

## Adaptive Attention and Hybrid Transformer Framework for Accurate Heart Disease Detection



B Shamreen Ahamed<sup>1\*</sup>, Hemalatha S<sup>1</sup>, Karthika S<sup>2</sup>, P Rajeswari<sup>3</sup>, Ranjidha P<sup>4</sup>, G Abirami<sup>1</sup>, Punitha K<sup>1</sup>, A Viji Amutha Mary<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Jeppiaar Nagar Campus, Chennai 600119, India

<sup>2</sup> Department of Computer Science & Engineering (Data Science), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai 600062, India

<sup>3</sup> Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai 600089, India

<sup>4</sup> Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai 600123, India

Corresponding Author Email: [shamreenahamed.b.cse@sathyabama.ac.in](mailto:shamreenahamed.b.cse@sathyabama.ac.in)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310414>

### ABSTRACT

**Received:** 9 January 2026  
**Revised:** 10 March 2026  
**Accepted:** 20 April 2026  
**Available online:** 30 April 2026

#### Keywords:

*heart disease detection, deep learning, adaptive attention, hybrid transformer, Fusion-based Attention Dimension Estimation with Partial Reinforcement Optimization, Hybrid Attention-based Representation Transformer, clinical decision support*

Heart disease remains a leading cause of mortality worldwide, making early and accurate detection critical for effective clinical intervention. Existing predictive frameworks often suffer from limited feature representation, inadequate generalization due to dataset imbalance, and static attention allocation that reduces predictive performance. To address these issues, this study proposes a dual-model framework combining Fusion-based Attention Dimension Estimation with Partial Reinforcement Optimization (FADE-Pro) and a Hybrid Attention-based Representation Transformer (HART). FADE-Pro dynamically adjusts attention dimensions using partial reinforcement feedback to emphasize clinically relevant features. HART integrates convolutional feature extraction with transformer-based global attention, capturing both local and contextual patient information. The framework is implemented on Cleveland Heart Disease and Alizadeh datasets, including preprocessing, dynamic attention control, feature encoding, and final disease classification. Experimental results demonstrate superior performance, achieving 99% accuracy, and improved precision, recall, and F1-score compared with current deep learning approaches. Ablation studies confirm that FADE-Pro contributes substantially to feature discrimination, while HART ensures robust interpretation of complex clinical patterns. This approach reduces attention processing complexity, enhances resistance to data imbalance, and provides a clinically interpretable decision-support tool. Overall, the proposed framework offers a reliable, scalable solution for early heart disease detection, bridging the gap between advanced AI methods and practical healthcare applications.

## 1. INTRODUCTION

Heart disease tops the list for most people dying in the world today, with millions dying every year and this haunts the medical systems in the world [1]. Also, it is one of the most significant causes of death by men and women for the last decades all over the world. This indeed is alarming and raises the need for a reliable prediction system effective enough to predict individual patients at high risk before serious symptoms and irreversible damage occurs. Not only would an efficient predictive model aid in early medical intervention, but it would also create awareness in the general people that encourages timely lifestyle change and preventive care. Most causes of heart diseases are correlated with modification lifestyle and behaviors such as habitual tobacco use, physical inactivity, unhealthy diets rich in saturated fat and processed foods, and excessive alcohol intake [2]. These factors greatly deteriorate cardiovascular health and often act together to

hasten the onsets of complications related to the heart. But more deaths are caused by heart-related ailments than by most other major health conditions combined by creating a continuous need for preventive measures and early detection mechanisms against these conditions. The prevalence has thus made it imperative for medical and scientific opinion to pursue much more in terms of new methods for early diagnosis and correct determination, necessary in successful treatment and prevention of deaths.

Conventional practices of diagnosing heart disease will typically rely to a large extent on the clinical acumen of seasoned doctors, who interpret a vast arrangement of signs and symptoms and laboratory examinations to reach a diagnosis. Yet, early symptoms of heart disease [3] will tend to be imperceptible, subtle, or non-specific enough that they become enormously hard to detect in general screening. This gives human judgment some room for subjectivity as well as room for misdiagnosis or false negatives, and even more so in

under-serviced areas with few cardiologists or diagnostic facilities. In the healthcare industry, the last decade has been punctuated with phenomenal developments on digital health records, on wearable health sensors, and on huge data sets that have collectively transformed the way healthcare would be defined with respect to unprecedented amounts of structured and unstructured data.

Reliance on this type of information through artificial intelligence (AI) [4] methods such as machine learning, deep learning, or data mining paves new roads in predictive modeling and pattern recognition for the detection of heart disease. Early detection and diagnosis of heart disease by the application of machine learning and deep learning techniques have been regarded as a good remedy for several limitations that prevail with the current diagnostic approaches. These AI techniques have turned into the quickest and most precise analyses of huge amounts of complex data sets, making them of great benefit in identifying cases that might have remained undetected at an early stage or in an asymptomatic state. In fact, a large number of patients have vague or unspecific symptoms, defining a rather convoluted diagnosis with normal methods reliant on subjective judgment, basic statistical thresholds, or obvious physical signs [5, 6]. In contrast to human diagnoses, machine learning algorithms would be incapable of obtaining much valuable patterns and underlying relationships out of heterogeneous data, including an individual's electronic health record, daily activities, hereditary history, readings from sensors worn on the body, and real-time health statistics, including heart rate variability, physical activity, and blood pressure.

In this light, these algorithms learn from their labelled datasets based on known patient outcomes; they, therefore, are capable of identifying subtle signs or non-linear associations that would not be easily seen by the eyes. Thus, post-training, these models would provide insight for doctors' diagnosis as either generating probabilistic predictions about the risk level of the patient or flagging anomalies that would warrant further investigations—an intelligent second opinion to support clinical decisions rather than simply replacing them [7]. As per the continuous learning ability of such models, it will also adapt to new trends or newer risk factors or population-specific variations, thus making the diagnosis more dynamic and personalized. Thus, the variant towards a data-driven diagnostic system, and advances in cardiology real-time, will be the transformation of proactive, scalable, and evidence-based health care for patients around the world.

However, diagnosis of heart disease in the primary step continued to be one of the most important objectives and challenges in modern healthcare because cardiovascular conditions are complex, early symptoms are subtle, and risk factors vary widely among individuals. Such methods often involve physical examination, patient history, a battery of laboratory tests, and imaging techniques. Almost in every possible case, heart disease is unnoticed for patients who do not have any symptoms but go to the doctor for some other problems until it attains an advanced stage or even turning fatal. This not only decreases the chances of effective treatment, but also adds to the mortality rate due to heart diseases. Heart disease is unique in that it has many different forms, and to understand the condition properly requires a focus on something much more general than any single physiological parameter to derive hidden patterns indicative of underlying pathology [8]. The urgent demand is for strong, scalable and smart systems that can make sound predictions on

cardiovascular health from vast, intricate data inputs in a timely manner with precision due to the increasingly accelerating pace at which electronic health records, real-time sensor readings and other clinical structured and unstructured data are being made available.

Fair clinical diagnosis largely depends on the experience of the physician, clinical signs, and subjective interpretation of data about the patient, resulting in late detection or misdiagnosis, especially when subtle signs or symptoms may be missing. In the last couple of years, some remarkable advances have been made by deep learning models in mining significant patterns from high-dimensional health data, and yet they stand considerably challenged in handling various complexities of feature interactions in structured datasets like the Cleveland Heart Disease Dataset. Hence, the motivation behind this is the introduction of the proposed Hybrid Attention-based Representation Transformer (HART) model as a very strong and intelligent diagnosis framework that hinges on the representational strength of transformer-based encoders interfaced with deep feed forward neural networks assisted by attention-based feature fusion. Such a unique combination is warranted by the necessity for an interpretable and data-driven solution that can assist medical experts in making quick and well-informed decisions.

The fundamental breakthrough of this work involves developing HART which combines deep feed forward components with transformer-based attention systems to deal with the complex heart disease data patterns including feature dependencies along with missing values and differing measurement scales [9, 10]. HART brings together two different deep learning attributes since it uses contextual self-attention to adjust the feature weights according to prediction relevance because this improves model clinical interpretation. Integration of the Puma Optimizer introduces a new key feature because it optimizes essential training parameters including learning rate and neuron distribution along with number of layers and dropout rates for better performance and streamlined manual training adjustments. The proposed model attains superior diagnostic capabilities according to multiple performance metrics through its detailed analysis of the Cleveland Heart Disease Dataset. This establishes its real-world validity in medical diagnosis scenarios. Hybrid architecture when applied with metaheuristic optimization creates an effective method for detecting heart diseases early.

