

A Review on Cardiovascular Diseases Risk Prediction Approches Based on Machine Learning and Evolutionary Algorithms

Syed Akbar Ali Shah ^{1*}, Imtiaz Ali Korejo ¹, Gulsher Laghari ¹, Kamran Borhi ¹, Ghulam Ali ²

¹Institute of Mathematics and Computer Science (IMCS), University of Sindh, Jamshoro;

²Information Technology Centre, Sindh Agriculture University Tando Jam Pakistan

Keywords: Machine Learning (ML), Hybrid ML-GA Models, Evolutionary Algorithm (EA), Genetic Algorithm, Evolutionary Machine Learning Models.

Journal Info:

Submitted:
March 05, 2025
Accepted:
April 28, 2025
Published:
May 08, 2025

Abstract

Across the world, 17 million people die from heart disease each year. Heart-related diseases were the main cause for about 19% percent of deaths in Pakistan in 2016, the same has since now risen to 29%. As per most recent WHO statistics regarding prevalence of heart attacks in Pakistan, approximately more than two hundred thousand (200000) persons died in Pakistan in 2020 due to coronary heart disease, making up 16.49 percent of all fatalities. With a death rate of 193.56 per 100,000 inhabitants, Pakistan is ranked at number 30 in the world. Rising death rate due to heart disease can be minimized through detection at early stage. Different data mining approaches have made early detection of cardiac disease possible. Certain datasets are being used to retrieve useful information. Several machine learning techniques / models have been proved to be the most effective, accurate and profitable to detect cardiovascular disease at an early stage. However, the approach of machine learning and Genetic Algorithm (GA) with feature selection may aid in lowering the computational complexity of GA and increasing the effectiveness of its search for ideal solutions. Hence, there is dire need to apply such hybrid approach to get much more effective and accurate results. The goal of this survey paper is to review different papers related to CVD prediction at early stage by applying hybrid approach of ML Techniques with Genetic Algorithm. Moreover, the results obtained by the authors in reviewed papers are also examined. In the end, this survey will showcase the importance of the hybrid approach in improving accuracy of ML results.

***Correspondence author email address:** syedakbars@hotmail.com

DOI: [10.21015/vtcs.v13i1.2091](https://doi.org/10.21015/vtcs.v13i1.2091)



1 Introduction

The cardiovascular system relies on the heart as its central organ, responsible for circulating blood throughout the body. Any decline in its function can lead to disorders involving the heart and blood vessels, collectively known as cardiovascular diseases (CVDs). Globally, CVDs remain the foremost cause of death [1]. Recent reports indicate that millions of individuals have lost their lives to these conditions in past years [1], with a substantial proportion of these deaths occurring in low- and middle-income countries. In 2019 alone, non-communicable diseases accounted for nearly 17 million premature deaths (before 70 years of age), and CVDs contributed a major share of this burden [2, 3].

A significant number of cardiovascular complications can be minimized or prevented by controlling modifiable lifestyle risks such as tobacco use, unhealthy diet, obesity, lack of physical activity, and excessive alcohol intake [4, 5]. Timely identification of CVDs plays a key role in initiating appropriate management through clinical advice and therapeutic intervention [4]. Various forms of cardiovascular diseases exist, each triggered by different underlying factors, and many can lead to early mortality if not recognized and treated promptly.

Death due to heart disease has been an issue around the world nowadays and different measures are being taken to overcome this issue. An early detection of heart disease may prevent this menace and reduce the ratio of death occurred due to this disease. Different Artificial Intelligence (AI) techniques are being deployed to overcome this menace, especially machine learning techniques have performed pivotal role in detecting disease at an early stage [6]. Machine Learning (ML) has proved to be cost effective, time-saving and less human intervention as compared to traditional diagnosis [7]. Previously, medical history, physical examination report and analysis of concerning symptoms of the patients were used for diagnosis of CVDs by a doctor. But the results acquired from this process were not accurate and satisfactory in identifying the patient of heart disease. Additionally, it was expensive and hard to analyze computationally [8]. ML is said to be the powerful tool because of its ML algorithms. These algorithms help ML model to learn, analysis and predict the results from data. Evolutionary Computation (EC) is also one of the categories of AI which empowers machine learning algorithms and improve its performance and get optimized results.

The purpose of this review is to have further understanding about the contribution of ML techniques and EC in disease diagnosis especially cardiovascular disease at an early stage individually. And, examine the results obtained by integrating EAs in different ML phases which is said to be an Evolutionary Machine Learning (EML) / hybrid approach to enhance the performance of ML in respect of accuracy, precision, recall, sensitivity, specificity, F1 Score etc.

The main contribution of this paper is to review the results of several related work, in detection of heart disease at an early stage using hybrid approach. The researchers carried out hybrid approaches in various literature review to fulfill the objective. During this survey, following few points were kept under consideration:

1. Basic concept of ML
2. Frequently applied modern model / techniques of ML in heart disease diagnosis.
3. Overview of Evolutionary Computation (EC)
4. Integration of ML and EC for early detection of heart disease.

This survey paper is likely to contribute in the ongoing research related to heart diseases using hybrid approach. The integration of hybrid approaches of ML and EC would help in getting better results after applying it on the data-set for early detection of heart diseases.

