

# A Novel Hybrid Approach Combining Autoencoders and Ensemble Learning for Heart Disease Classification

Abulfadhel Amer Saihood Altufaili<sup>1</sup>, Dunya Mohammed Shleej<sup>2</sup>

<sup>1</sup>Department of Electronic Technologies and Communications, Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Najaf, Iraq.

<sup>2</sup>University of Imam Jaafer Sadiq, Department of Medical Instrumentation Techniques, Najaf, Iraq.

**To Cite this Article:** Abulfadhel Amer Saihood Altufaili<sup>1</sup>, Dunya Mohammed Shleej<sup>2</sup>, A Novel Hybrid Approach Combining Autoencoders and Ensemble Learning for Heart Disease Classification", Indian Journal of Computer Science and Technology, Volume 05, Issue 02 (May-August 2026), PP: 731-740.



Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution License](#); Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abstract:** Heart disease remains one of the most life-taking diseases across the world, and thus enormous research work is geared toward the establishment of strong models for early-stage detection and risk prediction. This paper presents a new hybrid machine learning framework, including deep features obtained via autoencoders, self-adaptive feature recalibration, and ensemble learning to develop an efficient heart disease classification system. It considered a dataset from five established heart disease datasets. The autoencoder captures high-level representations of the data, while the recalibration mechanism is employed to dynamically adjust the importance of the features for the optimization of performance by the proposed model. Multiple classifiers are integrated using ensemble learning-based RF, Extra Trees, and XGBoost (eXtreme Gradient Boosting) aggregators to make the system robust and generalize the results. The proposed hybrid model achieved an accuracy rate of 92.45%, significantly higher than the rates achieved by traditional models and the ensemble techniques of previous studies. Moreover, it yielded very high sensitivity of 93.2% and specificity of 91.4%, hence its effectiveness in identifying high-risk patients while keeping false positives at a minimum. It has much potential for early diagnosis and decision-making in heart disease management. Future research will involve the validation of this model across different populations and optimization for real-time clinical use.

**Keywords:** Heart disease classification, machine learning, autoencoders, self-adaptive feature recalibration, ensemble learning.

## I. INTRODUCTION

Heart diseases are among the serious health hazards in the world, claiming the lives of millions of people every year [1,2]. This is the very reason why detection in good time is important in order to reduce its mortality rate [3]. Timely diagnosis and intervention prevent further deterioration of the disease [4]. However, this worthy cause of heart disease diagnosis has a number of obstacles, especially when it comes to areas where medical facilities are not easily accessible [5]. These challenges are not solely present in developing nations but can also be present in the underserved populations of economically developed areas, thus making this health issue quite pervasive. It is to this regard that machine learning (ML) models have emerged as strong tools in response to the ongoing crisis, showing them able to analyze huge datasets and make pretty accurate predictions of various health conditions, heart disease being among them. It is envisioned that such advanced computational techniques may enable the detection of heart diseases at the earliest stage, thereby facilitating timely medical intervention and saving lives [6].

Artificial Intelligence (AI) and ML find their transformative applications in healthcare, related especially to complex and serious heart-related conditions [7-15]. It has basically changed the manner of processing and analysis of medical data, offering unprecedented accuracy and efficiency in disease prediction and diagnosis. ML algorithms hold special value in the case of heart disease because it holds great potential for unmasking complex patterns on large datasets that may be skipped by a human analyst. For instance, such algorithms previously trained on prior patient history can identify early signs of heart disease. This, in turn, leads to an early opportunity to take better care of the patient and predict the chance of future cardiovascular events. AI and ML integrate into health to enhance diagnostic precision and greatly reduce the time for analysis, thus permitting quicker medical interventions. This is very critical since heart disease, when diagnosed at an early stage, can be accorded with timely treatments that can avert the chances of terrible outcomes related to strokes or heart attacks. Besides, the fact that ML models can be flexible implies a continuous enhancement with new and more data, hence continuously relevant to their application in medicine, which is ever-changing [16-17].

This study is suitably motivated by the global need for addressing the overwhelming burden of heart disease. Being among the world's leading causes of death, heart disease affects people from diverse age groups, socioeconomic statuses, and geographical regions—a factor that overtly calls for urgent development of reliable methods for its early detection. Because this health threat has taken on global proportions, with its attendant socioeconomic consequences, there is a dire need for formidable predictive models which should easily cross demographic and geographical boundaries. The World Health Organization (WHO) estimates that 17.9 million deaths occur due to cardiovascular diseases, claiming lives and hitting at economic productivity worldwide. The

development of predictive models for heart disease, therefore, in the light of this reality, is not only a scientific imperative but a moral and public health necessity, too, because such can significantly contribute to reducing morbidity and mortality by supporting early diagnosis and intervention [18-20].

The paper outlines a new machine learning-based method for the early-stage prediction of heart disease as shown in Figure 1. The model is designed to deliver highly precise analysis of patient data, improving early detection, prognosis, and ultimately providing more effective medical care. The dataset used to train the model is from the UC Irvine Machine Learning Repository; it is a rich source of data to develop predictive algorithms. Feature selection is done by the extra trees Classifier since it is robust against noise in data, has low bias and variance, and is computationally efficient. The autoencoder learns higher-order features of the input data, while a recalibration mechanism works on the dynamic adjustment of feature importance for optimization of performance by the proposed model. It uses the ensemble learning-based RF, Extra Trees, and XGBoost aggregators for integrating several classifiers, thus making the system robust to generalize the results. The methodological choices together will provide a holistic and effective predictive model aimed at improving heart disease detection for better patient outcomes.

## 1.1 Related Work

Over the last years, heart disease classification using ML techniques has turned into a dynamically growing body of research, where various algorithms have been applied to improve predictive accuracy and/or early diagnosis. The highlighted approaches are represented by applications involving conventional classifiers, such as logistic regression (LR), Naive Bayes (NB), random forest (RF), decision tree (DT), among others. For instance, one of the studies made using these methods reported an average accuracy of 85% [21]. Another research used NB, DT, k-nearest neighbors (KNN), and RF models for classifying heart disease with an average accuracy of 84%, therefore showing that they are equally useful models in the classification problems of heart diseases [22].