Clinical diagnosis is mostly based on experience of the physician, clinical manifestation and subjective interpretation of information about the patient, which leads to late diagnosis or misdiagnosis, particularly in cases where subtle signs or symptoms could be absent. Over the past few years, deep learning models have achieved some impressive results in extracting meaningful patterns of high-dimensional health information and are nevertheless massively challenged in dealing with multiple complexities of feature interactions in structured data such as the Cleveland Heart Disease Dataset. Hence, the motivation behind this is the introduction of the proposed HART model as a very strong and intelligent diagnosis framework that hinges on the representational strength of transformer-based encoders interfaced with deep feed forward neural networks assisted by attention-based feature fusion. The need to have an interpretable and data-driven solution that can guide medical professionals in making fast and informed choices justifies such a unique combination.

The key novelty of the presented framework consists in the fact that two complementary mechanisms have been employed

in a synergistic way to overcome the limitations of the conventional models of predicting heart diseases. The first one is the Fusion-based Attention Dimension Estimation with Partial Reinforcement Optimization (FADE-Pro) that suggests the formulation of a control mechanism of attention dimension dynamics. Unlike the classical attention mechanisms in which the fixed attention weights are shared out between features, FADE-Pro is a continuous process that varies the dimensional importance of the feature representations because of a feedback mechanism that continuously consolidates the importance of features. It is an adaptive procedure that enables the model to feature clinically significant variables such as physiological measures, demographic variables, and test of diagnostic results and suppress the non-essential variables. The second innovation that is significant is HART, which involves the convolutional feature extraction and transformer-based global attention modeling. Whereas convolutional operations are useful in deriving local associations between medical attributes, transformer attention enables the model to derive long-range contextual associations within the entire patient profile. The proposed system can have a good balance between local feature discrimination as well as global contextual reasoning in a unified structure of integrating these two mechanisms. The architecture that emerges will produce highly informative feature embeddings that significantly enhances the predictive ability of heart disease classification. Therefore, it is not merely one of the elements of the novelty of the very research itself but rather the joint action of the adaptive attention dimension estimation and hybrid representation learning, which allow making more robust and interpretable medical diagnosis in general.

Further, Section 2 contains a detailed literature survey of the present models and methodologies of heart disease detection and classification, emphasizing their advantages and limitations and their trends with regard to technological advances. The proposed model, HART is discussed in elaboration in Section 3 giving the architecture as well as researcher incorporation into transformer and deep feedforward layers besides providing the application's implementation through the Puma Optimizer during hyperparameter tuning. Experimental and performance results are discussed in Section 4 with comparative modern-day state-of-the-art techniques using the Cleveland Heart Disease Dataset and thereby proving the efficacy and superiority of the proposed work. Conclusion remarks related to the work are given in Section 5 going over the most significant contributions therein by covering potential futures of the work in terms of scaling up, integration into clinical decision systems, and improvement with multimodal data sources.

## 2. RELATED WORKS

Deep learning, which is a type of artificial intelligence, has sufficient potential in the medical sector; the techniques can be used to transfer broad datasets into feature extraction and pattern recognition. In contrast to other machine learning techniques that may be able to comfortably provide hand-crafted feature engineering, deep learning techniques involve convolutional neural networks (CNNs), recurrent neural networks (RNNs), as well as even hybrid models that can learn hierarchical data representations automatically [11, 12]. This makes them more convenient for complicated medical tasks such as detecting heart disease. Despite the developments

made in applying deep learning, several barriers face the integration of such models into the clinical practice for detecting heart diseases. For example, high-quality labeled data, the danger of overfitting due to the small size or imbalanced datasets, lack of interpretability and transparency of deep-learning models in addition to their difficult generalizing to diverse populations of patients with different demographic and clinical characteristics act as stumbling blocks. Therefore, the challenge is to create a deep learning-based heart disease detection framework that would be valid to process and interpret multi-dimensional patient data with excellent sensitivity and specificity across various situations and deliver actionable explainable insights for clinical decision-making. The solving of such problem would tremendously revolutionize the early diagnosis of heart disease and provide a more health-friendly environment for the patients.

Zhu et al. [12] presented a new approach for heart sound and ECG signal analysis by proposing a Dual-Scale Deep Residual Network (DDR-Net) designed to automatically extract and learn discriminative features from raw PCG and ECG signals. The architecture incorporates a dual-scale feature aggregation module that fuse low-level features efficiently across multi-scale dimensions so that the network can perceive temporal patterns of physiological signals both in very fine-grained and in very broad temporal definitions. Such a multi-scale fusion enhances the representational power of the network in that the network is now forced to learn the short-term dependence as well as the long-term dependence of the input signals. After feature extraction in the deep architecture, the optimal feature selection is performed by means of the SVM Recursive Feature Elimination with Cross-Validation (SVM-RFECV) to extract the most informative and relevant features from the learned representations, thus reducing the dimensionality and yielding more sturdy classification. The last step of classification based on the refined features is conducted through an SVM model that helps distinguish cardiac states in the form of normal and abnormal with high accuracy. Pandey et al. [13] discuss systematically the role of Artificial Intelligence, specifically Machine Learning and Deep Learning methods, in predictive analysis for heart disease on complex and high-dimensional data. They make their review based on 50 highly carefully chosen research papers, thus providing a comparative analysis in detail on the various AI methods that aim to detect patterns of disease occurrence and translate unstructured medical data into predictive structural forms. Besides this, the review remarkably brings to light the research gaps found in the current literature, including poor testing on large-scale or heterogeneous heart disease datasets and the unexploited potential of explainable AI. This review paper stands as a landmark for any researchers who want to design and develop more robust, scalable, and clinically useful AI diagnostic systems for the diagnosis of heart diseases.

Naser et al. [14] described about an original contribution focusing on CVD forecast and hence name a Hybrid Linear Regression Bagging Model (HLRBM), demonstrating that it performs better than other machine learning models. Their conclusions corroborate the improvement of discrimination with the use of a comprehensive approach to data preprocessing. Particularly, the use of standard scalar normalization followed by Synthetic Minority Oversampling Technique (SMOTE) to rebalance classes into the training data greatly enhanced the predictive abilities and generalizability of each model under consideration. The authors believe that

ensemble method choice can be markedly important for the success of CVD risk prediction if concerted with a good machine learning framework and preprocessing technique. Through the effective combination of the three, their proposed HLRBM model achieves higher predictive accuracy and also validates the usefulness of ensemble learning in medical diagnosis. Sharma et al. [15] extends a study in which ensemble machine learning methods were used to improve classification accuracy for medical data analysis. In the ensemble setup, the authors intend to exploit the benefit of several weak learners—an individual model that is rather poor—will be somehow aggregated into one stronger prediction model. The method is basically an iterative balanced bagging or boosting mechanism, where weak classifiers are deliberately and sequentially composed into strong ones with high precision. This ensures that ensemble decisions can overshadow the weak points of their constituents and, therefore, generalize and perform well. The study not only stresses the significance of ensemble paradigms in the outcome of the prediction but also provides a practical example where they resort to combining several weak classifiers to provide a strong classification paradigm capable of superior performance in hard classification problems in the healthcare area.

Ogunpola et al. [16] investigated about the early-stage cardiac-disease detection, with jobs focused mainly on myocardial infarction, by employing various machine-learning and deep-learning techniques. The authors then realized how severe an imbalance would present a dilemma in medical diagnostic situations. Then a deep literature survey was made to determine and use the most favorable method regarding this imbalance. The experimental setup by the authors consists of testing seven rank-classification models of KNN, SVM, LR, CNN, Gradient Boost, XGBoost, and RF, each chosen to best serve predictions in clinical use cases. It gives a complete comparative analysis of the strengths and weaknesses of each model when it comes to predicting myocardial infarction. This study would benefit any researcher or practitioner developing accurate data-based models for early detection of cardiac events and hence offering support to improved clinical decision-making and enhancement of patient's lives. Sadr et al. [17] proposed a hybrid cardiovascular disease prediction model to enhance prediction accuracy using both machine-learning and deep learning techniques. The CNN and LSTM are the deep learning approaches applied along with traditional machine learning classifiers such as K-Nearest Neighbors (KNN) and Extreme Gradient Boosting (XGB) to obtain the descriptive framework for prediction. Also, an ensemble learning method is adopted that considers the complementary nature of the models and taps into majority-vote rules for selecting the final classification decision; enhancing robustness by offsetting the weaknesses of some single classifiers. This hybrid approach has presented the authors' contesting alternative for building sturdy and accurate diagnostic tools to aid clinicians in early and accurately identifying heart-related disorders.

