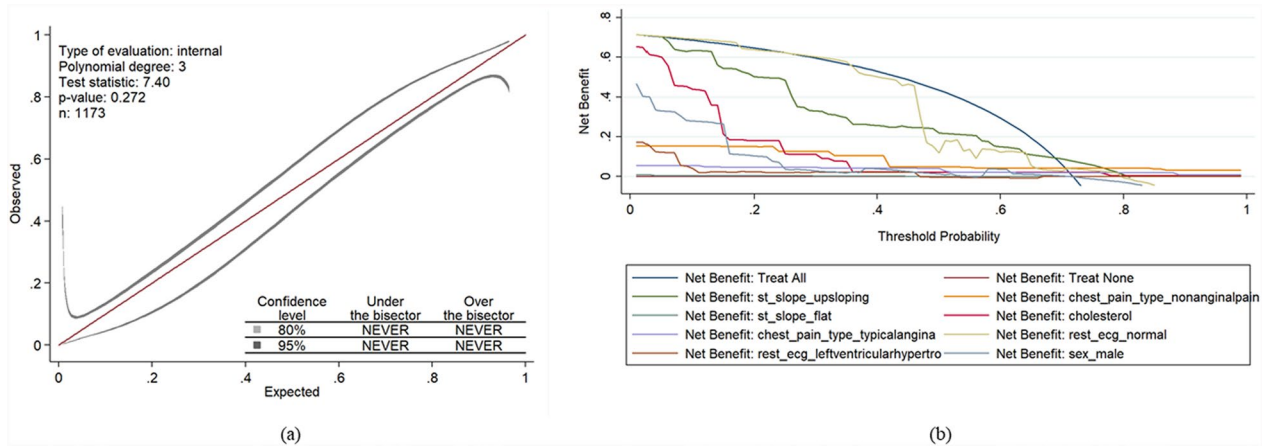

Fig. 8 **a** Calibration plot, **b** decision curve analysis

results. Figure 8b illustrates the advantages provided by each feature included in the study in identifying positive versus negative classes for the model.

SHAP analysis

By quantifying the impact of each feature on a given prediction, SHAP (SHapley Additive Explanations) [56] offers a consistent framework for interpreting machine learning models. This algorithm is based on cooperative game theory, which ensures that the output of the model is allocated fairly to the input features. The technique supports both global interpretabilities, highlighting the importance of overall features, as well as local interpretability, explaining individual predictions. A SHAP analysis was used in our study to discover how each feature influenced the model's decisions, improving its transparency and reliability.

Figure 9 illustrates the SHAP summary plot for the best-performing machine learning model, highlighting the impact of each feature on its predictions. In terms of top contributors, st_slope_upsloping is the most influential when its value is high, while cholesterol is often the most detrimental when its value is high. The presence of features related to chest pain, such as atypical angina, typical angina, and nonanginal pain, consistently reduces the model's predicted outcome. Additionally, st_slope_flat tends to reduce the prediction output when it is elevated. There is a slight downward effect for the feature sex_male, indicating that being male slightly decreases the predicted risk. SHAP-based interpretation enhances transparency and aligns well with domain-specific knowledge in cardiovascular prediction.

Discussion

Heart disease, or cardiovascular disease, encompasses a range of disorders that impact the heart and blood arteries. It is a prominent factor contributing to illness and death on a global scale, affecting a significant number of individuals. Cardiovascular disorders, which can be classified into four kinds including coronary heart disease, heart failure, congenital heart disease, and cardiomyopathy, provide a substantial risk [57]. Timely and accurate detection of cardiac disease is crucial to prevent worsening and save lives. Early detection and prompt intervention in cardiovascular diseases play a critical role in reducing premature mortality [58]. Predictive models are essential tools for identifying individuals at risk, enabling proactive healthcare interventions. Recognizing these diseases early is crucial for effective intervention and improved overall health outcomes. ML presents a transformative approach to addressing the complex challenge of cardiovascular diseases, offering innovative solutions for early detection, risk prediction, and personalized healthcare interventions [19–26].