This paper summarizes the various Machine Learning (ML) Techniques with Evolutionary Algorithm in diagnosis of Cardiovascular disease at an early stage. The rest of the paper is structured as follows: Section 2 discusses the basic concepts of ML, whereas Section 3 contains the overview of Machine Learning Algorithms and

its contribution in disease diagnosis and how evaluation metrics are measured. Section 4 describes evolutionary computation and hybrid approach models. Section 5 discusses the contributions of different authors and their experiments with results who most frequently utilized hybrid approach. Section 7 discusses the outcomes. Finally, Section 8 concludes the article with a general conclusion. Acronyms used in this paper are shown in Table 2.

Table 1. List of Acronym

Acronym	Description	Acronym	Description
GA	Genetic Algorithm	ML	Machine Learning
WHO	World Health Organization	CVD	Cardiovascular Disease
EC	Evolutionary Computation	EML	Evolutionary Machine Learning
PCA	Principal Component Analysis	DQN	Deep Q Network
DT	Decision Tree	SVM	Support Vector Machine
UCI	University of California Irvine	EAs	Evolutionary Algorithms
SIAs	Swarm Intelligence Algorithms	KNN	K-Nearest Neighbor
LR	Logistic Regression	MOEFC	Multi-Objective Evolutionary Fuzzy Classifier
GFS-LB	Genetic Fuzzy System-LogitBoost	FURIA	Fuzzy Unordered Rule Induction
FH-GBML	Fuzzy Hybrid Genetic Based Machine Learning	ANN	Artificial Neural Network
BPNN	Back Propagation	MLAs	Machine Learning Algorithms
NB	Naive Bayes	ACO	Ant Colony Optimization
ROC	Receiver Operating Characteristics	TPR	True Positive Rate
AdaBoost	Adaptive Boosting	RF	Random Forest

2 Basic Concepts of Machine Learning

Machine Learning (ML) is a branch of artificial intelligence which lets computers learn from data and make predictions based on that data without having explicitly programmed [9]. The aim of ML is to create models that can recognize patterns in data and predict or decide based on that data. Applications including image identification, natural language processing, and predictive analytics frequently use ML techniques[10]. The deployment of statistics and computer science aid in making accurate decision [11]. It extracts patterns utilizing knowledge discovery, algebra, statistics and optimization strategies. Figure. 1 depicts the fundamental taxonomy of ML, which is divided into three primary categories i.e Supervised, Unsupervised and Reinforcement Learning (RL).

2.1 Supervised Learning

Supervise learning is a type of machine learning where the algorithm learns from labeled data. In this approach we have input data (features) along with corresponding output label (target variable).

1. **Labeled Data:** The dataset used for training contains known output labels.
2. **Prediction Task:** The goal is to predict or classify new, unseen data based on the learned patterns from historical data, e.g. Predicting house prices based on features like area, number of bedrooms, etc. Classifying emails as spam or not.
3. **Algorithms:** Naive Bayes, Support Vector Machines (SVM), Logistic Regression, etc.

2.2 Unsupervised Learning

Unsupervised learning deals with unlabeled data. The algorithm identifies patterns, structures, or relationships within the data without any predefined output labels.

1. **No Labels:** The input data lacks explicit output labels.

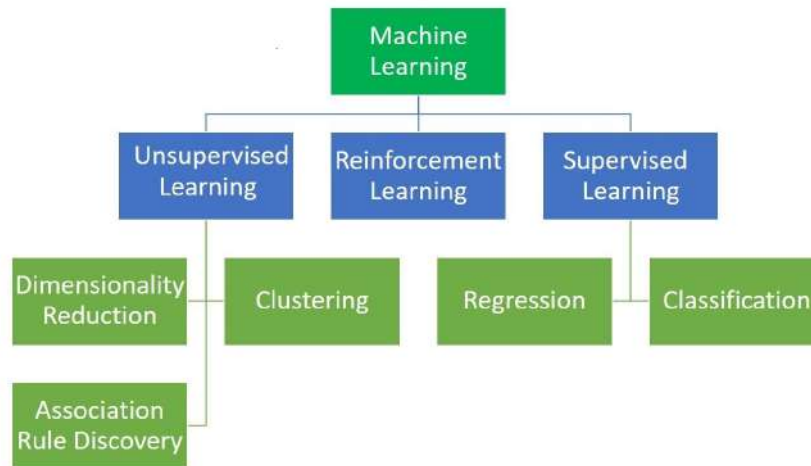


Figure 1. Taxonomy of machine learning

2. **Clustering and Dimensionality Reduction:** Common tasks include clustering similar data points and reducing the dimensionality of the data, e.g. Grouping similar customer profiles for targeted marketing. Discovering hidden patterns in customer behavior.
3. **Algorithms:** K-Means, Hierarchical Clustering, Principal Component Analysis (PCA), etc.

2.3 Reinforcement Learning

Reinforcement learning involves an agent interacting with an environment to learn how to take actions that maximize cumulative rewards.