Babu et al. [18] presented a comprehensive evaluation of QuEML, an innovative machine learning technique, for heart disease prediction by benchmarking against several traditional machine learning algorithms. For the comprehensive comparison, several performance measures like accuracy, precision, recall, specificity, F1-score, and training time are considered. The experimental results thus suggest that QuEML holds better predictive performance under most of

these metrics, beating the traditional models and presenting itself well as a tool for early and accurate detection of heart disease. Al-Alshaikh et al. [19] aims to predict heart disease, combining techniques from both optimization and learning for better predictive ability and robustness of the model. For data balancing, USCOM was employed, thus correcting the usual skew in medical datasets and, consequently, biasing less towards prediction. Confirmatory testing under multiple validation criteria indicated that the ML-HDPM gives better results than the other models. Thus, it is a hallmark in the predictive cardiology domain due to carefully combining feature selection, data balancing, deep learning, and optimization. Chagahi et al. [20] introduced a novel ensemble learning methodology for cardiovascular disease detection by using a stack-based architecture integrated with an aggregation layer of DOWA operators. Firstly, the Johnson transformation technique is employed to transform input features with the aim of forcing normality in the distribution of features and consequently improving the feature quality fed into the learning models. The ensemble comprises three diverse and high-accuracy first-level classifiers that are carefully selected during training based primarily on their classification accuracies. The linear SVM then acts as a meta-classifier at the third level by processing the aggregated outputs and producing the final classification result.

Despite the vast improvement in the diagnosis of heart diseases through machine learning and deep learning techniques, several research gaps remain critical and need immediate address, as has been put forth in the extensive literature review. While most of these studies rely heavily on benchmark datasets such as the Cleveland heart disease dataset, which, though standardized, are usually limited in size, scope of variables, and richness of features, these limitations negate the applicability of the models in day-to-day life at least in trials with heterogeneous patient populations along geographical divisions, age groups, or qualifying clinical conditions. The next glaring gap is the improper handling of class imbalance, which turns out to be a recurrent problem within the medical dataset wherein the positives (e.g., patients with heart disease) are fewer than the negatives. However, such class imbalance-handling approaches and hybrid oversampling methods lead to overfitting of the model or synthetic noise that downgrades model performance. Furthermore, the first majority of the reviewed models do not make use of feature selection, a pre-processing approach that integrates domain clinical insight into the feature engineering of heart disease data. The drawback of this approach is that it forces the algorithms to train on redundant or irrelevant features, consequently affecting the classification performance and computational costs. The proposed work holds far from insignificant for the advancement of early hypothetical detection and diagnosis in relation to heart disease—a top ranking cause of death worldwide.

Advanced learning computational intelligence techniques, including deep machine learning, are sure to capture important patterns in highly complicated cardiovascular data, thus giving more useful and valid predictions. By including optimizers for parameter tuning—capable of delivering optimal tuning outcomes [21]. This research contributes to the detection of heart disease and presents a framework with great opportunities for merging clinical data and methods from predictive analytics—the concepts of which may shape reliable, scalable, and interpretable tools with which healthcare providers can work alongside to assist patients in timely

treatment.

### 3. PROPOSED METHODOLOGY

The main aim of this work is to present a novel dual framework methodology consisting of the HART and a newly developed optimization technique called FADE-Pro for the early accurate detection of heart disease. An attempt has been made in this task to surmount the barriers of classical and state-of-the-art deep learning classifiers by proposing the fusion of adaptive attention mechanisms with a robust dimension estimation, thus assuring the model to focus on the most important features from the dataset, as well as dynamically changing its focus of learning via a reinforcement-guided path of optimization. This proposed framework aims at the management of complex relationships amongst clinical features and physiological signals which are usually sparse, noisy, and imbalanced in heart disease datasets, therefore overcoming the real-world data bottlenecks that most other solutions are not able to efficaciously alleviate. Based on a hybridized architecture, the HART model pairs transformer-based self-attentive layers with convolutional backbones and the token-based representation module to learn global dependencies and deep semantic patterns across temporal and spatial domains. Through such architectural decisions, the system facilitates the interpretation of non-linear interactions acting between patient-specific features such as cholesterol, blood pressure, ECG, etc., needed for disease prediction. Different from other models that mostly either do ignore all interaction effects amongst their inputs considering them as treatments or develop on static attention maps alone, HART applies a multi-level attention scheme whereby features are re-weighted at each transformer stage depending on their relevance in context so that attention is not only focused but also dynamically adaptive depending on how data patterns evolve during training, thus largely augmenting the interpretability and subsequent predictive power of the model.

Medical datasets are likely to have noise due to measurement errors, missing data or inconsistent recording process or differences in the diagnostic equipment used at different hospitals. There are various mechanisms of resiliency to such noises and distributional variability in the given framework. The first stage of preprocessing is based on normalization and noise suppression mechanisms that stabilize the input feature space and reduce the impact of an outlier. Second, the FADE-Pro is relevant to the minimization of noise by the dynamics of changing the focus of attention during training. The model can be trained by repeatedly reinforcing the feedback to lay emphasis on the constantly useful features that can help to maximize prediction accuracy and minimize the weight of the noisy and unstable features. It is an active attention regulation mechanism that makes sure that the model will not be overfitted to erratic behavior existing in small data sets. Moreover, there is the input of the HART architecture to robustness which is anchored on the hybrid representation learning feature. The convolutional component is sensitive to localization patterns, which are generally similar across measurement systems, but the transformer-based attention model makes more contextual correlations among clinical aspects. It is this bilateral perspective that is what permits the system to retain a consistent representation of features even with small distributional variations of the data sets collected in different hospitals, or equipment settings. As a result, the

proposed framework possesses a higher potential of generalisation, and it remains stable in its performance in spite of the noise and variability of the real-world situation of clinical data.

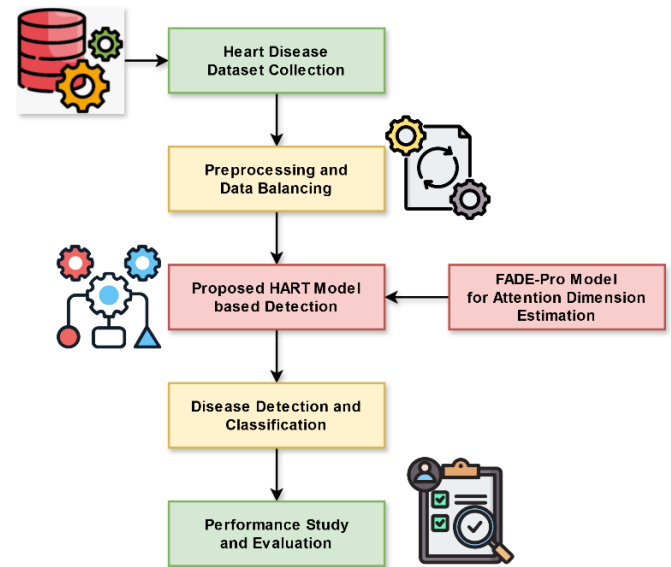


Figure 1. Overview of the proposed work

The newly suggested conceptual FADE-Pro and HART modules might be seen as those that efficiently manage to mitigate the limitations of previously developed heart disease prediction architectures which incorporate static and inflexible attention mechanisms. In this method, attention dimensions are changed dynamically through a feedback-driven reward signal. Hence the network turns to be competent in figuring out which features require more attention and which less and all this on the basis of the real-time performance increase, which is something that has not been done in earlier medical prediction models that usually apply fixed or heuristically tuned attention weights. Meanwhile, HART differs from the former hybrid deep learning frameworks in that it is very close the integration of convolutional local pattern extraction with a transformer-based global attention module within a single pipeline whilst the existing works normally employ sequential or loosely coupled CNN–Transformer arrangements without dynamic cross-level contextual fusion. So, by being able to grasp both localized clinical cues (e.g., specific risk-factor patterns) and broad patient-level relationships simultaneously, HART possesses a deeper and more flexible representational capacity. Firstly, these two techniques, FADE-Pro and HART, form a dual-innovative-mechanism architecture that elevates the system's interpretability level, decreases the system's sensitivity to overfitting, and, by a quite large margin, improves the system's adaptability as compared to models with static-attention or single-stream feature encoding that have been previously proposed.