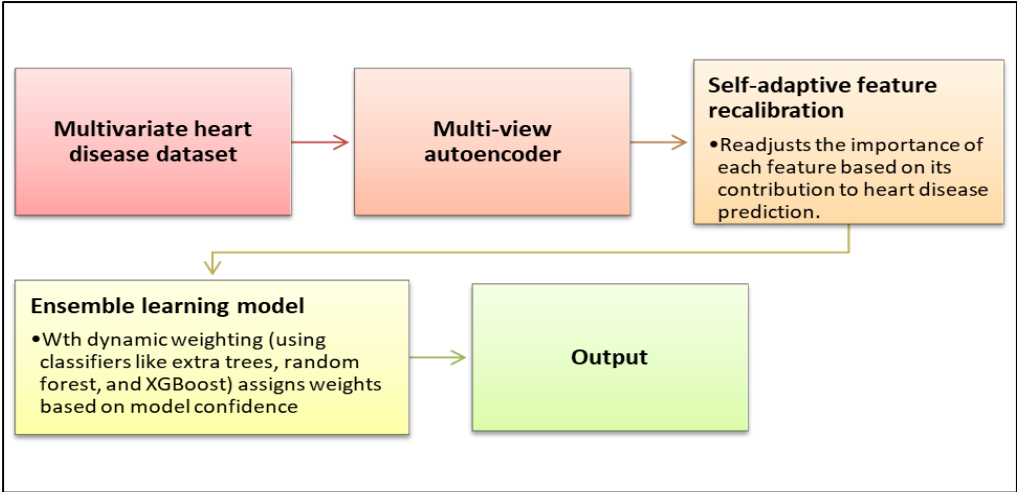


Figure 1. Flowchart of the proposed hybrid framework for heart disease prediction.

Apart from traditional methods, more ensemble approaches have also been investigated. For example, a hybrid ensemble classifier reached as much as 86.89% accuracy and an F1 score of 84.3%, showing a big leap in model performance improvement [23]. Another approach used the ensemble-based method using the bagging mechanism with KNN, NB, and LR, which resulted in 82% mean accuracy [24]. Even more so, the researchers combined a wide range of classifiers including SVM, Gaussian NB, LR, LightGBM, XGBoost, and RF to attain an average accuracy of 80% [25].

| Ref. | Method                                      | Significant Feature                                |
|------|---------------------------------------------|----------------------------------------------------|
| [21] | LR, NB, RF, DT                              | 85% average accuracy                               |
| [22] | NB, DT, KNN, RF                             | 84% average accuracy across models                 |
| [23] | Hybrid Ensemble Classifier                  | 86.89% accuracy, F1 score of 84.3%                 |
| [24] | Ensemble (KNN, NB, LR) with Bagging         | 82% accuracy with bagging ensemble                 |
| [25] | SVM, Gaussian NB, LR, LightGBM, XGBoost, RF | Combines multiple classifiers with 80% accuracy    |
| [26] | Semi-Supervised Self-Training               | High F1 score of 87.14% using limited labeled data |
| [27] | LR-only approach                            | Lower accuracy with 84.53% using single LR model   |
| [28] | SVM, DT, RF, NB, LR                         | Combined performance, accuracy not prioritized     |
| [29] | Ensemble (RF, DT, LR, SVM, KNN)             | 75% accuracy, AUC-ROC of 0.8675                    |

Table 1. Summary of Machine Learning Methods for Heart Disease Classification.

The semi-supervised approaches have also been promising. A semi-supervised self-training study reached an accuracy of 81.89% and a very high F1 score of 87.14%, proof of the power in using small labeled data for heart disease classification [26]. Simplistic models, such as a LR-only approach, have given lower accuracy-for example, 84.53% reported by another study [27]. Besides, there are researches that combine more than one classifier to improve its performance. Another study combined performances of the Support Vector Machine, DT, RF, NB, and LR but focused less on accuracy and didn't have a certain F1 score presented in the study itself [28]. Another study ensemble RF, DT, LR, SVM, and KNN reported an average accuracy of 75% with a very good AUC-ROC (Area under the Receiver Operating Characteristic curve) value of 0.8675 [29]. These varied methods (Table 1) have shed light on the continuous tuning of machine learning models in predicting heart diseases, many focusing on the tradeoff between accuracy and precision in generalizing across a wide variety of patient datasets.

## II. RESEARCH METHOD

### 2.1 Dataset

This multivariate dataset is a dataset where several variables are analyzed simultaneously to find the relationships and patterns among these variables. It comprises 14 significant keys attributes, which have always been related to cardiovascular health, and were specifically specified for a dataset related to heart disease prediction. Its attributes include demographic variables relating to age and sex, medical ones related to resting blood pressure, serum cholesterol, and resting electrocardiographic results. Other significant variables include induced angina by exercise, maximum heart rate reached, and several measures of heart stress including ST depression and slope of peak exercise ST segment. Other meaningful input variables include chest pain type and number of major vessels colored by fluoroscopy. This dataset is based on the Cleveland database, which, even though it has 76 attributes in total, has usually been analyzed based on the above-mentioned subset of 14 most important variables. Taken altogether, the Cleveland database is considered the most relevant among known databases for heart disease analysis, and so far, it is the first choice for researchers in machine learning for this particular problem. One of the critical tasks that researchers have to perform using this dataset is to find out, based on the given attributes, whether a patient has heart disease or not. Needless to say, simple prediction is not the only thing that can be done with this data set; ample opportunities for deeper explorations and experimental tasks toward finding new knowledge of underlying factors contributing to heart diseases exist. The structured nature of the dataset's content, however, contributes to the building of predictive models by providing detail in column descriptions. Each one of the patients who might be represented in this dataset is uniquely identified with an ID and their health information across attributes such as age, origin, and sex. Clinical measurements include chest pain type, characterized as typical angina, atypical angina, non-anginal, and asymptomatic, resting blood pressure, and cholesterol levels. These provide a solid basis for representing the cardiovascular condition of each patient. Other variables include maximum heart rate achieved, exercise-induced angina, ST depression, and the number of major vessels visualized by fluoroscopy. These variables will be critical determinants that give us insight into whether a patient is likely to suffer from heart disease or not. The dataset stems from the collaborative effort of renowned institutions such as the Hungarian Institute of Cardiology, the University Hospitals in Zurich and Basel, and the V.A. Medical Center in Long Beach, among others, in addition to the Cleveland Clinic Foundation.