1. **Reward-Based Learning:** The agent receives rewards or penalties based on its actions.
2. **Exploration and Exploitation:** Balancing exploration (trying new actions) and exploitation (choosing known actions) is crucial, e.g Training game-playing agents (e.g., AlphaGo, chess engines). Autonomous vehicles learning to navigate.
3. **Algorithms:** Q-Learning, Deep Q Networks (DQN), Policy Gradient methods

Machine learning techniques has demonstrated value in providing a huge platform in the medical industry so that healthcare concerns may be treated easily and quickly [12, 13]. Healthcare is one of the top industries where data and technology may be readily enhanced. It is not possible to manage the enormous amount of medical data without passing through computation processes [14]. It has been a challenge for medical experts and doctors to properly investigate signs and symptoms and recognize diseases at an early stage in a short span of time efficiently and accurately [15]. Figure.2 shows the schematic of model building using ML.

Reducing the complexity and improving the efficiency of the problem has been an issue of Machine Learning (ML) all the time, but with the help of Genetic algorithms (GA), ML can perform this task very effectively and efficiently during the optimization process [16].

Machine learning is being used to develop new and innovative ways to diagnose diseases, predict patient outcomes and even develop new treatments [17, 18]. Table. ?? shows the year wise contributions of some of the authors in ML for real time applications and are currently being used in different fields.

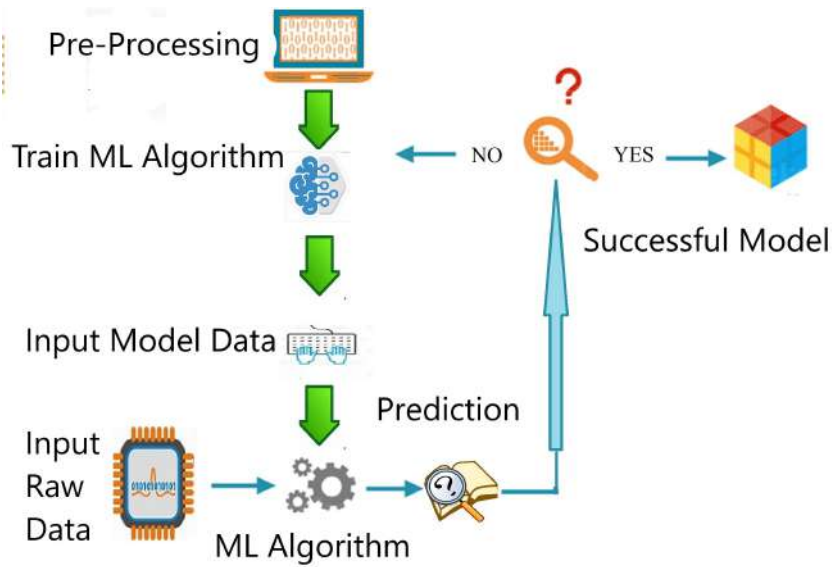


Figure 2. Schematic of model building using ML[?]

Table 2. Contribution of authors in ML for real time application

Acronym	Description	Acronym	Description
Author	Applications	Year	
Sergio Francisco Da Silva et. Al [20]	Disease Diagnosis	2011	
P. A. Castillo,Maribel Garcia Arenas et. Al [21]	Cancer Classification	2006	
Halina Kwa?nicka and Kajetan ?witalski et.al [22]	Medicine	2006	
Patrick C. H. Ma [23]	Healthcare	2006	
Veenu Mangat and Renu Vig [24]	Gene Expression and Data Analysis	2014	
Qin Song [25]	Gastrointestinal Infection Prediction	2017	
Chun-Hui Wu [26]	Cardio vascular Disease Detection	2019	
Cuong To and Tuan D. Pham et.al [27]	Medicine	2009	
Gunjan Chugh et.al [28]	Breast Cancer Diagnosis	2021	

3 Machine Learning Algorithms

This section reflects comprehensive review regarding ML algorithms mostly used in disease diagnosis [19].

3.1 K-Nearest Neighbor

K-Nearest Neighbor was invented in 1951 by Evelyn Fix and Joseph Hodges and is one of the supervised machine learning algorithms and mostly used in regression and classification problems [20]. It is sometime called as lazy learner algorithm due to the holding of data and perform operation on it instead of learning directly from trained dataset [21].

3.2 Decision Tree

The divide-and-conquer rule is used in decision tree (DT). In decision tree, the fields are converted into a tree where leaves indicate classes while branches represent the mixture of attributes of those classes [22]. The trees are constructed based on high entropy inputs [23] as shown in equation (1).

$$Entropy = \sum_{j=0}^m p_{ij} \log_2 p_{ij} \quad (1)$$

where m represents possible outcome, j describe the index which repeats on each possible results, i implies to the index which interprets different conditions related with each outcome, p_{ij} correspond to joint probability of outcome (j) happened under situation (i) while \log_2 is the measurement unit of entropy.

3.3 Support Vector Machine

It is a type of supervised machine learning and used to resolve regression and classification problems [24]. Hyper-plane is sought to divide data into different classes. SVM is also broadly used in the field of computation biology and other classifications i.e text and image etc.

3.4 Logistic Regression

For binary classification, logistic regression is widely recognized as one of the most effective machine learning techniques. Because of its simplicity, which may be used to a wide range of issues and offers appropriate solutions [25].