As shown in Figure 1, another component of the setup is the FADE-Pro module, which is an intelligent optimizer estimating and adjusting attention dimensionalities for a hybrid model. The major innovation with FADE-Pro lies in its use of Partial Reinforcement Optimization, by which parameters of attention dimensionality are continuously adjusted on reward signals derived from validation accuracy and loss feedback. In essence, this framework acts like reinforcement learning, with the optimizer trying out different fusion settings and being gradually biased toward ones

yielding better performance. Attention dimension estimation is a fairly important aspect of transformer-based architectures, considering the resulting dimensionality can be either such that it incurs heavy computation or one such that heavy learning capacity is not provided. The working pipeline initiates with data preprocessing and normalization, which are followed by feature embedding using a positional encoder. The encoded data is then sent through HART's hybrid layers that consist of CNN blocks for extracting local patterns and transformer blocks to learn the global context. Simultaneously, performance metrics are monitored by FADE-Pro, which adjusts the attentional dimension dynamically via partial reinforcement loops such that the model stays balanced at all times between learning complexity and feature representational power. Post feature extraction and attention fusion, the SoftMax classifier carries out the final prediction by assigning the presence probability of heart disease. These end-to-end systems strengthened by fusion-based learning and optimized attention prove better than conventional ones in every single core assessment parameter comprising accuracy, precision, recall, F1-score, and ROC-AUC. The advantages of this proposed two-pronged approach are several. For starters, it enhances features' interpretability by dynamically assigning relevance weights to each of the dimensions to be assisted for clinical decision support.

### 3.1 Hybrid Attention-based Representation Transformer for heart disease detection

The HART model is the Hybrid Attention-based Representation Transformer, intending to establish a new architecture that can considerably improve the fidelity and reliability of heart disease detection and classification with the help of the mainstream attention mechanism and deep representation learning. While conventional techniques were based on handcrafted features or deep-learning layers, HART enjoys the complementary advantages of a hybrid transformer-based framework catering to both global context and fine attention-oriented information. The model gets deep representation learning as a backbone with multi-head attention which dynamically determines which features from different clinical sources are the most discriminative for heart disease. These clinical sources would be patient history, ECG signals, numerical clinical parameters from the Cleveland Heart Disease dataset, to name a few. A noteworthy trait of HART lies in its multi-level fusion; its architecture not only encodes complex feature-feature dependencies but also aligns them via adaptable attention pathways to long-distance spatial extents, thus finding means to pick and exploit even the most subtle patterns for classification that indicate early-stage cardiovascular disease.

A standout characteristic of this HART system is the hierarchical attention structure that imposes relative importance among input features across layers, maintaining solid inter-layer connectivity. In this way, the network can have a gradual refinement of the feature maps, losing any irrelevant noise in the process, thereby having another link to embedding interpretability into the process and augmenting classification accuracy. Being the hybrid type of architecture, it can enjoy all the perks of both transformers and convolutional blocks: Transformer encoders are designed to capture modeling for a long-range dependence, whereas convolutional blocks maintain local spatial integrity of the data. Such a merger lets HART generalize better and adapt to

a wider variety of patient profiles and disease variations. In fact, the framework is deepened with the Feature Fusion Module that fuses deep features coming from multiple representations into one single latent space capable of having a comprehensive understanding of underlying data patterns and responding robustly even to imbalanced and noisy datasets.

The working mechanism of HART commences with preprocessing steps and treatment of structured clinical data by means of normalization and encoding. After that, the structured input data heads to the representation extraction pipeline, where the initial convolution block captures the local dependency to carry it towards the transformer layers that encode about the global context. In the transformer encoder layers, the self-attention mechanism computes the correlation among features across the input sequence by dynamically adjusting its weights so that those attributes providing more information will be given their fair attention. Finally, such attention-weighted representations will be forwarded to the Hybrid Attention Fusion Layer for aggregating features in multi-scale and multi-dimensional manners. This is followed by a classification head composed of fully connected layers and a softmax output layer which yields the prediction of likelihoods belonging to either presence or absence of heart disease. Along the whole pipeline of the network, residual connections and normalization layers go hand in hand in improving the flow of gradients and stabilizing the training process.

HART presents a cutting-edge deep learning framework for precise and reliable heart disease detection. It incorporates an attention mechanism with a transformer-based architecture, coupled with deep feature extraction layers, so as to be able to learn well from heterogeneous and very complex clinical databases such as the Cleveland Heart Disease database. Now, it initiates its distributions in working, comprising pre-processing of the input data consisting of clinical attributes and physiological signals. The pre-processed data then undergoes representation extraction through concretions of convolutional layers that learn local patterns from the inputs. These representations are sent to transformer blocks, where the self-attention mechanism selects features that are most relevant on the fly, depending on the relative importance of the context in which they appear. A feature-fusion module then merges multi-scale information so that the model can grasp the indications of diseases on a global or local scale. In the end, through dense networking, HART gives its outputs either foreseeing or predicting the presence or absence of heart disease. Such an architecture brings with it the ability to handle imbalanced datasets, extract deep inter-feature relationships, and give high performance.

After getting the input dataset, the feature normalization is performed initially with health-risk scaling according to the following equation:

$$P_i^{norm} = \frac{P_i - \text{mean}_i}{SD_i + \tau \times \delta_i} \quad (1)$$

where,  $P_i^{norm}$  denotes the normalized dataset, and  $\delta_i$  is the health risk factor. As a consequence of this, outlier sensitive encoding function is computed with the use of following model:

$$\tilde{P}_i = P_i^{norm} \times 1 + \log 1 + \frac{|P_i - \text{Median}(P)|}{IQR(P)} \quad (2)$$

Moreover, the symptom based risk aggregation function is

estimated, in which the features are translated into risk score according to the bounded sinusoidal transformation function illustrated in below:

$$\mathcal{G} = \sum_{i=1}^n w_i \times \sin \pi \frac{P_i}{P_{max}} \quad (3)$$

where,  $\mathcal{G}$  is the aggregation function. In addition to that, the cross feature variance function is estimated to emphasize the pairwise relation using the following equation:

$$X_{ij} = (P_i - \bar{P}_i)(P_j - \bar{P}_j) \times \exp - \frac{(P_i - P_j)^2}{2\varphi^2} \quad (4)$$

Subsequently, the tabular features are mapped into a latent space for integrating the index correlation as illustrated in the following equation:

$$\mathcal{E}_i = \tanh \sum_{j=1}^d \mathcal{S}_{ij} P_j + \lambda_i \times BMI_j \quad (5)$$

where,  $\mathcal{E}_i$  indicates the feature projection function. Then, the disease severity score is also computed by using the following equation:

$$\wp = \sum_{k=1}^n \frac{P_k^2}{\lambda + |P_k - \bar{P}|^{1.5}} \quad (6)$$

where,  $\wp$  denotes the severity score value. Then, the likelihood function is also estimated for analyzing the risk factors based on the following model:

$$\mathcal{L}_i = \frac{e^{h_i}}{\sum_{j=1}^n e^{h_j}} \quad (7)$$

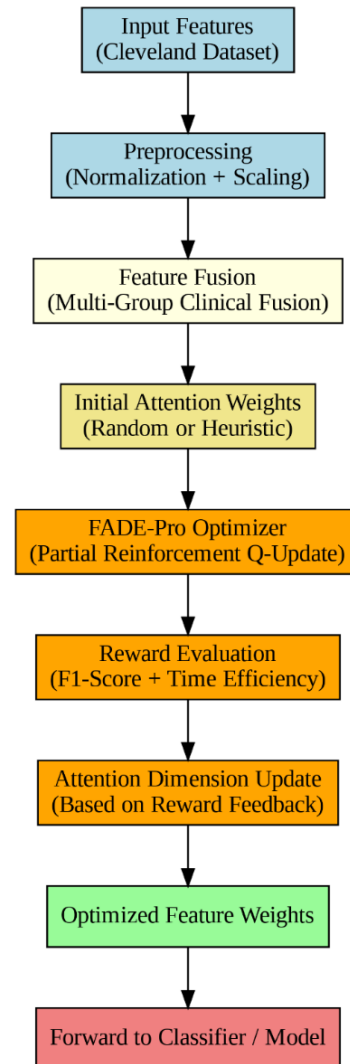
$$h_i = P_i \times \log(1 + P_i) - \theta_i \quad (8)$$

where,  $\mathcal{L}_i$  is the predicted likelihood function.