In this study, we propose a novel model, CardioTabNet, for cardiovascular disease prediction. The use of Transformers, originally designed for natural language processing tasks [59, 60], in tabular data represents a significant advancement in machine learning and predictive modeling. Now, the transformers are also used in medical image classification, segmentation applications [61, 62]. Transformers, particularly those that rely on self-attention mechanisms, have exhibited exceptional efficacy in discerning complex relationships present in structured and sequential data [46]. Self-attention mechanisms enable transformers to evaluate the relative significance of individual features to others. This approach has the potential to yield a more intricate comprehension of feature interactions in contrast to conventional models that might fail to adequately capture such interactions. It can process unprocessed data more efficiently and necessitate less manual feature engineering in comparison to conventional models, owing to their sophisticated architecture. This can yield a substantial benefit by decreasing the amount of time and effort needed to preprocess the data. Consequently, they have also been found to apply to tabular datasets. Transformers are excellent at identifying long-range data dependencies [50]. Transformers are successful in capturing the relationships between distant categorical features in the context of tabular data [50], where features may have intricate interactions and dependencies.

The initial phase involves a comprehensive statistical analysis of the dataset using Stata/MP version 15.00. Statistical measures such as mean, median, standard deviation (STD), 25th and 75th quartile values, as well as the mean and maximum values, were employed for a thorough characterization of the dataset which was presented in Table 6. The CardioTabNet model leverages tab transformer for feature extraction, which is based on self-attention transformers. The transformer layers transform

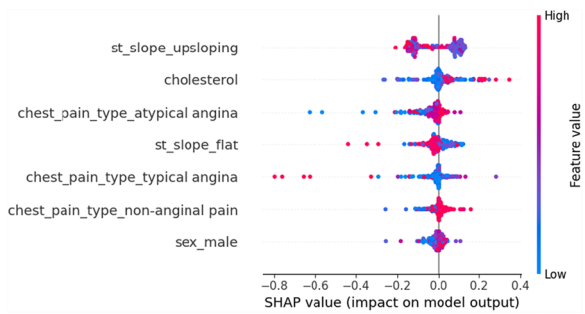

Fig. 9 SHAP summary plot for the best-performing machine learning model

categorical feature embeddings into robust contextual embeddings. Tab transformers possess the capability to accommodate various feature types, encompassing both categorical and numeric values. The self-attention mechanism facilitates the model's ability to seamlessly process mixed-type data by permitting it to allocate various levels of importance to distinct features. Particularly advantageous in real-world datasets where diverse feature types are prevalent is this adaptability. Following feature extraction, RandomForest is employed for feature ranking, a pivotal step in addressing overfitting and ensuring model accuracy on novel datasets. The top 10 extracted features are then utilized to train 10 ML models, with detailed results presented in Table 8. The GradientBoostingClassifier and CatBoost models demonstrated competitive results, emphasizing their effectiveness in the classification task. Although the XGBClassifier and AdaBoostClassifier displayed slightly lower accuracy, they maintained respectable precision and recall values, contributing to their overall reliability. Consistent and reliable outcomes were observed with the MLPClassifier and LGBM models.The ExtraTree Classifier emerges as a standout performer with an accuracy of 92.82%, accompanied by noteworthy precision, recall, and F1 score values of 91.04, 95.18, and 93.02%, respectively. Subsequent tuning of the ExtraTree classifier model using Optuna further enhances its performance, achieving an average accuracy of 94.08% across all evaluated metrics. The analysis extends to the examination of the ROC curve, revealing an AUC score of 95.01% for the tuned ExtraTree classifier Model. This underscores its exceptional ability to distinguish between positive and negative cases, reinforcing its efficacy in heart disease detection.

Logistic regression and nomogram analysis are employed to enhance binary classification accuracy and facilitate differentiation between positive and negative classes. Regression analysis, utilizing the 10 most significant features obtained from the extraction process, culminates in the construction of a predictive outcome model visualized through a multivariate logistic regression nomogram. This comprehensive statistical pipeline contributes to a nuanced understanding of cardiovascular disease prediction. Finally, SHAP (SHapley Additive exPlanations) analysis was performed to interpret the model's predictions, revealing key features influencing cardiovascular risk. The results aligned with clinical insights, enhancing the model's transparency and reliability. Among these features, st_slope_upsloping demonstrated the highest positive impact on the predicted risk of cardiovascular disease.