3.5 Naive Bayes

One of the greatest and strongly scalable machine learning classification algorithms that makes use of the Bayesian method is the Naive Bayes classifier, Naive Bayes classification requires linear variables in the form of predictor variables in the problem statement. For the classification task, Naive Bayes is a probabilistic machine learning method based on Bayes' theorem [26].

3.6 Random Forest

This approach for supervised learning can be applied to regression as well as classification. It is an ensemble of several decision trees that uses each tree's judgment to classify samples. A certain class is assigned to the provided input sample by majority voting procedures [27]. Random forest has got popularity due to dealing with missing data in biomedical research [28].

3.7 Boosting Algorithms (Gradient Boost, AdaBoost and XGBoost)

3.7.1 Gradient Boosting:

is one of the machine learning algorithms usually used for tabular datasets [29]. By adjusting the weights each estimator is added progressively as the weak learners are trained one after the other. The goal of gradient boosting is to reduce the discrepancy between the expected and actual values by forecasting the residual errors of earlier estimators [37].

3.7.2 AdaBoost:

In [30], the authors have developed adaptive Boosting, popularly known as AdaBoost. AdaBoost adjusts all the weights without considering the prior knowledge of weak learners. The estimator's error rate during model training serves as a proxy for the base learner's weakness. Using the AdaBoost algorithm, decision tree terms are frequently employed to address classification and regression issues [37].

3.7.3 XGBoost:

XGBoost [38] is a type of gradient boosted decision trees. Decision Trees are created sequentially in this algorithm [39]. XGBoost computes the similarities scores independently by merging various decision tree types (weak learners) through adaptation of the gradient decent and regularization process. It aids in overcoming the issue of over fitting during the training phase [37].

3.8 Performance Metrics of Machine Learning:

This section describes the performance metrics of Machine Learning Algorithm used by the authors in literature review to measure the performance of algorithms based on confusion matrix. These performance indicators i.e Accuracy, Precision, F1 Score, Sensitivity, Specificity, Error Rate, ROC curve etc are mostly used to evaluate performance in disease diagnosis [9].

3.8.1 Accuracy:

Accuracy is one of the performance metrics of machine learning which is calculated to point out the accuracy rate of algorithm [40]. Formula for calculating Accuracy of an algorithm is shown in equation (2):

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (2)$$

Where Tp denotes number of True Positive, Tn represents number of True Negative while Fp implies number of False positive and Fn number of False negative.

3.8.2 Precision:

Precision is used to measure the proportion of positive observations out of precisely predicted observations, see equation (3).

$$\text{Precision} = \frac{Tp}{Tp + Fp} \quad (3)$$

3.8.3 Recall:

The recall is calculated to get the proportion of overall relevant results that the algorithm properly recognizes is referred to as recall. See equation (4)

$$\text{Recall} = \frac{Tp}{Tp + Fn} \quad (4)$$

3.8.4 Sensitivity:

Sensitivity denotes only true positive measure considering total instances and is measured as follows in equation (5):

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn} \quad (5)$$

3.8.5 Specificity:

It identifies how many true negatives are appropriately identified and calculated as follows in equation (6):

$$\text{Specificity} = \frac{Tn}{Tn + Fp} \quad (6)$$

3.8.6 F-measure:

The F-measure, also known as F1 score. F1 score is the mean of accuracy and recall in a harmonic manner. The highest F score is 1, indicating perfect precision and recall score as in equation (7):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Where "2" is a constant value and used as multiplier to balance precision and recall.

3.9 Machine Learning Contribution

3.9.1 Disease Diagnosis:

ML algorithms can be trained on large datasets of medical records and images to learn how to identify patterns that are associated with different diseases [41]. This information can then be used to develop new and more accurate diagnostic tools [18, 30].

3.9.2 Classification:

ML can be used to classify different types of cancer based on their genetic or molecular profiles [42]. This information can help doctors to choose the best course of treatment for each patient [43].

3.9.3 Gene Expression and Data Analysis:

ML can be used to analyze large datasets of gene expression data to identify patterns that are associated with different diseases [42]. This information can help researchers to understand the underlying causes of disease and to develop new diagnostic and therapeutically strategies [43].

3.9.4 Gastrointestinal Infection Protection:

ML can be used to predict which patients are at risk of developing gastrointestinal infections. This information can help to improve early detection and treatment of these infections [45].

3.9.5 Cardiovascular disease detection:

ML can be used to detect early signs of cardiovascular disease such as atherosclerosis. This information can help to prevent or delay the onset of cardiovascular events. Overall, the machine learning is a powerful tool that is being used to develop new or innovative ways to improve the diagnosis and treatment of diseases [46, 47].

4 Evolutionary Computation:

Evolutionary Computation (EC) is one of the core branches of artificial intelligence and computational intelligence, which is inspired by the natural selection mechanism of "Survival of the fittest" and it is the process of biological evolution. This approach has been successfully applied for solving many optimization problems due to the properties of being easy-to-use and its robustness for determining potential solution for complex problems and many application of EC in the various areas such as medical science, engineering, business and social science. The main branches of EC is shown in Figure. 3, there are two main branches of EC: Evolutionary Algorithms (EAs) and Swarm Intelligence Algorithms(SIAs)[48].

EC uses rules of evolution and genetics to reduce the complexity of the problem. GA is a branch of EA which is used to reduce the complexity and improve the performance of accuracy of Machine Learning.