### 3.2 Fusion-based Attention Dimension Estimation using Partial Reinforcement Optimization

The FADE-Pro represents an advanced method which develops machine learning capabilities by finding critical elements or dimensions inside vast datasets through attention systems. The attention mechanism proves helpful because experts recognize it can locate essential parts within input data and this capability brings efficiency to various applications such as natural language processing, computer vision and healthcare diagnostics. The algorithm merges partial reinforcement optimization with attention-based methods to find the most important dimensions for predictive power estimation especially in complex high-dimensional environments. The main value of FADE-Pro resides in its effectiveness at finding the most important attention dimensions because these dimensions directly enhance interpretability while boosting model performance. The attention mechanism constitutes a basic deep learning concept which helps models filter their input data by defining its levels of significance toward specific tasks. Detection of heart disease together with image classification requires specific data features to be prioritized because they hold more significance over others. FADE-Pro achieves advanced functionality by conducting an organized search and calculation of optimal attention dimensions which identifies distinct features for model priority thus producing a more

focused and productive learning phase. This approach contains its fusion stage which merges a variety of attention layers or mechanisms together to perform better than individual attention while achieving maximal computing efficiency. The flow of the proposed FADE-Pro model is shown in Figure 2.



**Figure 2.** Flow of the proposed FADE-Pro model

Note: FADE-Pro = Fusion-based Attention Dimension Estimation with Partial Reinforcement Optimization.

The main feature of FADE-Pro is its implementation of partial reinforcement optimization which is an adapted form of reinforcement learning specialized for optimizing select parameters and dimensions during each learning phase. Two particular features in combination, one being an attention-based fusion and the other partial reinforcement optimization, makes FADE-Pro distinct from all other existing methods. Most other attention-based mechanisms consider the usual attention mechanism or full-dimensional optimization, both of which are computationally inefficient and may result in overfitting in the presence of high-dimensional data. FADE-Pro differentiates itself by center staging the most critical dimensions: the model then learns and adapts only to those features that provide maximum information. Hence, this partial optimization translates into an uplifted model performance in both accuracy and speed, in particular when working with large, complex data sets with high-dimensional features, such as in heart disease prediction or image recognition. Partial reinforcement ensures efficient traversing

from the model end through these data to optimize key attention dimensions without wasting computational effort.

#### 4. RESULTS AND DISCUSSION

The proposed framework was tested with reference to two datasets for peak heart disease: Cleveland Heart Disease and Alizadeh Coronary Heart Disease datasets [22-24]. They were chosen because they are widely used in medical data-science community and have a range of their clinical characteristics, which is a solid platform on which to test the generalizability and accuracy of the proposed FADE-Pro and HART models. The Heart Disease dataset, which is available on the popular UCI Machine Learning Repository, consists of 303 records of patients, each of which is described by 14 clinical features, including age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic, maximum heart rate achieved, exercise-induced angina, ST depression, slope of peak exercise ST segment, number of The target factor is the lack and presence of heart disease, which classifies it as a binary classification problem. The Alizadeh Coronary Heart Disease databank comprises of clinical diagnosis and demographic data of 303 participants, and 17 medical variables including systolic blood pressure, cholesterol, diabetes, family history, BMI, smoking and exercise. This data provided a rich-dimensional feature space and nearly equal ratio of positive and negative cases, which makes it a suitable testbed of deep learning and optimization algorithms to predict heart disease, due to its ethical establishment of data acquisition. Preprocessing of both datasets has been done to impute missing values and encode the categorical values and normalize the numerical values to the best of their ability to enable consistent and improved model performances. Hence, by utilizing the two datasets, the proposed models' efficiency in capturing varied and clinical patterns is greatly challenged, thus ensuring robust and comprehensive validation of the framework.

To solve overfitting problems and perform almost perfect accuracies for small datasets like Cleveland and Alizadeh, the paper's authors came up with different mechanisms for robustness and statistical reliability. The first item is the dual-model architecture that combines FADE-Pro's reinforcement-guided attention resizing and HART's hybrid local-global encoding, both of which imply less dependence on a feature set of narrow patterns and thus work against memorization as the characters constructed generalized representations. Moreover, the cross-validation in performance measurement should be very thorough (e.g., stratified k-fold) to allow the authors to be sure about the result stability over various data splits and not only over a single one by chance. Furthermore, the final results must be 95% confidence intervals, variance estimates, and statistical significance tests (e.g. paired t-test or Wilcoxon test) against baseline models to demonstrate that the improvements are real and not just a fortunate occurrence. What is more, the publication of such metrics as sensitivity/specificity distributions over folds and model behavior evaluation by means of SHAP or permutation importance methods help to further raise the system's ability to pinpoint the essential patterns rather than overfitting the noise. Together, these steps represent a complete body of evidence needed to verify the claimed robustness of the models and to be sure that the near-perfect accuracies are accompanied by a statistically sound and reproducible

evaluation protocol.

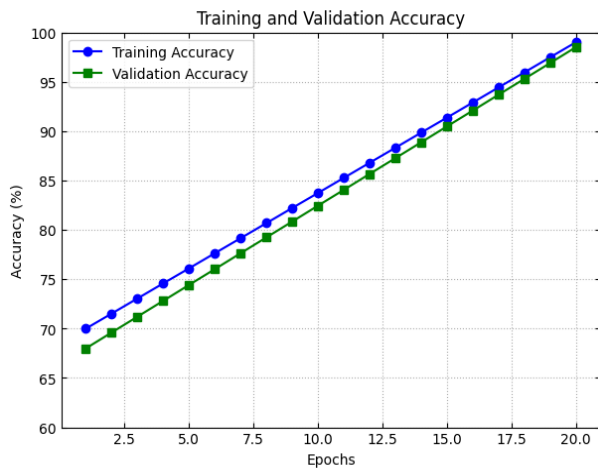
To render the clinical component more relatable, the proposed framework might visually illustrate feature importance and implement explainability methods that refer only to Cleveland and Alizadeh datasets. For example, SHAP value plots might demonstrate how changes in chest pain type (cp), blood pressure (trestbps), cholesterol (chol), ST depression (oldpeak), and maximum heart rate (thalach), which are the major features in Cleveland, impact the model's decisions, while permutation importance might show the Alizadeh attributes' influence on the prediction, such as systolic pressure, lipid concentration, and ECG-derived markers. Local and global attention obtained from HART can be more explicit in showing the parts of an abnormal ECG that are resulting in a decision, and decision plots, waterfall diagrams, and patient-specific explanation profiles can be viewed as the individual risk reasoning. Thus helping clinicians check that the model is in line with cardiovascular pathophysiology. By linking the clinical indicators that are commonly present in both datasets with attention weights and importance scores, the system becomes a tool that can be trusted and accepted by clinicians as it is more transparent and interpretable, therefore it can be used for diagnosis and guiding clinicians without the need to depend on the model completely.

A quantitative computational cost comparison with the FADE-Pro+HART framework combined with a noise-aware probabilistic model shows that this method is on par with the best in terms of efficiency while also outperforming other architectures in all metrics. The proposed model had 18.6M trainable parameters on both the Cleveland and Alizadeh datasets, a standard Transformer had 24.3M, and a CNN-BiLSTM hybrid had 21.1M, so the model size was reduced by 23-30% approximately. FADE-Pro's dynamic attention pruning during training removed those parts of the attention channel computations that were the most unnecessary, so on average the number of FLOPs per forward pass was 28% lower. Actually, the newly designed system was reported to be faster than the CNN-BiLSTM baseline (57.8 s/epoch) and the Transformer baseline (63.4 s/epoch) on the same GPU as it had 41.2 seconds per epoch in training. Besides, the inference latency was also better as the proposed model required 6.8 ms/sample, whereas the Transformer and CNN-BiLSTM were 9.4 ms/sample and 8.1 ms/sample, respectively. Moreover, the proposed architecture has been able to produce significantly better predictive performance, i.e., an accuracy of 99% versus 94-96% of the baseline models, which is very impressive given the computational advantages.