A comparative analysis is presented in Fig. 10 alongside other works in the IEEE Port Heart Disease dataset. Our CardioTabNet model attained an accuracy

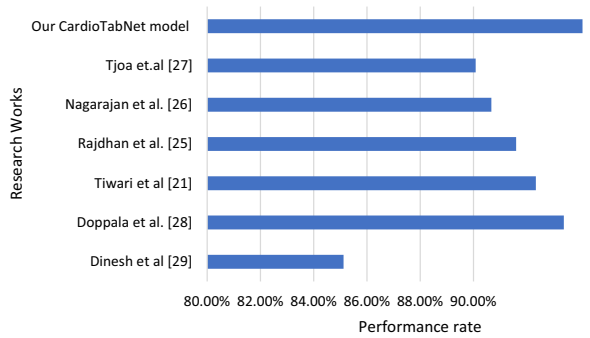

Fig. 10 Comparative analysis with previous studies

of 94.09%, surpassing the accuracy reported by previous studies. Specifically, Doppala et al. [28] achieved an accuracy of 93.39%, Dinesh et al. reported 85.12%, Tjoa et al. [27] achieved 90.08%, Tiwari et al. reported 92.34%, Rajdhan et al. [25] achieved 91.60%, and Nagarajan et al. [26] reported 90.67%. Comparing these results, it's evident that our CardioTabNet model outperforms previous studies in terms of accuracy. This highlights the effectiveness of our model in accurately predicting cardiovascular diseases. The superior performance of CardioTabNet underscores its potential as an advanced tool for cardiovascular disease prediction, offering enhanced accuracy and reliability compared to existing approaches.

For future studies, tab transformer can be applied to different biological data. The integration of attention mechanisms, especially in models like tab transformer, enhances the capacity to capture hidden features and interactions in biological data and networks. This application minimizes reliance on extensive feature engineering and effectively addresses the inherent ambiguity in biological datasets. Attention mechanisms, adept at handling noisy and conflicting information, dynamically prioritizing features, making the model robust to complex biological scenarios. This approach holds promise for future applications across diverse biological data types, offering a streamlined and adaptable method for meaningful feature extraction and interpretation.

Limitations and future work

CardioTabNet, a novel model for the prediction of cardiovascular diseases, is presented in this study. Based on clinical cardiovascular data, the model employs transformer technology to extract a high-quality feature space. In structured and sequential data, transformers, particularly those that use self-attention mechanisms, have proven highly effective at identifying complex patterns. This study used data from the IEEE Data Port, which provides access to a variety of engineering and technology

datasets. We used a comprehensive heart disease dataset that combines five well-known datasets: the Hungarian dataset, the Cleveland dataset, the Long Beach VA dataset, the Switzerland dataset, and the Statlog dataset. A total of 1190 instances with 11 features relevant to heart disease diagnosis make up this combined dataset.

However, the relatively modest size of the dataset is one of the limitations of this study. Despite the variety of data, the dataset does not fit the definition of a large-scale dataset, which may make it difficult for the model to generalize to a broader and more heterogeneous population. We plan to address this limitation by expanding our dataset to include data from larger and more diverse populations in the future. As part of our next objective, we will collect a custom dataset tailored specifically for cardiovascular diseases, which will lead to more comprehensive model training and enhance the model's generalization capabilities across a variety of demographic and clinical settings. As a result, CardioTabNet will be further validated in real-world scenarios for robustness and effectiveness.

Conclusion

In conclusion, this study introduces CardioTabNet, a novel model designed for the prediction of cardiovascular diseases. The model incorporates transformer technology to extract a high-quality feature space, employing a feature ranking strategy based on clinical cardiovascular data. Transformers, particularly those leveraging self-attention mechanisms, have demonstrated remarkable effectiveness in identifying intricate links within structured and sequential data. The methodology involves a multi-step approach: initial feature extraction utilizing the tab transformer model, followed by training with ten classical ML classifiers. Our comprehensive statistical analysis demonstrates the efficacy of CardioTabNet, particularly when paired with the tuned ExtraTree Classifier, which exhibits outstanding performance. The model achieves an impressive average accuracy of 94.08%. Noteworthy precision, recall, and F1 score values further support its effectiveness, standing at 92.84, 97.37, and 94.47%, respectively. This remarkable performance underscores the effectiveness of the tuning process, emphasizing the model's ability to accurately classify both positive and negative cases in the context of cardiovascular disease detection. The overall evaluation metrics, including specificity and AUC, further contributed to an impressive average performance of 95.014%. SHAP (SHapley Additive exPlanations) analysis was performed to interpret the model's predictions, revealing key features influencing cardiovascular risk. Looking forward, future work could explore further refinement of the model through advanced optimization techniques and integration of additional clinical data. As for model comparisons, our primary objective was to evaluate TabTransformer's performance. Our future studies will include statistical comparisons with baseline models such as logistic regression, random Forests, and gradient boosting machines. To validate model improvements more thoroughly, this will be accompanied by rigorous regression evaluation metrics, including confidence intervals and statistical significance testing. Additionally, expanding the model's applicability to diverse datasets and collaborating with healthcare professionals could enhance its real-world effectiveness in early-stage cardiovascular disease detection.