GA and ML are two different techniques but have some commonalities to solve the complex problems [36]. Hybrid approach ML-GA is being applied to seek optimized solution of the problem and predict coronary heart

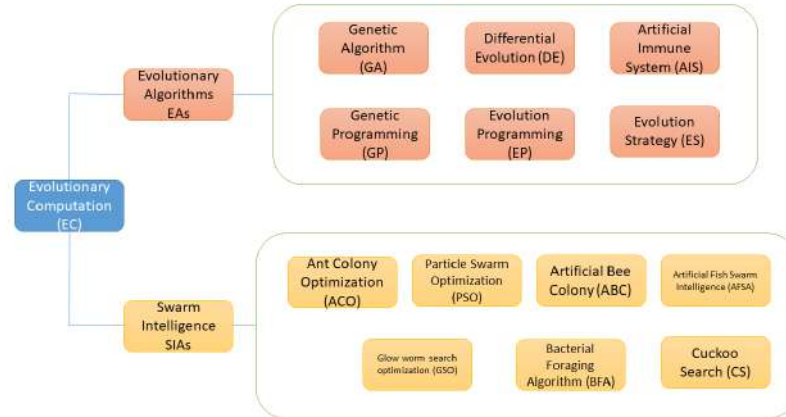


Figure 3. Evolutionary Computation

disease at an early stage by analyzing Physiological and Demographical features selection or reduction data of a person available at machine learning repository developed at University of California Irvine (UCI).

Dealing with high-dimensional data and sophisticated models is one of the primary issues in machine learning. Finding the best answers becomes challenging as the number of variables or features in a model rises since the search space expands exponentially [49]. By producing a smaller collection of potential solutions that are more likely to be optimal, genetic algorithms can aid in reducing the complexity of the problem space. This is accomplished by developing new candidate solutions based on the top-performing solutions from the previous generation using procedures including crossover, mutation, and selection [9].

Moreover, ML and GA may be coupled to produce hybrid algorithms that make use of each technique's advantages. The fitness function in GA, which determines the quality of potential solutions, may be learned using machine learning [50]. This may aid in lowering the computational complexity of GA and increasing the effectiveness of its search for ideal solutions [51].

4.1 Evolutionary Machine Learning (EML):

The term EML is used when the evolutionary algorithm is applied with Machine Learning techniques to improve the performance of ML [16]. ML contains three parts: Pre-processing, learning and Post processing and EML can be applied in each part to get better results as shown in figure. 4.

5 Literature Review:

The authors in [52] used genetic algorithm (GA) for prediction of heart disease using dataset on heart disease available on UCI Machine learning repository. Feature engineering techniques i.e MSE and RMSE were used to measure the performance and it was found that female patients were more affected than male by the heart disease.

In [53], the authors proposed a system and used traditional machine learning techniques with one of the Boosting Algorithm i.e Gradient Boosting and also used GA for optimization and reduced the parameter by 20% and also implemented hyper parameter optimization techniques for further improvement of the performance of the predictive model.

The authors in [54] used RBF network and GA for enhancement of prediction of heart disease. RBF was used to create the classifier while the authors used GA for minimization of features and the features were minimized from

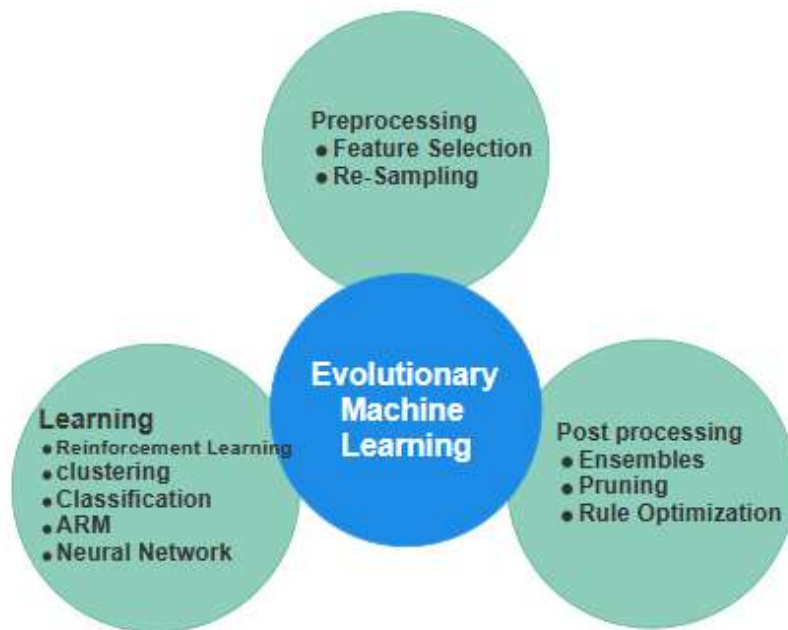


Figure 4. Evolutionary Machine Learning Classification Approach

14 to 9 features which ultimately increased the accuracy of model upto 85.48% using proposed system which was more accurate than Naive Bayes and J48.

The authors in [4] tested other data sets taken from UCI repository and applied a hybrid algorithm which combines KNN with genetic algorithm for optimized classification that helped the doctors to classify and diagnose heart disease efficiently with minimum features.