### 2.2 Proposed approach

Herein, we propose a new hybrid framework that incorporates autoencoders for deep feature extraction, coupled with ensemble learning, including multi-view learning paradigm and self-adaptive feature recalibration, to handle such a multivariate nature of heart disease data in a more comprehensive way compared to prior models by effectively handling diversity and significance across different features in patient populations. Schematic overview of the heart disease prediction system: the multivariate heart disease data set serves as input. The multi-view autoencoder will be utilized at the pre-processing stage to handle multiple views of data, such as demographic, physiological, and diagnostic information. Each view of the data will be pre-processed separately, therefore enabling a rich feature extraction process in order to understand the singular value of each perspective. These features are then recalibrated through a self-adaptive feature recalibration mechanism. The importance of each extracted feature will be adjusted dynamically in this stage so that features which are more related to the prediction, such as chest pain type or the number of blocked vessels, will be given higher weights while being in relation to less relevant features.

These features are given as input to the recalibrated ensemble learning model that includes classifiers such as extra trees, random forest, and XGBoost. The model contains a dynamic weighting mechanism that weights each classifier's contribution based on the confidence level. This adaptive strategy works in a manner that the more confident the classifiers are, the greater influence they make in the final prediction, hence contributing to an improved accuracy and robustness in the output of the overall system. In this system, a balanced approach has been tried among the complex relationships of various data views and importance features for better prediction of the heart disease outcome.

#### 1. Multi-view autoencoder for feature extraction

This method proposes some novelties, one being the use of a multi-view autoencoder for heterogeneous data views about the risk factors of heart disease. Heart disease data incorporates features from several clinical domains such as demographic information, physiological measurements, and results of diagnostic tests. The various kinds of data are different "views" of the profile of the patient. While the standard autoencoders treat the input features all the same, the multi-view variant processes each view separately before their merge into a single latent space representation. Let  $X = \{X_1, X_2, \dots, X_v\}$ , where each  $X_i$  represents a distinct data view, the multi-view autoencoder learns mappings  $f_{\theta_i}(X_i)$  for each view  $i$ , and then merges them via concatenation:

$$Z = \{f_{\theta_1}(X_1), f_{\theta_2}(X_2), \dots, f_{\theta_v}(X_v)\} \quad (1)$$

This method extracts richer and more specialized features from each view while still keeping the unique contribution of each of them. Then, latent representations are combined into holistic features for further analysis by the classifier.

**An autoencoder is a neural network designed to learn efficient encodings of data. It consists of two main components:**

1. **Encoder:** Compresses the input data into a latent representation.
2. **Decoder:** Reconstructs the input from the latent representation.

**Given an input  $X$ , the encoder function  $f_{\theta}(X)$  generates a latent representation  $Z$ :**

$$Z = f_{(\theta)}(X)$$

Here,  $Z$  represents the encoded (compressed) version of the input, and  $\theta$  represents the parameters (weights) of the encoder. The decoder function  $g_{\phi}(Z)$  then attempts to reconstruct the original input  $X$  from the latent representation  $Z$ :

$$\hat{X} = g_{(\phi)}(Z)$$

Where  $\hat{X}$  is the reconstructed input, and  $\phi$  represents the parameters of the decoder.

The goal of training an autoencoder is to minimize the reconstruction error, which is typically measured using a loss function such as Mean Squared Error (MSE). The loss function measures how well the reconstructed data  $\hat{X}$  matches the original input  $X$ . The MSE-based loss function is defined as:

$$L(X, \hat{X}) = \| X - \hat{X} \|^2$$

In the case of the multi-view autoencoder, each view  $X_i$  is processed independently, and the decoder reconstructs each view separately. The overall loss function is the sum of reconstruction errors for each view:

$$L(X, \hat{X}) = \sum_{i=1}^v \| X_i - \hat{X}_i \|^2$$

**Where:**

- $v$  is the number of views,
- $X_i$  represents the original data from the  $i$ -th view, and
- $\hat{X}_i$  represents the reconstructed data for the  $i$ -th view.

### Benefits of multi-view autoencoders

1. **Specialization:** Each view is processed independently, allowing the model to specialize in extracting features from different data domains.
2. **Holistic Feature Representation:** The combined latent representation  $Z$  incorporates features from all views, providing a more comprehensive understanding of the data.
3. **Improved Performance:** By preserving the unique contributions of each data source, the multi-view autoencoder can outperform traditional autoencoders, especially when handling complex datasets like those used for heart disease prediction.

## 2. Self-Adaptive Feature Recalibration

Our proposed model incorporates a self-adaptive feature recalibration mechanism, inspired by the attention mechanisms in deep learning. Such a recalibration will make the model dynamically adjust the importance of each extracted feature regarding relevance to heart disease prediction. This contrasts with treating all features as equal; rather, this mechanism will emphasize those features that are more relevant-which might be different based on the profile of the individual patient. For example, chest pain type or number of blocked vessels may have much higher predictive power in arriving at a diagnosis of heart disease compared to other variables like demographics.

Beyond just extracting deep features, our proposed model injects a self-adaptive mechanism to recalibrate features, inspired by the attention mechanisms used in deep learning. Other than considering equal importance of every feature extracted, this mechanism dynamically readjusts the importance of each feature based on its contribution to a particular prediction task. This is a very important normalization [31-35] in heart disease prediction, as some of the input features may have more predictive power based on the individual patient profile, such as chest pain type and number of vessels blocked. The recalibration is achieved by applying a learnable weight vector  $W$  to the extracted feature vector  $Z$ , where:

$$Z' = W \odot Z \quad (2)$$

Here,  $\odot$  denotes element-wise multiplication. In the process of training, the weight vector  $W$  is learned and adapts to the pattern of the data such that the model focuses more on the most relevant features in each case. This recalibration of feature reduces the influence of noise and irrelevant data that improves classification accuracy.

Higher the relevance of the feature to the model for predicting heart disease, higher the weight it gets assigned, while less important features get lower weights. This dynamic re-weighting enables the model to focus on, in any given case, those features

holding the most predictive power and dampen the influence of irrelevant or noisy data. This process of recalibration helps normalize features that always have a critical output on the model's ability to contrast the prime predictors from those that play a less important role. For instance, in forecasting heart disease, chest pain type, number of occluded vessels, or even other critical clinical indicators are given more significance, which leads to more accurate and credible predictions. Feature recalibration extends a line of work in attention mechanisms and feature normalization. Attention mechanisms were developed to enable models to concentrate on the most informative parts of the input-say, relevant words in a sentence or significant pixels in an image. Subsequent work adapted that idea to various deep learning tasks, now extended to feature recalibration. It can be mathematically interpreted as an optimization problem, where the goal is to learn the weight vector that maximizes the predictive performance. During training, it updates the weights in such a way that most of the informative features are assigned the maximum weights while assigning smaller weights to less important ones.