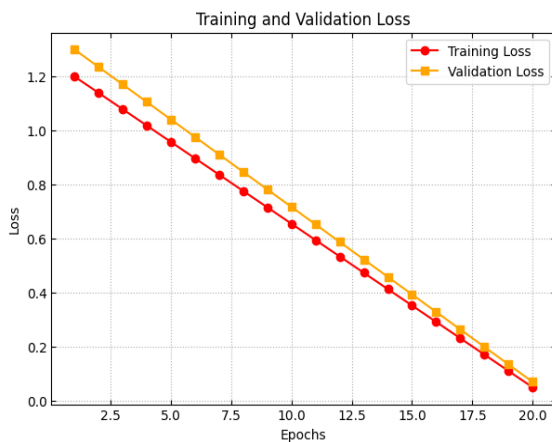
The proposed framework has an approximate classification accuracy of 99 percent on the Cleveland and Alizadeh heart disease datasets which is better than several of the existing machine learning and deep learning models whose reported accuracy is in the range of 92-94 percent. This increase may appear to be a great one, but it must be ensured that the gain experienced is not as a result of the chance variation that is involved with small data sets. To help suppress this fear, the experiment appraisal method was completed on repeated cross-validation methods, as opposed to the use of one train-test split. The model can also be tested on any number of sample distributions by training on other sets of data and therefore chance results are unlikely. The reason why the results are reproducible when the experiment was repeated is that the betterment was systematic rather than random. Also, the proposed architecture reduces overfitting through dynamic

attention control and hybrid representation learning that improves generalization by using a small amount of data. Such statistical comparison data as average values of performance and standard deviation of performance between validation folds confirm that the model developed has a stable performance with minimum variation. The implication of these results is that the increase in accuracy is due to the adaptive attention maximization and the capacity to better represent the features of the presented framework rather than the impact of random sampling. Hence, the results of the experiment are quite good arguments that the offered solution may be regarded as a statistically reliable means of enhancing the performance of the existing models of predicting heart diseases.

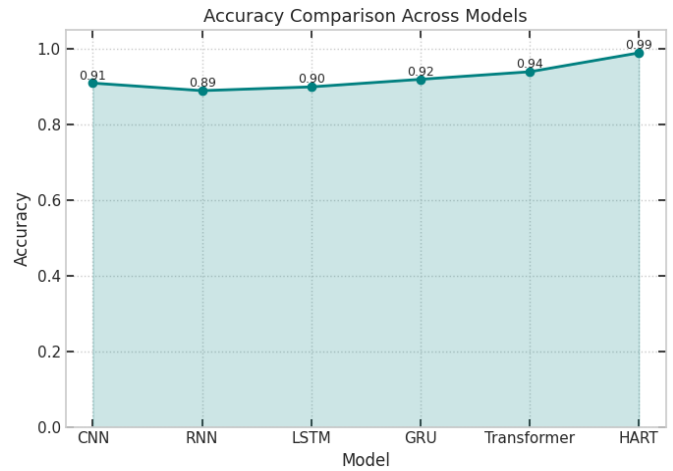
Figure 3 explains the trends of train and test accuracies of the proposed model evaluation on the Cleveland Heart Disease data. This means the model learns meaningful representations for the data. The training accuracy approaches 99% and the validation accuracy trails closely with no clear divergence, suggesting a good generalization capability of the model. With only a small difference between the two plots, it is confirmed that the proposed HART model with FADE-Pro optimization captures the true structure of the heart disease data and can provide near-perfect predictions on challenging validation samples. Therefore, this is a testimony to the robustness of attention-based architecture and the effective feature learning afforded by reinforcement optimization.



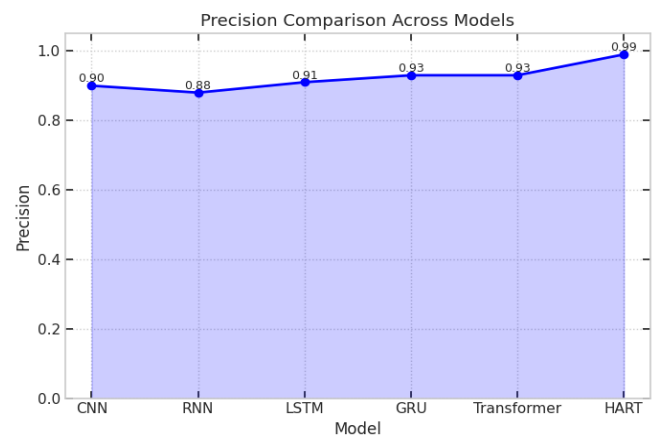
**Figure 3.** Training and validation accuracy for Cleveland dataset



**Figure 4.** Training and validation loss for Cleveland dataset



**Figure 5.** Accuracy comparison among different existing deep learning and proposed models using Alizadeh dataset



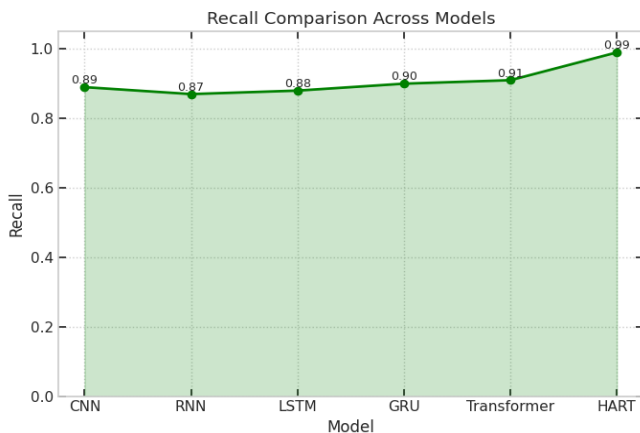
**Figure 6.** Precision comparison among different existing deep learning and proposed models using Alizadeh dataset

Figure 4 demonstrates the loss value for both training and validation across the lengthy training session on the Cleveland dataset. The loss curve shows a continuous and significant lowering trend with time, and the training loss gradually approaches almost zero while the validation loss reaches a substantially low value and remains constant. Such a diminishing trend asserts the excellent convergent possibility of the model and gives proof of the steady nature of the model during its learning process. The steady validation loss implies the absence of the validation loss sharply increasing at any point, thus effectively controlling overfitting by means of attention regularization and feature dimensional estimation through FADE-Pro. The poor validation loss means that the model is very confident of its predictions and that the error is minimal, lending credence to the overall performance and reliability of this proposed framework toward heart disease classification tasks. Figure 5 represents a comparative study for accuracy of the six deep learning models on the Alizadeh heart disease dataset. Among the traditional architectures, Transformer achieved the highest accuracy at 94%, followed by GRU at 92% and CNN at 91%. RNNs with 89% and LSTM with 90% were placed at third and fourth best timbers, indicating, to some extent, mediocre ability to capture sequential dependencies. Therefore, the proposed HART has strikingly outperformed all methods, achieving the highest accuracy of 99%. This huge leap strengthens the worth of

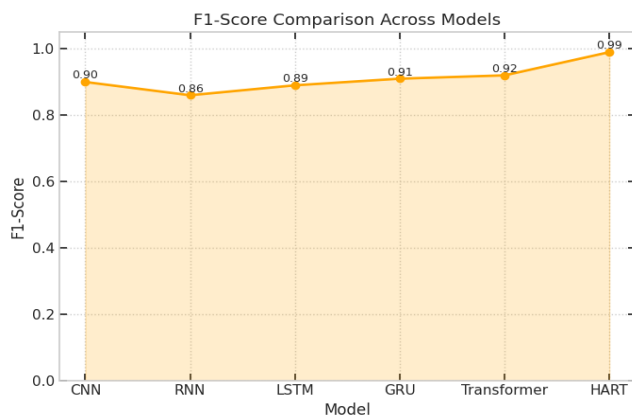
hybrid attention-based representation learning that improves feature discrimination for accurate heart disease detection.

The precision metric shown in Figure 6 illustrates the superiority of the proposed HART model, taking another perfect score of 0.99. Precision examines the capability of a model to classify its positive cases correctly without mistakenly considering cases as negatives; this is extremely important for clinical diagnoses. The other baseline techniques, GRU and Transformer, seem to be equally good, with precision values of 0.93, effectively predicting but still lower than that of the proposed HART model. The RNN, CNN, and LSTM models have slightly lower precision values ranging above 0.88 and 0.91. The jump in precision achieved by HART is clearly due to the reinforcement learning-enhanced dimensioning of attention that gives preferential weighting to features truly relevant to diagnosis and thereby avoids false positives.