A hybrid algorithm proposed by the authors [9] wherein they utilized KNN to extract and SVM for classification to enhance efficiency in results during prediction analysis. The authors investigated collected datasets of real time patients of heart disease and got 75% percent accuracy in KNN and 76 percentage in SVM while in proposed hybrid approach got accuracy of 81%.

The authors in [55] examined a diverse set of classification techniques, such as Logistic Regression (LR), Adaptive Boosting (AdaBoost), the Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), Fuzzy Unordered Rule Induction (FURIA), the Genetic Fuzzy System combined with LogitBoost (GFS-LB), and the Fuzzy Hybrid Genetic-Based Machine Learning approach (FH-GBML). These techniques were integrated as a hybrid framework, and their comparative outcomes were analyzed to determine the most suitable model for cardiovascular disease (CVD) prediction. The authors also employed Weka and Keel tools to compute accuracy, sensitivity, specificity, and error rate [56].

In [57], an Advanced Hybrid Machine Learning (AHML) model was introduced for precise detection of CVD. The proposed system operates in four major stages: (i) balancing the raw dataset using various sampling strategies, (ii) applying Chi-square statistics to extract the most significant features, (iii) utilizing a Stacking Ensemble Model (SEM) composed of Random Forest Trees (RFT), K-Nearest Neighbors (KNN), and AdaBoost as base classifiers, and (iv) evaluating the model using accuracy, sensitivity, specificity, F1-score, and ROC-AUC. The authors reported that their hybrid method outperformed existing comparable approaches.

In [58], the researchers proposed a hybrid classification model based on Support Vector Machines (SVM) combined with an Artificial Neural Network (ANN), measuring both accuracy and testing time. They further in-

corporated a Back-Propagation Neural Network (BPNN) for classification and assessed its training and testing outcomes across multiple sample sizes, noting improved accuracy with increased data. Another hybrid model, HRFLM—merging Random Forest (RF) and Linear Methods (LM)—was shown to be highly reliable for heart-disease prediction. The authors recommended additional experimentation on real clinical datasets rather than relying primarily on controlled simulations.

A UCI dataset with 303 records and 76 attributes was used in [59], from which 14 relevant features were selected for evaluation. Various supervised learning models were tested, including K-Nearest Neighbor (KNN), Random Forest, Decision Tree Classifier, Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes. The authors experimented with KNN using eight neighbors and compared its performance—through metrics such as sensitivity, precision, accuracy, and F1-score—with that of the other classifiers, including Naïve Bayes, linear-kernel SVM, and Decision Trees with 4 and 18 selected features, as well as Random Forest [60].

The authors of [13] applied four distinct machine-learning methods to perform a comparative analysis and observed positive predictive performance across the models. They concluded that machine learning offers an effective pathway for heart-disease prediction and classification. Using the UCI heart-disease dataset, they compared metrics such as precision, F1-score, accuracy, and recall across all models and found KNN to perform best. The authors suggested that subsequent research could extend the comparison to additional classifiers and explore datasets covering other biological conditions.

The study in [56] emphasized the critical role of early detection and timely treatment in managing CVD. The authors assessed ten ML algorithms on four datasets—two for CVD prediction and two for CVD diagnosis. They adopted both train-test splitting and k-fold cross-validation to analyze the performance of the algorithms using accuracy, precision, recall, and F1-score. Among the evaluated metrics, accuracy emerged as the most effective performance indicator [61].

Their findings demonstrated significant potential for enhancing CVD diagnosis in regions with limited medical infrastructure. Two datasets (Cleveland and Faisalabad) were examined for CVD diagnosis, while the Framingham and South African datasets were assessed for CVD prediction.

Across these analyses, three key risk factors—Chest Pain Type (cp), Serum Creatinine, and Ejection Fraction—were consistently identified as major indicators of cardiovascular abnormalities such as arrhythmia and tachycardia. These attributes could therefore support early-stage screening and timely clinical intervention.

The authors further recommended applying a similar methodological framework to investigate and predict other diseases, including diabetes and breast cancer, either through clinical systems or mobile-based diagnostic solutions.

The authors developed a system that used the MICE model to resolve the missing data problem in the pre-processing step and applied XGboost, Adaboost, gradient boosting, extra trees, light gradient boosting Lightgbm, SGDC, Nu SVM and the stacking algorithm in the classification step and got the accuracy score of 95.83% by using the stacking algorithm [62].

In future study, the author will take his research towards development of a system to improve medical treatment and predict effectively with minimum expenses. Other disorders, such as diabetes, that are clinically related to heart disease will be considered and the prediction system will be integrated with other systems to support the healthcare system's security foundations.

In his study, data gathered from four hospitals of four distinct locations and investigated which strategy would be the best for forecasting the diseases. The efficiency of various data mining methods was compared by the authors to predict cardiac disease, which included K-nearest neighbors, Random Forest, and Multi-layer Perceptron, Logistic Regression. After examining the Data Mining techniques, it is reported on how well each method for prediction performed [63]. The results showed that more than 97% percent of heart problems may be predicted.