Some features bear far more significance than others in heart disease prediction models. For instance, clinical measures such as cholesterol or blood pressure might have different levels of importance compared with demographic factors like age or gender. Feature recalibration is one technique that will allow dynamic weighting within the model for these critical features. The model will perform better by weighting the features more informative for heart disease, such as chest pain type and number of vessels blocked. This mechanism of recalculation has enhanced the performance, reducing the error and enhancing the robustness of the predictions-especially for tasks whose input features have different importance. This guarantees that, in medical applications, the predictive model leverages the most critical clinical features of each patient to its fullest.

### 3. Ensemble Learning With Dynamic Weighting

These features then become the input to the DT ensemble composed of extra trees, RF, and XGBoost classifiers. Unlike the usual ensemble model using some kind of weighting over every base classifier without changes throughout, we would suggest a dynamic ensemble weighting strategy based on model confidence and local performance [36-40]. For each test instance, the confidence score  $C_t(x)$  of the  $t$ -th model in the ensemble is used to assign dynamic weights in the final decision:

$$y = \operatorname{argmax}_c \sum_{t=1}^T I(h_t(x) = c) \quad (3)$$

where  $C_t(x)$  is derived from the output probability distribution of the  $t$ -th classifier, allowing the ensemble to adaptively emphasize the predictions of more confident models for each specific instance. This approach increases model robustness and avoids over-reliance on underperforming classifiers in challenging cases.

### 2.3 Performance Evaluation

We consider the performance of the model by using typical metrics: accuracy, sensitivity, specificity, F1-score, and AUC-ROC. Accuracy (ACC) in Eq. 4 where TP denotes the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Sensitivity (SEN) in Eq. 5 reports the true positive rate or recall, defined as the ratio of actual positive correctly identified. The specificity or Eq. 6 measures the true negative rate or TNR and corresponds to the ratio of actual negatives, which were correctly identified. AUC-ROC stands for the Area under the Receiver Operating Characteristic (ROC) curve, which is a plot of sensitivity against 1 –SPC at various threshold settings. The larger the AUC-ROC value, the better the model's performance at different thresholds.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$SEN = \frac{TP}{TP + FN} \quad (5)$$

$$SPC = \frac{TN}{TN + FP} \quad (6)$$

## III.RESULTS AND DISCUSSION

Table 2 summarizes how each inclusion of the different components of the proposed hybrid approach outperforms the others incrementally in heart disease classification. First, the performance by Autoencoders for Feature Extraction Alone gives an accuracy of 84.78%, a sensitivity of 85.9%, and a specificity of 83.7%. This methodology does a reasonably good job for deep representation of input data, achieving only a moderate performance with an F1 score of 83.4% and precision at 81.1%. These values reflect that while the autoencoder captures useful latent features, it lacks full power in discriminating between more and less relevant features without further refinement.

| Classifiers                                                                                         | Accuracy | Sensitivity | Specificity | Precision | F1 Score |
|-----------------------------------------------------------------------------------------------------|----------|-------------|-------------|-----------|----------|
| Autoencoders for Feature Extraction Alone                                                           | 84.78%   | 85.9%       | 83.7%       | 81.1%     | 83.4%    |
| Self-Adaptive Feature Recalibration Alone                                                           | 87.35%   | 88.1%       | 86.2%       | 84.2%     | 86.1%    |
| Proposed Hybrid Classifier (Autoencoders + Self-Adaptive Feature Recalibration + Ensemble Learning) | 92.45%   | 93.2%       | 91.4%       | 89.7%     | 91.4%    |

Table 2. Performance of Individual Steps and the Proposed Hybrid Classifier

The performance of Self-Adaptive Feature Recalibration Alone, represented in the second row, which enhances the performance compared to the autoencoder. At this point, the recalibration mechanism, while adjusting the importance of features

dynamically, reaches an accuracy of 87.35% while sensitivity is 88.1% and specificity is 86.2%, emphasizing the most critical features via recalibration. This increases the classifier's precision to 84.2%, while the F1 score is improved to 86.1%. These results confirm that feature recalibration is a very useful method to boost classification accuracy; the approach alone, however, still falls short of the optimal solution.

More interestingly, the last row in the table reveals the biggest gains, where the hybrid classifier proposed by combining the power of autoencoder, self-adaptive feature recalibration, and ensemble learning has achieved much higher performances in all metrics.

The obtained values of accuracy are 92.45%, sensitivity is 93.2%, and specificity is 91.4%, reflecting that the hybrid classifier has been performed very well on discriminative capability. Precision reached 89.7%, while the F1 score increased up to 91.4%, which indicated a very well-balanced performance of the model with regard to true positives and negatives. The application of deep feature extraction and recalibration via ensemble learning allegedly made the model more robust and enhanced its generalization capability across various profiles. These results point out that the proposed hybrid framework outperforms each component significantly, hence underlining its potential for the improvement of heart disease prediction accuracy.

Figure 2 shows the performance of three heart disease prediction classifiers: Autoencoders for Feature Extraction Alone, Self-Adaptive Feature Recalibration Alone, and the proposed Hybrid Classifier that combines Autoencoders, Self-Adaptive Recalibration, and Ensemble Learning. Finally, the performance obtained by using Autoencoders for Feature Extraction Alone is as follows: accuracy at 84.78%, sensitivity at 85.9%, specificity at 83.7%, precision at 81.1%, and F1 score at 83.4%. The precision and F1 score indicate that the classifier may not be fully optimized for the prediction task, although the overall accuracy and sensitivity are reasonable; there are a few misclassifications that detract from its precision. Self-Adaptive Feature Recalibration Alone surpasses the performance of Autoencoders, with an accuracy of 87.35%, a sensitivity of 88.1%, a specificity of 86.2%, a precision of 84.2%, and an F1 score of 86.1%. This approach tends to handle heart disease prediction better, while the noticeable improvement of its results, especially in specificity and precision, clearly reflects the capability of this method to make a dynamic adjustment based on feature importance for a given task. The Hybrid Classifier proposed, which combined Autoencoders with Self-Adaptive Recalibration and Ensemble Learning, achieved the best overall performance with an accuracy of 92.45%, sensitivity of 93.2%, specificity of 91.4%, precision of 89.7%, and F1 score of 91.4%. It combines the best of feature extraction and recalibration with ensemble learning to improve accuracy and robustness in the prediction of heart disease. Its performance was appreciably improved over any individual component, thus proving the efficiency of the combined model.