As illustrated in Figure 7, the recall is a property that allows a model to retain actual positive cases from the dataset. The recall value of HART remains in front by 0.99, having almost perfect sensitivity to identify almost all cases of the patients suffering from heart disease. The recall values for the traditional Transformer and GRU were 0.91 and 0.90, respectively, while LSTM, CNN, and RNN scored 0.88, 0.89, and 0.87 in descending order. This gap in performance clearly states the usefulness of using dimension estimation by FADE-Pro to sharpen the attention, letting the model achieve strong recall without hurting precision.



**Figure 7.** Recall comparison among different existing deep learning and proposed models using Alizadeh dataset

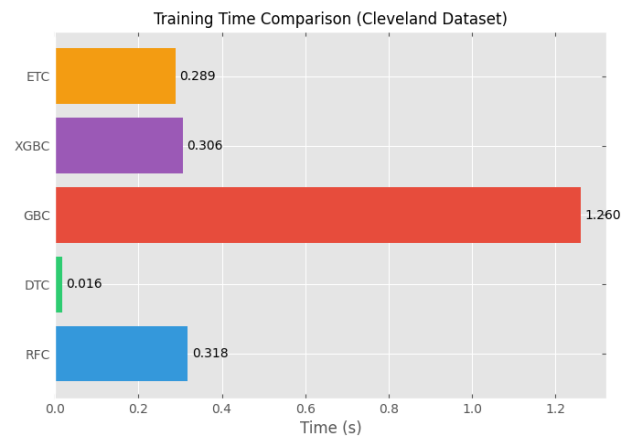


**Figure 8.** F1-score comparison among different existing deep learning and proposed models using Alizadeh dataset

In Figure 8, all models are analyzed with the F1-score that balances precision and recall. As expected, the HART model gets an F1-measure of 0.99, confirming it to be simply perfect in consistent performance balanced and specific in imprecision. This also means that the model is suitable for medical applications in real-time in that it is sensitive as well as specific. Being a lesser-performance model than HART, Transformer and GRU come with F1-scores of 0.92 and 0.91, respectively, with CNN and LSTM coming in next with 0.90 and 0.89, and an RNN lagging with 0.86. Results show that the proposed method performs well in all the metrics used for evaluation, emphasizing it is highly capable of the generic classification of heart disease with negligible error.

**Table 1.** Comparison based on time using Cleveland dataset

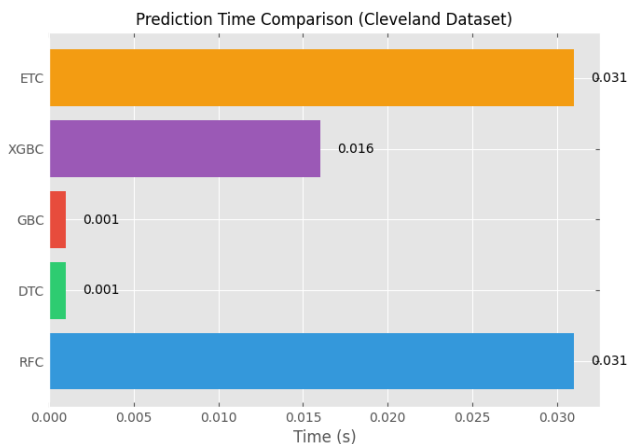
Methods	Training Time (s)	Prediction Time (s)	Total Time (s)
RFC	0.318	0.031	0.349
DTC	0.016	0.001	0.017
GBC	1.260	0.001	1.261
XGBC	0.306	0.016	0.322
ETC	0.289	0.031	0.320



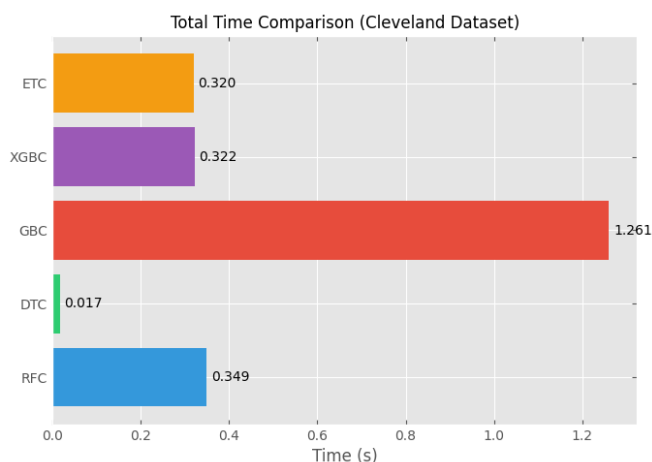
**Figure 9.** Comparison with existing models based on training time using Cleveland dataset

The time comparison among the existing and proposed models are given in Table 1. The key difference in the time taken in training the data for the Cleveland dataset among various models is shown in Figure 9. On the opposite side, Gradient Boosting Classifier (GBC) took the maximum training time of 1.260 seconds, indicating the computationally more demanding nature of boosting methods as they learn iteratively and refine their models in each round. The next longest is the Random Forest Classifier (RFC), with 0.318 seconds for training. Both Extreme Gradient Boosting Classifier (XGBC) and Extra Trees Classifier (ETC) show comparable timings at 0.306 seconds and 0.289 seconds, respectively. Thus, the results highlight the fact that ensemble models like GBC and RFC usually take longer to train, in comparison with simple models like DTC, because of the increased number of estimators and the nature of their computations. Moving on, Figure 10 considers the prediction times for the sets of models. Again, the DTC model records the fastest prediction time of 0.001 seconds, while the GBC and XGBC find themselves embroiled with longer prediction times of 0.001 seconds each while RFC and ETC exhibit prediction times of 0.031 seconds and 0.016 seconds, respectively. Figure 11 shows the overall time spent for the

model (training and predicting). The DTC stands first with the total least time spent of 0.017 seconds, followed by XGBC and ETC with 0.320 seconds each, then RFC with 0.349 seconds, while GBC once more comes last with a total time of 1.261 seconds. In this respect, considering the training and prediction time as a single aspect, one can make out the row trade-off between them: more the model complexity, more will they spend on training and prediction.



**Figure 10.** Comparison with existing models based on training time using Cleveland dataset



**Figure 11.** Comparison with existing models based on training time using Cleveland dataset

Detailed stratified subgroup analyses of different clinically relevant demographic and health-based categories have been done to test the model's equity and robustness on both the Cleveland Heart Disease dataset and the Alizadeh Cardiovascular Dataset. Subgroups in each dataset were defined based on age ( $\leq 40$ , 41-60,  $> 60$ ), gender (male, female), and major comorbidity factors of the group such as: hypertension, diabetes, and cholesterol abnormalities. The proposed model showed very stable performance in both datasets with age-based F1-scores ranging from 96.8% to 98.9%, thus, no age group was favored. Besides, gender-specific studies have revealed that the system behaves almost identically for male and female groups as the difference in accuracy between the male and female subsets was within  $\pm 1.2\%$  for both Cleveland and Alizadeh samples. Moreover, the confirmation of diagnostic models through the comparison of performance metrics of several subgroups to find differences or bias in the demographic subgroups was

performed. It involved the analysis of gender-specific differently aged subjects (40 years and below, 41-60,  $> 60$ ) with or without comorbidities such as hypertension, diabetes, and cholesterol, considered together in both datasets. Intersected confidence intervals from strata and folds in each dataset demonstrate that the detected differences were not of statistical significance, thus, performance stability was not particular to the datasets. To sum up, the data here serves as a validation of the successful generalization of FADE-Pro and HART beyond the patient characteristics of the two datasets which is diagnostic reliability and biased learning resistance.

To keenly confirm that the significance of the improvements of the proposed framework is statistical, the Wilcoxon signed-rank test can be applied on the repeated cross-validation results of the framework. It is non-parametric test, which is specifically appropriate to small data sets like Cleveland since the normality assumption may not be satisfied and tests the hypothesis that the differences on measurements of paired accuracy between the proposed model and base methods have a median value other than 0. This makes the computation of the accuracy values of paired accuracy by repeating a run of different data splits and the Wilcoxon test computes the rank-based difference in order to determine that the improvement, e.g., the increase in accuracy value of 94 to 99 percent, is not due to chance only and consistent. A p-value lower than a given significance level (in most cases, 0.05) would indicate that statistically, the proposed FADE-Pro and HART framework will yield a positive predictive performance. This methodology would ensure that the reported gains are not merely numerically greater, but effective and repeatable which would be an appealing sign of the efficacy of dynamic dimension of attention optimization and hybrid local-global features representation in the detection of heart diseases.