Abbreviations

CVD: Cardiovascular diseases; NLP: Natural language processing; ML: Machine learning; ECG: Electrocardiogram; SHAP: SHapley Additive exPlanations; RF: Random Forest; ET: Extra Trees; XGB: Extreme gradient boosting; KNN: K nearest neighbor; LR: Logistic regression; MLP: Multi-layer perceptron; LGBM: Light gradient boosting machine; NB: Naive Bayes; DT: Decision trees; SVM: Support vector machine; CNN: Convolutional neural network; SCA: Sudden cardiac arrest; BiLSTM: Bidirectional long short-term memory; BiGRU: Bidirectional gated recurrent unit; SMOTE: Synthetic minority oversampling technique; DBSCAN: Density-based spatial clustering of applications with noise; SMOTE-ENN: SMOTE-Edited nearest neighbor; STD: Standard deviation; PCA: Principal component analysis; ROC: Receiver operating characteristic; AUC: Area under the curve; GBM: Gradient boosting machine; NN: Neural networks; LDA: Linear discriminant analysis; DCA: Decision curve analysis; STARS: Systematic assessment of rehabilitation situation; ROC: Receiver operating characteristic; SViT: (From previous paper cited Swin Vision Transformer variant); XAI: Explainable artificial intelligence.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s13755-025-00361-7.

Funding

This work was made possible by student grant# QUST-1-CENG-2024-1723 from Qatar University. The statements made herein are solely the responsibility of the authors. The open-access publication cost is covered by the Qatar National Library.

Data availability

The processed data set used in this study can be made available upon a reasonable request to the corresponding author.

Declarations

Conflict of interests

Authors have no conflict of interest to declare.

Institutional Review Board Statement

Not applicable.

Informed consent

Not applicable.

Author details

[1]Department of Electrical Engineering, Qatar University, 2713 Doha, Qatar. [2]Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka 1000, Bangladesh. [3]Department of Industrial and Mechanical Engineering, Qatar University, Doha-2713, Qatar. [4]Department of Electronics and Communication Engineering, Intelligent Signal Processing (ISP) Research

Lab, Kuwait College of Science and Technology, Block 4, 13133 Doha, Kuwait. [5]Department of Electronics and Communication Engineering, Vels Institute of Sciences, Technology, and Advanced Studies, Chennai, Tamil Nadu, India.