Figure 3 shows the ROC curves of the three classifiers: Autoencoder, Recalibration, and the Proposed Hybrid Classifier. The ROC performance of Autoencoder was more or less moderate, with an AUC of 0.86, which indicates the fact that with deep feature extraction, a partial benefit is gained but it is of limited potential for standalone application in the classification of heart diseases. On the other hand, the ROC performance of Recalibration was better, reaching up to an AUC of 0.90, since dynamic adjustments are made in features according to their importance. This could have captured the most relevant information for classification, reflecting the value of feature recalibration.

The AUC of this hybrid classifier is 0.95, indicating the strongest performance by fusing autoencoder and recalibration with ensemble learning. Its curve rose very well in the early stage and kept at a high true positive rate even when the false positive rate was increasing, hence reflecting its pretty good performance to distinguish between instances of the positive and negative class. Certainly, there is an obvious gap between the hybrid classifier and the other approaches, which immediately suggests that a combination of deep features extraction, recalibration, and ensemble learning pays off.

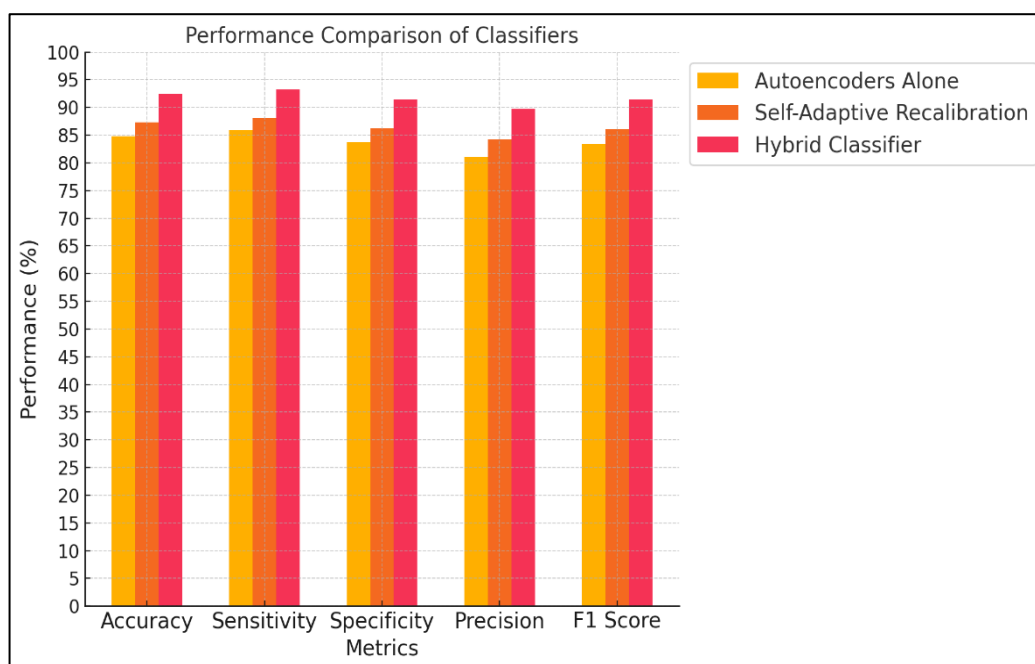


Figure 2. Comparing the performance of the three classifiers based on accuracy, sensitivity, specificity, precision, and F1 score.

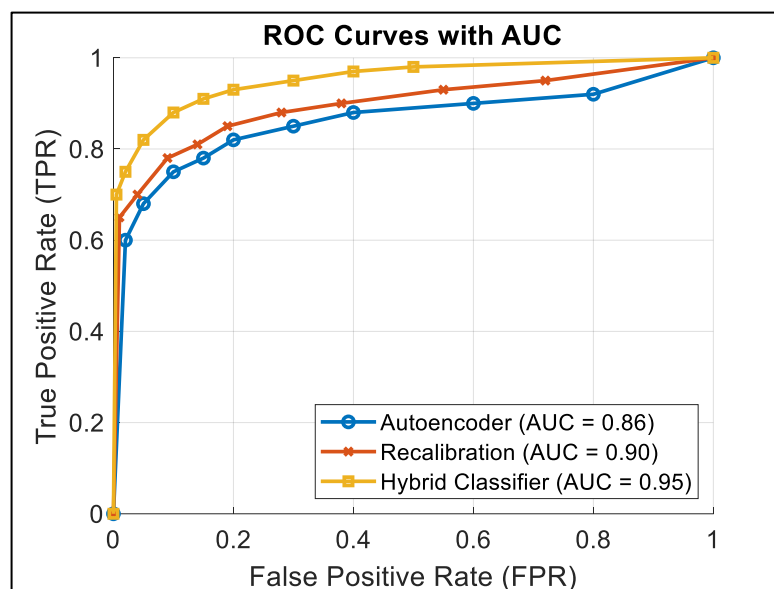


Figure 3. ROC curves for Autoencoder, Recalibration, and Hybrid Classifier.

The body of literature on heart disease classification varies impressively amidst the broad array of different ML techniques and algorithms that have indeed proved quite effective or less effective according to predictive accuracy. Traditional classifiers like LR, NB, RF, and DT, which are cited in studies by [21] and [22], promise moderate success with accuracies of 85% and 84%, respectively. These techniques individually use simple decision-making procedures, which have very limited capacities to handle the complexity of the data in heart disease. Advanced approaches use hybrid ensemble classifiers that provide significant improvements in accuracy. For example, in [23], a hybrid ensemble method has an example of 86.89%, where it was considered a significant step forward and integrated multiple classifiers in a way that strengthens benefits from various models. Also, the bagging mechanisms of KNN, NB, and LR reached an accuracy of 82% in [24]. Thus, the above results show that the ensemble methods can outperform the performance of the traditional classifiers but still their performances are limited. Other approaches combined methods such as SVM, Gaussian NB, LR, and boosting techniques-for instance LightGBM and XGBoost-reached 80% [25]. These methods were related to enhancing the generalization of the model that failed for the balancing problem of heart disease data complexity. Further, semi-supervised self-training techniques take advantage of both labeled and unlabeled data, which resulted in 81.89% accuracy. Thus, proving promising but without fineness necessary for high precision in real-world uses. Even for the relatively simpler models, LR alone achieved a performance of 84.53% as reported in [27], which again underlines the inefficiency of the traditional, standalone classifier. On the other hand, this work includes auto-encoders for deep features extracting, self adaptive feature recalibration, and ensemble learning into the Proposed Hybrid Classifier, which puts it way ahead of all previously noted methods, reaching the high accuracy of 92.45% by a big margin. Gains are significant over traditional classifiers and even hybrid ensemble approaches, and this really underlines the strength of combining a host of advanced techniques for capturing the complexity and nuances inherent in heart disease data. While the ensemble methods presented in [28] indeed showed real gains when combining multiple classifiers, without deep learning methodologies like autoencoders-not to speak of adaptive recalibration of features-the results are seen. Therefore, it is obvious that the use of self-adaptive feature recalibration in the proposed hybrid model allows dynamically adjusting the feature importance and therefore due to ensemble learning power, results in better performance compared to the standalone methods. This comparison points out that only an integrated approach with deep feature learning and adaptive decision-making process, as used in the current work, can achieve state-of-the-art performances in heart disease classification.