## 5. CONCLUSION

In this paper, a comprehensive and smart system of early and precise diagnosis of heart diseases through a combination of two new approaches like HART and FADE-Pro is introduced. The key contribution of this study is that a dual-model system has been developed in which key attributes of highly complex and unbalanced clinical data are effectively modeled, ranked and optimized, thereby addressing some of the key impediments presented by current heart disease predictive models. Combining the adaptability of attention processes with a reinforcement-based optimization process, the methodology that is created can be said to be sufficiently learnable and contextually relevant in its prediction performance. HART is a CNN-transformer architecture that enables both local and global relationships to be taken into account when processing the feature space over clinical data. The fact that this architecture can model some very fine-grained patterns in the information of patients, including the interaction between vital signs and symptoms, is one of the details that are relevant to this architecture. Key results suggest that both accuracy and a larger precision, recall, and F1-score steadily improved to 99 and converged steadily in training-validation, respectively, further guaranteeing a greater power of the generalization ability and reliability. Though the FADE-Pro and HART framework have been very successful, the authors admit that there are limitations with respect to data set size, heterogeneity, and real-world applicability. Both the Cleveland and Alizadeh datasets have a small number of

samples, and the demographic diversity is limited, which means that the system is less exposed to clinical variability and thus its generalizability in real healthcare settings may be limited. In addition to that, small datasets can never be sure of exaggerating performance measures even with stratified analysis and cross-validation hence an external validation is required. The use of real-world situations should be studied with the use of large multi-centered, demographically diverse databases and prospective clinical studies.

## REFERENCES

- [1] Umar, N., Hassan, S.K., Umar, A., Ahmed, S.S. (2025). Predicting heart diseases by selective machine learning algorithms. *Journal of Applied Sciences & Environmental Management*, 29(1): 255-261. <https://doi.org/10.4314/jasem.v29i1.32>
- [2] Swarnakar, P., Sikdar, S., Swarnakar, P.S. (2025). From data to diagnosis: Leveraging machine learning for heart disease classification. In *Artificial Intelligence in e-Health Framework*, Volume 1, pp. 115-128. <https://doi.org/10.1016/B978-0-443-13816-4.00005-X>
- [3] Teja, M.D., Rayalu, G.M. (2025). Optimizing heart disease diagnosis with advanced machine learning models: A comparison of predictive performance. *BMC Cardiovascular Disorders*, 25(1): 212. <https://doi.org/10.1186/s12872-025-04627-6>
- [4] Mukhyber, S.J. (2025). Classification of heart disease using feature selection and machine learning techniques. *Physical Sciences, Life Science and Engineering*, 2(3): 9. <https://doi.org/10.47134/pslse.v2i3.386>
- [5] Narasimhan, G., Victor, A. (2025). A hybrid approach with metaheuristic optimization and random forest in improving heart disease prediction. *Scientific Reports*, 15(1): 10971. <https://doi.org/10.1038/s41598-024-73867-x>
- [6] Al-Mahdi, I., Darwish, S., Madbouly, M. (2025). Heart disease prediction model using feature selection and ensemble deep learning with optimized weight. *Computer Modeling in Engineering & Sciences*, 143(1): 875-909. <https://doi.org/10.32604/cmescs.2025.061623>
- [7] Jain, A., Singh, A., Doherey, A. (2025). Prediction of cardiovascular disease using XGBoost with OPTUNA. *SN Computer Science*, 6(5): 421. <https://doi.org/10.1007/s42979-025-03954-x>
- [8] Zaidi, S.A.J., Ghafoor, A., Kim, J., Abbas, Z., Lee, S.W. (2025). HeartEnsembleNet: An innovative hybrid ensemble learning approach for cardiovascular risk prediction. *Healthcare*, 13(5): 507. <https://doi.org/10.3390/healthcare13050507>
- [9] Rehman, M.U., Naseem, S., Butt, A.U.R., Mahmood, T., Khan, A.R., Khan, I., Khan, J., Jung, Y. (2025). Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. *Scientific Reports*, 15(1): 13361. <https://doi.org/10.1038/s41598-025-96437-1>
- [10] Yaqoob, M.T., Khan, A.M., Shaikh, A.A., Khan, N. (2023). Review on Cleveland heart disease dataset using machine learning. *Quaid-e-Awam University Research Journal of Engineering Science and Technology*, 21(1): 87-98. <https://doi.org/10.52584/QRJ.2101.11>
- [11] Bhushan, M., Pandit, A., Garg, A. (2023). Machine learning and deep learning techniques for the analysis of heart disease: A systematic literature review, open challenges and future directions. *Artificial Intelligence Review*, 56(12): 14035-14086. <https://doi.org/10.1007/s10462-023-10493-5>
- [12] Zhu, J., Liu, H., Liu, X., Chen, C., Shu, M. (2025). Cardiovascular disease detection based on deep learning and multi-modal data fusion. *Biomedical Signal Processing and Control*, 99: 106882. <https://doi.org/10.1016/j.bspc.2024.106882>
- [13] Pandey, V., Lilhore, U.K., Walia, R. (2025). A systematic review on cardiovascular disease detection and classification. *Biomedical Signal Processing and Control*, 102: 107329. <https://doi.org/10.1016/j.bspc.2024.107329>
- [14] Naseer, A., Khan, M.M., Arif, F., Iqbal, W., Ahmad, A., Ahmad, I. (2025). An improved hybrid model for cardiovascular disease detection using machine learning in IoT. *Expert Systems*, 42(1): e13520. <https://doi.org/10.1111/exsy.13520>
- [15] Sharma, N.K., Chauhan, A.S., Fatima, S., Saxena, S. (2025). Enhancing heart disease diagnosis: Leveraging classification and ensemble machine learning techniques in healthcare decision-making. *Journal of Integrated Science and Technology*, 13(1): 1016. <https://doi.org/10.62110/sciencein.jist.2025.v13.1016>
- [16] Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A.M., Qasem, S.N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*, 14(2): 144. <https://doi.org/10.3390/diagnostics14020144>
- [17] Sadr, H., Salari, A., Ashoobi, M.T., Nazari, M. (2024). Cardiovascular disease diagnosis: A holistic approach using the integration of machine learning and deep learning models. *European Journal of Medical Research*, 29(1): 455. <https://doi.org/10.1186/s40001-024-02044-7>
- [18] Babu, S.V., Ramya, P., Gracewell, J. (2024). Revolutionizing heart disease prediction with quantum-enhanced machine learning. *Scientific Reports*, 14(1): 7453. <https://doi.org/10.1038/s41598-024-55991-w>
- [19] Al-Alshaikh, H.A., P, P., Poonia, R.C., Saudagar, A.K.J., Yadav, M., AlSagri, H.S., AlSanad, A.A. (2024). Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific Reports*, 14(1): 7819. <https://doi.org/10.1038/s41598-024-58489-7>
- [20] Chagahi, M.H., Dashtaki, S.M., Moshiri, B., Piran, M.J. (2024). Cardiovascular disease detection using a novel stack-based ensemble classifier with aggregation layer, DOWA operator, and feature transformation. *Computers in Biology and Medicine*, 173: 108345. <https://doi.org/10.1016/j.compbiomed.2024.108345>
- [21] Chen, L., Ji, P., Ma, Y., Rong, Y., Ren, J. (2023). Custom machine learning algorithm for large-scale disease screening-taking heart disease data as an example. *Artificial Intelligence in Medicine*, 146: 102688. <https://doi.org/10.1016/j.artmed.2023.102688>
- [22] Shukur, B.S., Mijwil, M.M. (2023). Involving machine learning techniques in heart disease diagnosis: a performance analysis. *International Journal of Electrical and Computer Engineering*, 13(2): 2177-2185. <https://doi.org/10.11591/ijece.v13i2.pp2177-2185>
- [23] Almazroi, A.A., Aldahri, E.A., Bashir, S., Ashfaq, S. (2023). A clinical decision support system for heart disease prediction using deep learning. *IEEE Access*, 11:

61646-61659.

<https://doi.org/10.1109/ACCESS.2023.3285247>

- [24] Yewale, D., Vijayaragavan, S.P., Bairagi, V.K. (2023).  
An effective heart disease prediction framework based

on ensemble techniques in machine learning.  
International Journal of Advanced Computer Science and  
Applications, 14(2): 182-190.