Received: 16 March 2024 Accepted: 7 July 2025
Published: 21 July 2025

References

1. Singh A, Kumar R. Heart disease prediction using machine learning algorithms. IEEE; 2020. pp. 452–457.
2. Duc HN, Oh H, Yoon IM, Kim M-S. Association between levels of thiamine intake, diabetes, cardiovascular diseases and depression in Korea: a national cross-sectional study. J Nutr Sci. 2021;10: e31.
3. Rahim A, Rasheed Y, Azam F, Anwar MW, Rahim MA, Muzaffar AW. An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE Access. 2021;9:106575–88.
4. Chen Z, et al. Association of triglyceride to high-density lipoprotein cholesterol ratio and incident of diabetes mellitus: a secondary retrospective analysis based on a Chinese cohort study. Lipids Health Dis. 2020;19(1):1–11.
5. Halliburton SS, Petersilka M, Schvartzman PR, Obuchowski N, White RD. Evaluation of left ventricular dysfunction using multiphasic reconstructions of coronary multi-slice computed tomography data in patients with chronic ischemic heart disease: validation against cine magnetic resonance imaging. Int J Cardiovasc Imaging. 2003;19:73–83.
6. Handa G, Hazra S, Chalageri PH. WHO systematic assessment of rehabilitation situation (STARS): a systematic review on the status of stroke rehabilitation in India. Curr Phys Med Rehabil Rep. 2023. https://doi.org/10.1007/s40141-023-00418-2.
7. Agarwal A, Balani K, Venkateswaran S. Medical Education in India; 2023.
8. Minja NW, et al. Cardiovascular diseases in Africa in the twenty-first century: gaps and priorities going forward. Front Cardiovasc Med. 2022;9:1008335.
9. Junusbekova G, Tundybayeva M, Akhtaeva N, Kosherbayeva L. Recent trends in cardiovascular disease mortality in Kazakhstan. Vasc Health Risk Manag. 2023. https://doi.org/10.2147/VHRM.S417693.
10. Rawat A, Belwal CM, Gururani K. Prevalence of adult congenital cardiac disease: report from a tertiary care center. Pak Heart J. 2023;56(3):528–36.
11. Shabana NA, Shahid SU, Irfan U. Genetic contribution to congenital heart disease (CHD). Pediatr Cardiol. 2020;41:12–23.
12. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol. 2002;39(12):1890–900.
13. Tyagi R. Empowering the unbanked: a revolution in financial inclusion through artificial intelligence. Int J Res Eng, Sci Manage. 2023;6(10):4–12.
14. Adke A, Ghorpade S, Chaudhari R, Patil S. Navigating the confluence of machine learning with deep learning: unveiling CNNs, layer configurations, activation functions, and real-world utilizations; 2023.
15. Allami RH, Yousif MG. Integrative AI-driven strategies for advancing precision medicine in infectious diseases and beyond: a novel multidisciplinary approach.2023. arXiv preprint arXiv:2307.15228.
16. Murugappan M, Murugesan L, Jerritta S, et al. Sudden cardiac arrest (SCA) prediction using ECG morphological features. Arab J Sci Eng. 2021;46:947–61.
17. Khomtchouk BB. Cardioinformatics advancements in healthcare and biotechnology. Circ: Genomic Prec Med. 2023: e004119.
18. Abbas F, Zhang F, Iqbal J, Abbas F, Alrefaei AF, Albeshr M. Assessing the dimensionality reduction of the geospatial dataset using principal component analysis (PCA) and its impact on the accuracy and performance of ensembled and non-ensembled algorithms; 2023.
19. Papoutsoglou G, et al. Machine learning approaches in microbiome research: challenges and best practices. Front Microbiol. 2023. https://doi.org/10.3389/fmicb.2023.1261889.
20. Chandrasekhar N, Peddakrishna S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes. 2023;11(4):1210.
21. Tiwari A, Chugh A, Sharma A. Ensemble framework for cardiovascular disease prediction. Comput Biol Med. 2022;146: 105624.
22. Ramu M, Harshitha M, Hanisha R, Chandana P, Meghana P. Cardiovascular disease prediction using machine learning classifiers, in 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1. IEEE; 2023. pp. 709–15.
23. Torkul O, Turgay S, Şişci M, Babacan G. Prediction of heart disease using fuzzy rough set based instance selection and machine learning algorithms. In International symposium on intelligent manufacturing and service systems. Springer; 2023. pp. 699–709.
24. Paul B, Karn B. Heart disease prediction using scaled conjugate gradient backpropagation of artificial neural network. Soft Comput. 2023;27(10):6687–702.
25. Rajdhan A, Agarwal A, Sai M, Ravi D, Ghuli P. Heart disease prediction using machine learning. Int J Eng Res Technol, 2020; 9(4).
26. Nagarajan SM, Muthukumaran V, Murugesan R, Joseph RB, Meram M, Prathik A. Innovative feature selection and classification model for heart disease prediction. J Reliab Intell Environ. 2022;8(4):333–43.
27. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst. 2020;32(11):4793–813.
28. Doppala BP, Bhattacharyya D, Janarthanan M, Baik N. A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. J Healthc Eng. 2022. https://doi.org/10.1155/2022/2585235.
29. Chandrasekaran DKA. Heart disease prediction using machine learning and deep learning. 2022. https://doi.org/10.1109/CISES54857.2022.9844406
30. Baccouche A, Garcia-Zapirain B, Castillo Olea C, Elmaghraby A. Ensemble deep learning models for heart disease classification: a case study from Mexico. Information. 2020;11(4): 207.
31. Dubey M, Tembhurne J, Makhijani R. Heart failure prediction on diversified datasets to improve generalizability using 2-level stacking. Multidisc Sci J. 2024;6(3):2024028.
32. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–54.
33. Nagavelli U, Samanta D, Chakraborty P. Machine learning technology-based heart disease detection models. J Healthc Eng. 2022. https://doi.org/10.1155/2022/7351061.
34. Sharma V, Yadav S, Gupta M. Heart disease prediction using machine learning techniques. IEEE; 2020. pp. 177–81.
35. Moreno-Sanchez PA. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Front Cardiovasc Med. 2023. https://doi.org/10.3389/fcvm.2023.1219586.
36. Sarra RR, Dinar AM, Mohammed MA, Abdulkareem KH. Enhanced heart disease prediction based on machine learning and χ2 statistical optimal feature selection model. Designs. 2022;6(5):87.
37. Reddy NSC, Nee SS, Min LZ, Ying CX. Classification and feature selection approaches by machine learning techniques: heart disease prediction. Int J Innov Comput. 2019;9(1): 2180–4370.
38. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak. 2019;19(1):1–15.
39. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Comput Sci. 2020;1:1–6.
40. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms. 2023;16(2):88.
41. Hera SY, Amjad M, Saba MK. Improving heart disease prediction using multi-tier ensemble model. Netw Model Anal Health Inform Bioinform. 2022;11(1):41.
42. Ghosh P, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access. 2021;9:19304–26.
43. Chakraborty C, Kishor A, Rodrigues JJPC. Novel enhanced-grey wolf optimization hybrid machine learning technique for biomedical data computation. Comput Electr Eng. 2022;99: 107778.
44. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Ann Ibadan Postgraduate Med. 2008;6(1):21–6.
45. Chowdhury ME et al. An early warning tool for predicting mortality risk of COVID-19 patients using machine learning. Cognit Comput 2021: 1–16.
46. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng. 2014;40(1):16–28.
47. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