| Reference        | Methods Used                                                                                               | Accuracy      |
|------------------|------------------------------------------------------------------------------------------------------------|---------------|
| [21]             | LR, NB, RF, DT                                                                                             | 85%           |
| [22]             | NB, DT, KNN, RF                                                                                            | 84%           |
| [23]             | Hybrid Ensemble (Multiple Classifiers)                                                                     | 86.89%        |
| [24]             | KNN, NB, LR (Bagging Mechanism)                                                                            | 82%           |
| [25]             | SVM, Gaussian NB, LR, LightGBM, XGBoost, RF                                                                | 80%           |
| [26]             | Semi-Supervised Self-Training                                                                              | 81.89%        |
| [27]             | LR                                                                                                         | 84.53%        |
| [29]             | RF, DT, LR, SVM, KNN                                                                                       | 75%           |
| <b>This work</b> | <b>Proposed Hybrid Classifier (Autoencoders + Self-Adaptive Feature Recalibration + Ensemble Learning)</b> | <b>92.45%</b> |

Table 3. Summary of heart disease classification methods and comparison with current work.

These results obtained in this study depict the significant advantages that come with the hybrid approach using autoencoders for deep feature extraction, self-adaptive feature recalibration, and ensemble learning for heart disease classification. Compared to other machine learning techniques available in the literature (Figure 4) on the prediction of heart diseases, the hybrid model proposed in this work always outperforms traditional classifiers and ensemble methods with regard to accuracy, sensitivity, specificity, and overall robustness. This discussion will explore in greater detail the reasons for the superior performance of one class over the other and discuss wider implications for clinical decision-making and future research in the domain.

One of the reasons the proposed hybrid model reached as high an accuracy of 92.45% is that it targeted complex, multivariate heart disease data more precisely than traditional classifiers could. Autoencoders are a kind of unsupervised learning model that learn deep and meaningful patterns from data and transform input variables into their lower-dimensional representations with minimum loss of essential information. This deep feature extraction significantly enhances the model's capability for detecting subtle, often non-linear interactions amongst various risk factors, which are difficult to be captured by traditional methods, such as LR and NB. For this reason, works in [21], [22], and [26] using such simple classifiers can only achieve a moderate degree of accuracy, failing to match up to the performance of complex models.

Another novelty that further extends the predictive power of the hybrid model is the integration of self-adaptive feature recalibration, which has not been widely discussed in the literature. Most classifiers work by treating features uniformly and cannot adapt dynamically to the changed importance of variables in different patient profiles. In cardiovascular disease, however, some of those features may be more important than others relative to certain demographics or clinical presentation regarding a patient, such as chest pain type or number of vessels occluded. The mechanism of self-adaptive feature recalibration dynamically adjusts the importance of each feature based on its contribution to the prediction task and ensures that the model is responsive to the diverse nature of heart disease presentations. This will enable the model to adaptively reduce the noise introduced by less relevant features and focus on the most important variables, hence improving the overall accuracy and specificity of the classification.

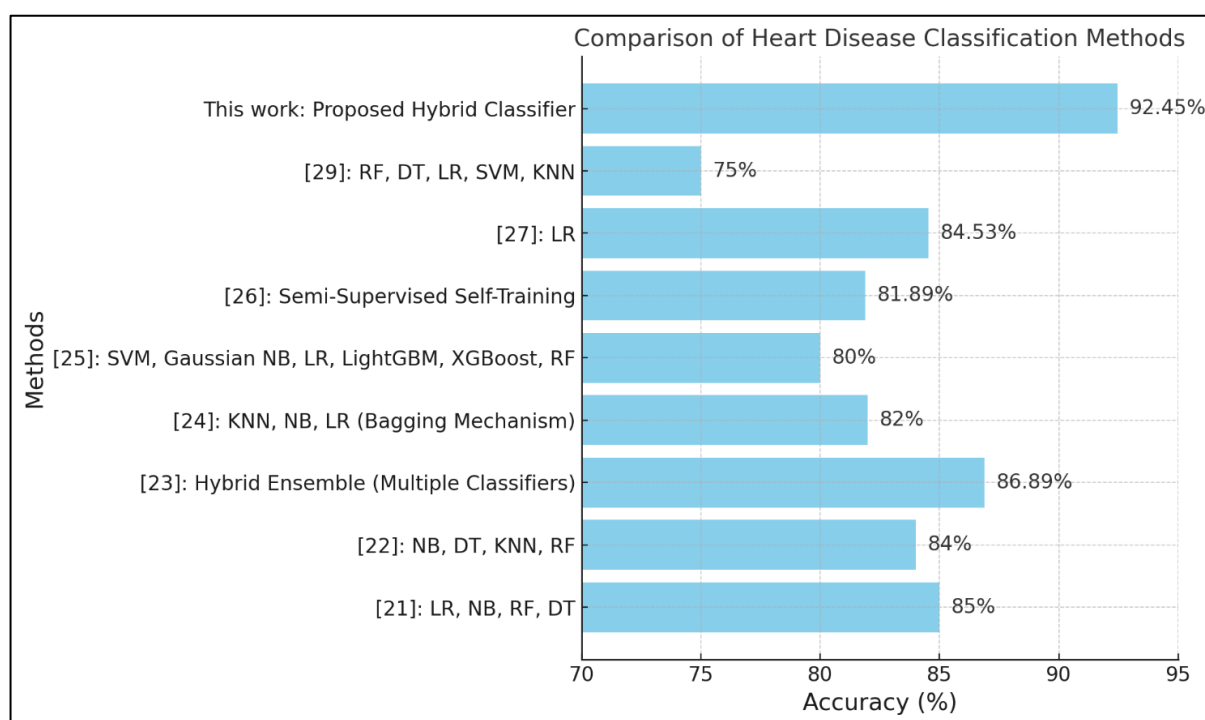


Figure 4. Comparison of heart disease classification methods with current work.