Information mining involves examining unstructured data to identify crucial designs that might guide future applications. To predict cardiac diseases, this technique makes use of many categorization strategies. Our study strategy incorporates SVM classification to build a model for predictive analysis and uses KNN to extract attributes from an infinite dataset. In comparison to developed computations like KNN and SVM, our suggested technique produces common results. Our analysis demonstrates that the crossbreed strategy outperforms the execution of other algorithms on the same dataset, achieving an 81% exactness rate in forecasting heart disease prognosis [63].

Data mining and machine learning approaches were used by the author to examine the state of the art of several clinical decision support systems for the prediction of cardiac disease. The author used numerous classification methods, with varying degrees of accuracy, to predict cardiac illnesses, including the Naive Bayes (NB), Decision Tree (DT), and Artificial Neural Network (ANN). To boost the accuracy of forecasting the early beginning of the illness, more complicated models that combine many geographically different data sources were required. As a result, only a limited success was obtained in the building of such predictive models for heart disease patients [63].

The review of existing literature indicates that a large proportion of studies have relied almost exclusively on the Cleveland Heart Disease dataset, which contains only 303 instances and 14 attributes. Owing to its limited size and narrow regional representation, this dataset does not adequately capture the diversity found in broader populations. Although a few studies incorporated additional data sources, the overall feature set and sample variability remained restricted. Consequently, the reported classification accuracies differ substantially across studies and cannot be reliably generalized. To address these shortcomings, our future work aims to utilize multiple heart disease datasets collected from geographically varied populations and enriched with a wider range of clinical features. Such an approach is expected to support the development of more robust machine learning models capable of improved disease categorization and early prediction, ultimately contributing to reduced morbidity and mortality associated with cardiovascular disorders.

Recent investigations also show that evolutionary computation (EC) techniques are increasingly being used to enhance the performance and output quality of machine learning models [16]. EC methods contribute across the entire ML workflow, including preprocessing activities (such as feature selection and resampling), the learning phase (such as parameter tuning, adjusting membership functions, and optimizing neural network topology), and post-processing tasks (such as pruning of decision trees or support vector models, and improving ensemble approaches). The study presented in [1] elaborates on how EC algorithms can be utilized to resolve a variety of ML-related challenges.

Research in EC has largely focused on mitigating specific challenges encountered in traditional ML tasks. As highlighted in [1], EC algorithms have gained prominence due to their adaptable design components, including representation schemes, search strategies, and fitness evaluation mechanisms. The authors identified nine categories of ML tasks where EC-based techniques have demonstrated significant impact. In evolutionary machine learning (EML), problem representation typically follows three main formats: a vector-based encoding—common to Genetic Algorithms, Particle Swarm Optimization, and Artificial Bee Colony—or a graph-based structure, suitable for Ant Colony Optimization and canonical Genetic Programming.

Furthermore, the authors of [1] summarized key contributions of evolutionary computation across various ML applications. These contributions were grouped into three central components: (i) formulating ML problems as optimization problems through appropriate representations; (ii) determining an effective search mechanism suitable for the target ML task; and (iii) establishing a fitness function that guides the generation of improved candidate solutions.

5.1 Research Articles Reviewed

This section outlines the total number of research articles considered in this comprehensive review. The analysis covers publications from 2003 to 2023 that specifically examined cardiovascular disease prediction using evolutionary machine learning approaches. Articles were sourced from multiple academic databases, and only relevant, full-text studies were downloaded and reviewed. Table 3 presents the publishers whose articles were selected based on the relevance of their contributions to this topic.

5.2 Review Analysis:

This section represents the methodology we have adopted to test Machine Learning models in our environment to observe performance evaluation metrics and the variation therein of results obtained by the above authors.

5.2.1 Methodology for cardiovascular disease risk prediction:

Initially, Installation of Anaconda3 distribution and Jupyter Notebook for using Python 3.9 was done. Later, Cleveland Dataset downloaded from UCI Machine Learning Repository . Feature-Class correlation was calculated, and Correlation Matrix was developed to examine the relation between features and target class. Features and Target class was separated to x and y variable respectively. The data was normalized in order to improve the performance of machine learning algorithms. Data was divided into Train and Test dataset. Then deployment of classification techniques for model training was implemented and finally got the results and confusion matrix was derived and represented into graph and then performance was evaluated.

5.2.2 Data Source:

Dataset of heart disease was downloaded from UCI Machine Learning repository [64] which consists of 303 samples and 13 features, the details of features are shown in Figure. 5 where target feature shows the presence of disease in the patient; 1 mean positive and 0 mean negative.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2
...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2

303 rows × 13 columns

Figure 5. Heart Disease Dataset showing 13 features

5.2.3 Correlation Matrix:

Correlation in machine learning is used for feature selection and mostly used to calculate correlation between features and class to drop one of the two highly correlated features which has same impact on dependent variables [65]. Figure. 6 shows the correlation matrix of feature-class correlation.