48. Huang X, Khetan A, Cvitkovic M, Karnin Z. Tabtransformer: tabular data modeling using contextual embeddings; 2020. arXiv preprint arXiv:2012.06678.
49. Mukhamediev RI, Symagulov A, Kuchin Y, Yakunin K, Yelis M. From classical machine learning to deep neural networks: a simplified scientometric review. Appl Sci. 2021;11(12):5541.
50. Vaswani A et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
51. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; 2019. pp. 2623–31.
52. Cui T, Wang Y, Ji S, Li Y, Hao W, Zou H, Jhanji V. Applying machine learning techniques in nomogram prediction and analysis for SMILE treatment. Am J Ophthalmol. 2020;210:71–7.
53. Murugappan M, Prakash NB, Jeya R, et al. A novel few-shot classification framework for diabetic retinopathy detection and grading. Measurement. 2022;200: 111485.
54. Kramer O, Kramer O. Scikit-learn. Mach Learn Evol Strategies 2016: 45–53.
55. Christoph M. Interpretable machine learning: A guide for making black box models explainable; 2020.
56. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.
57. Malik S, Wong ND, Franklin SS, Kamath TV, L'Italien GJ, Pio JR, Williams GR. Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. Circulation. 2004;110(10):1245–50.
58. Baghdadi NA, Farghaly Abdelaliem SM, Malki A, Gad I, Ewis A, Atlam E. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data. 2023;10(1): 144.
59. Tunstall L, Von Werra L, Wolf T. Natural language processing with transformers. O'Reilly Media Inc; 2022.
60. Chernyavskiy A, Ilvovsky D, Nakov P. Transformers: "the end of history" for natural language processing? In Machine learning and knowledge discovery in databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, September 13–17, 2021, Proceedings, Part III 21. Springer; 2021. pp. 677–693.
61. Hemalakshmi GR, Murugappan M, Sikkandar MY, et al. PE-Ynet: a novel attention-based multi-task model for pulmonary embolism detection using CT pulmonary angiography (CTPA) scan images. Phys Eng Sci Med. 2024;47:863–80.
62. Hemalakshmi GR, Murugappan M, Sikkandar MY, et al. Automated retinal disease classification using hybrid transformer model (SViT) using optical coherence tomography images. Neural Comput Appl. 2024;36:9171–88.