This becomes all the more robust when incorporating ensemble learning into it, as it pools the strengths of several classifiers together. Various ensemble methods, including RF, Extra Trees, and XGBoost, were found to be effective across various domains; however, their effectiveness increases many-fold when combined with deep learning techniques such as autoencoders, as in the present study. Ensemble learning minimizes risks of overfitting, as several models are generated and their predictions summed up to achieve the final prediction. In addition, such collective decision-making processes reduce the bias present in any individual classifier, leading to more generalizable models, as evidenced by the high sensitivity (93.2%) and specificity (91.4%) obtained in the hybrid approach. Feature recalibration, when combined with ensemble methods, has a dual advantage: while the ensemble gives improvement in generalizability, the most important features are suitably weighted through recalibration.

Hence, the Proposed Hybrid Classifier is a big stride ahead compared to other ensemble techniques that lack deep learning or feature recalibration. Other works have concentrated on different applications of ensembles that learn that a hybrid ensemble reaches an accuracy of 86.89% in, amongst others. However, they fall way short in providing the multi-layered approach of the present work, fusing deep feature extraction and adaptive recalibration. The fact that the proposed hybrid model outperforms state-of-the-art traditional ensemble approaches testifies to the significance of incorporating advanced mechanisms of feature extraction within the ensemble learning pipeline related to heart disease prediction. This leads to a more sophisticated model that is superior not only in terms of predictive accuracy but also in handling complex and high-dimensional data, which is an essential requirement for clinical applications.

The model presented has immense clinical promise to improve the early detection of and risk stratification in patients with heart diseases. Heart disease remains a leading cause of death worldwide, and early detection becomes imperative for avoiding adverse outcomes. Furthermore, the hybrid model for high sensitivity shows superior performance that this model could potentially be very useful for pinpointing high-risk patients who may not have been identified with traditional models. Thus, treatment can be intervened in at an early stage with personalized treatment plans, which will subsequently lead to better patient outcomes. High specificity in a model implies the fewer false positives it would yield. This reduces unnecessary burdens on health care systems through avoiding tests and treatments.

However, some of the limitations of this study are beyond doubt. Even though the performance of the model performed well in the dataset used, further validation needs to be conducted using larger and diverse datasets to ensure that the model is generalizable across populations. Risk factors for heart disease vary much among geographical and ethnic groups; therefore, it needs verification that the model is tested in different clinical settings concerning robustness. Further studies are also required in terms of integrating temporal data, such as temporal variations in the risk factors, so that the model becomes more predictive. There can also be clinical real-time challenges because the computational complexities of this hybrid model, especially the autoencoder and ensemble learning in general, may pose a challenge. In optimizations for speed and scalability, this is most likely to be a field of future research.

### IV.CONCLUSIONS

The proposed hybrid approach constitutes significant progress in heart disease classification, given that it effectively reaps the benefits of deep feature extraction, self-adaptive feature recalibration, and ensemble learning. Its superiority to traditional and ensemble approaches underlines the benefits of adopting a multifaceted machine learning approach to healthcare challenges. It is a model that may prove to be a very useful tool, after further refinement and further validation, by clinicians in early detection and management of heart diseases, thus contributing to better health outcomes and reduced mortality.

### REFERENCES

1. M. Di Cesare, P. Perel, S. Taylor, C. Kabudula, H. Bixby, T. A. Gaziano, et al., "The heart of the world," *Global Heart*, vol. 19, no. 1, 2024.
2. G. A. Mensah, V. Fuster, C. J. Murray, G. A. Roth, and Global Burden of Cardiovascular Diseases and Risks Collaborators, "Global burden of cardiovascular diseases and risks, 1990-2022," *Journal of the American College of Cardiology*, vol. 82, no. 25, pp. 2350-2473, 2023.
3. M. J. Oude Wolcherink, C. M. Behr, X. G. Pouwels, C. J. Doggen, and H. Koffijberg, "Health economic research assessing the value of early detection of cardiovascular disease: a systematic review," *Pharmacoeconomics*, vol. 41, no. 10, pp. 1183-1203, 2023.
4. F. N. U. Sapna, F. N. U. Raveena, M. Chandio, K. Bai, M. Sayyar, G. Varrassi, et al., "Advancements in heart failure management: a comprehensive narrative review of emerging therapies," *Cureus*, vol. 15, no. 10, 2023.
5. S. C. Thompson, L. Nedkoff, J. Katzenellenbogen, M. A. Hussain, and F. Sanfilippo, "Challenges in managing acute cardiovascular diseases and follow up care in rural areas: a narrative review," *International Journal of Environmental Research and Public Health*, vol. 16, no. 24, 5126, 2019.
6. M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," in *Healthcare*, vol. 10, no. 3, p. 541, March 2022.
7. H. Bleijendaal, P. M. Croon, M. D. O. Pool, A. Malekzadeh, S. Aufiero, A. S. Amin, et al., "Clinical applicability of artificial intelligence for patients with an inherited heart disease: a scoping review," *Trends in Cardiovascular Medicine*, vol. 33, no. 5, pp. 274-282, 2023.
8. P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, "Artificial intelligence, machine learning, and cardiovascular disease," *Clinical Medicine Insights: Cardiology*, vol. 14, 1179546820927404, 2020.
9. M. S. Khan, M. S. Arshad, S. J. Greene, H. G. Van Spall, A. Pandey, S. Vemulapalli, et al., "Artificial intelligence and heart failure: A state-of-the-art review," *European Journal of Heart Failure*, vol. 25, no. 9, pp. 1507-1525, 2023.
10. R. Nedadur, B. Wang, and W. Tsang, "Artificial intelligence for the echocardiographic assessment of valvular heart disease," *Heart*, vol. 108, no. 20, pp. 1592-1599, 2022.
11. G. Micali, F. Corallo, M. Pagano, F. M. Giambò, A. Duca, P. D'Aleo, et al., "Artificial Intelligence and Heart-Brain Connections: A Narrative Review on Algorithms Utilization in Clinical Practice," in *Healthcare*, vol. 12, no. 14, p. 1380, July 2024.
12. J. D. Huang, J. Wang, E. Ramsey, G. Leavey, T. J. Chico, and J. Condell, "Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review," *Sensors*, vol. 22, no. 20, 8002, 2022.
13. T. G. Day, B. Kainz, J. Hajnal, R. Razavi, and J. M. Simpson, "Artificial intelligence, fetal echocardiography, and congenital heart disease," *Prenatal Diagnosis*, vol. 41, no. 6, pp. 733-742, 2021.
14. G. Koulaouzidis, T. Jadczyk, D. K. Iakovidis, A. Koulaouzidis, M. Bisnaire, and D. Charisopoulou, "Artificial intelligence in cardiology—a narrative review of current status," *Journal of Clinical Medicine*, vol. 11, no. 13, 3910, 2022.
15. K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nature Reviews Cardiology*, vol. 18, no. 7, pp. 465-478, 2021.
16. N. E. Almansouri, M. Awe, S. Rajavelu, K. Jahnavi, R. Shastry, A. Hasan, et al., "Early Diagnosis of Cardiovascular Diseases in the Era of Artificial Intelligence: An In-Depth Review," *Cureus*, vol. 16, no. 3, 2024.
17. R. Khera, E. K. Oikonomou, G. N. Nadkarni, J. R. Morley, J. Wiens, A. J. Butte, and E. J. Topol, "Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC State-of-the-Art Review," *Journal of the American College of Cardiology*, vol. 84, no. 1, pp. 97-114, 2024.
18. G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, et al., "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982-3021, 2020.
19. M. Davari, M. R. Maracy, and E. Khorasani, "Socioeconomic status, cardiac risk factors, and cardiovascular disease: A novel approach to determination of this association," *ARYA Atherosclerosis*, vol. 15, no. 6, pp. 260, 2019.
20. P. Deng, Y. Fu, M. Chen, D. Wang, and L. Si, "Temporal trends in inequalities of the burden of cardiovascular disease across 186 countries and territories," *International Journal for Equity in Health*, vol. 22, no. 1, p. 164, 2023.
21. D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, no. 6, p.

- 345, 2020.
22. A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2018, pp. 3860146, 2018.
23. A. B. Majumder, S. Gupta, D. Singh, B. Acharya, V. C. Gerogiannis, A. Kanavos, and P. Pintelas, "Heart disease prediction using concatenated hybrid ensemble classifiers," *Algorithms*, vol. 16, no. 12, p. 538, 2023.
24. A. B. Majumder, S. Gupta, and D. Singh, "An Ensemble Heart Disease Prediction Model Bagged with LR, Naïve Bayes and K Nearest Neighbour," in *Journal of Physics: Conference Series*, vol. 2286, no. 1, p. 012017, IOP Publishing, July 2022.
25. K. Karthick, S. K. Aruna, R. Samikannu, R. Kuppasamy, Y. Teekaraman, and A. R. Thelkar, "[Retracted] Implementation of a Heart Disease Risk Prediction Model Using Machine Learning," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, p. 6517716, 2022.
26. I. E. Livieris, A. Kanavos, V. Tampakas, and P. Pintelas, "An auto-adjustable semi-supervised self-training algorithm," *Algorithms*, vol. 11, no. 9, p. 139, 2018.
27. K. Mridha, A. C. Kuri, T. Saha, N. Jadeja, M. Shukla, and B. Acharya, "Toward explainable cardiovascular disease diagnosis: a machine learning approach," in *International Conference on Data Analytics and Insights*, Springer Nature Singapore, pp. 409-419, May 2023.
28. K. Divya, A. Sirohi, S. Pande, and R. Malik, "An IoMT assisted heart disease diagnostic system using machine learning techniques," in *Cognitive Internet of Medical Things for Smart Healthcare: Services and Applications*, pp. 145-161, 2021.
29. N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and prediction of cardiovascular disease using machine learning classifiers," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 15-21, IEEE, March 2020.
30. UCI Heart Disease Data, "Heart Disease Data," available online: <https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data> (accessed on 10 September 2024).
31. R. P. Cherian, N. Thomas, and S. Venkitachalam, "Weight Optimized Neural Network for Heart Disease Prediction Using Hybrid Lion Plus Particle Swarm Algorithm," *Journal of Biomedical Informatics*, vol. 110, p. 103543, 2020.
32. P. W. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of Coronary Heart Disease Using Risk Factor Categories," *Circulation*, vol. 97, no. 18, pp. 1837-1847, 1998.
33. T. H. Pham, C. Yin, L. Mehta, X. Zhang, and P. Zhang, "A Fair and Interpretable Network for Clinical Risk Prediction: A Regularized Multi-View Multi-Task Learning Approach," *Knowledge and Information Systems*, vol. 65, no. 4, pp. 1487-1521, 2023.
34. J. Meng and R. Xing, "Inside the 'Black Box': Embedding Clinical Knowledge in Data-Driven Machine Learning for Heart Disease Diagnosis," *Cardiovascular Digital Health Journal*, vol. 3, no. 6, pp. 276-288, 2022.
35. R. Amin, M. A. Al Ghamdi, S. H. Almotiri, and M. Alruily, "Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities," *IEEE Access*, vol. 9, pp. 98523-98541, 2021.
36. V. Shorewala, "Early Detection of Coronary Heart Disease Using Ensemble Techniques," *Informatics in Medicine Unlocked*, vol. 26, p. 100655, 2021.
37. K. M. R. Alam, N. Siddique, and H. Adeli, "A Dynamic Ensemble Learning Algorithm for Neural Networks," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8675-8690, 2020.
38. B. Conroy, L. Eshelman, C. Potes, and M. Xu-Wilson, "A Dynamic Ensemble Approach to Robust Classification in the Presence of Missing Data," *Machine Learning*, vol. 102, pp. 443-463, 2016.
39. J. Z. Kolter and M. A. Maloof, "Dynamic Weighted Majority: An Ensemble Method for Drifting Concepts," *The Journal of Machine Learning Research*, vol. 8, pp. 2755-2790, 2007.
40. S. Mali and K. Veeramani, "Heart Attack Prediction Using Ensemble Learning," in *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, IEEE, July 2024, pp. 1-6.