Table 3. List of publishers with No. of articles reviewed

Publisher	Number of Articles
ACM Computer Survey	1
Elsevier	4
Hindawi	1
IEEE Access	3
IEEE Explore	10
IEEE Transactions on Evolutionary Computation	1
Journal of Intelligent and Fuzzy Systems	1
MDPI	4
Research Gate	3
Science Direct	4
Springer	1
Springer Link	11
Procedia Technol	1
J. Ambient Intell. Humaniz. Comput	1
Int. J. Comput. Inf. Manuf.	1
Malaysian Journal of Computer Science	1
Internal Journal of Computer Applications	1
IJCTE	1
EAI Endorsed Transactions on Pervasive Health and Technology, 2023	1
Proceedings of the 22nd ACM SIGKDD international conference on knowledge 2016	1
SAE International journal of commercial vehicles, 2018	1
Bulletin of Electrical Engineering and Informatics, 2023	1
American Journal of Artificial Intelligence, 2020	1
International Journal of Advanced Computer Science and Applications, 2023	1

5.2.4 Extraction of Features and Target Class:

Target Feature was extracted from x-variable and copied to y-variable so that the data may be split into train and test accordingly. Figure. 7 shows the dataset without having a target feature.

5.2.5 Normalization:

To increase the performance of Machine Learning Model, the Data was normalized to avoid data redundancy and remove outliers.

5.2.6 Splitting of Data into Train and Test:

The dataset consisted of 303 patients records out of them 242 records of patient were split as trained and 61 records of patients were considered as Test data which made 80% of trained and 20% of test data respectively.

5.2.7 Deployment of Classification techniques for model training:

Machine Learning Models i.e KNN, Decision Tree, Support Vector Machine (SVM), Random Forest, Gaussian Naive Bayes, AdaBoost, Gradient Boost and XGBoost were implemented and following results were obtained as shown in table 4.



Figure 6. Correlation Matrix of feature-class matrix

Table 4. Performance Evaluation Metric obtained after implementing ML Models

Classification Technique	Acc.(%)	Prec.	Recall (Sensitivity)	Spec.	F1 Score	Error Rate
KNN	82	83	85	85	84	0.18
Decision Tree	80	84	79	79	82	0.20
Support Vector Machine	82	81	88	88	85	0.13
Naive Bayes	85	84	91	91	87	0.15
Random Forest	87	91	85	85	88	0.13
AdaBoost	84	90	79	79	84	0.16
Gradient Boost	80	82	82	82	82	0.20
XGBoost	84	85	85	85	85	0.16

5.2.8 Confusion Matrix:

A Confusion Matrix also known as an error matrix. It is a specific table layout used in the field of machine learning to visualize the performance of an algorithm typically in supervised learning tasks [66] as shown below in figure.8

5.2.9 Performance Evaluation:

The performance of the Machine Learning algorithm is measured through performance evaluation metrics i.e Accuracy, Precision, Recall, Specificity, F1 Score and Error rate etc as shown in table 4 obtained in response to the implementation of classification techniques.

5.2.10 Receiver Operating Characteristics (ROC) Curve:

ROC curve is calculated by contrasting the true positive rate (TPR) against the false positive rate (FPR) at various threshold levels and efficiently divides the signal from the noise as shown in figure.9 [67].

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows x 14 columns

Figure 7. Heart Disease Dataset without target feature



Figure 7. XGBoost Confusion Matrix Figure 8. SVM Confusion Matrix Figure 9. RF Confusion Matrix Figure 9. NB Confusion Matrix

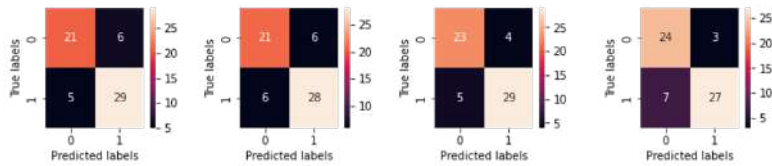


Figure 10. KNN Confusion Matrix Figure 11. GB Confusion Matrix Figure 12. DT Confusion Matrix Figure 13. AB Confusion Matrix

Figure 8. Confusion Matrix

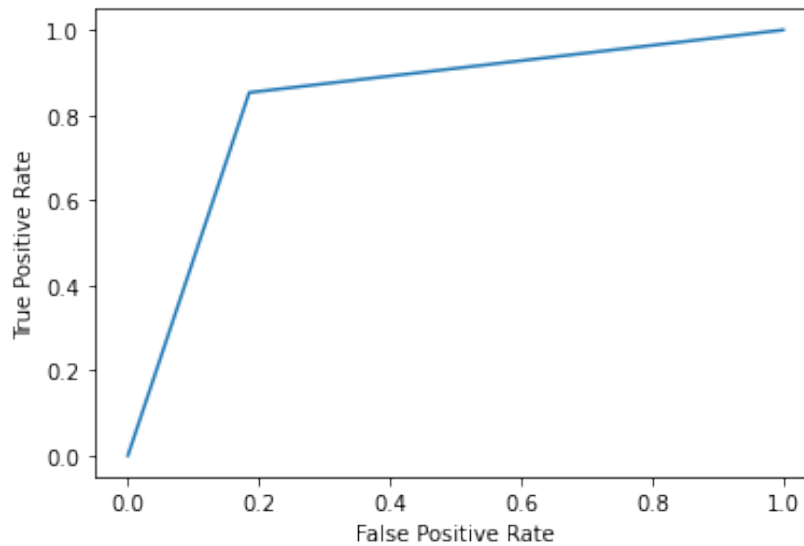


Figure 9. Receiver Operating Characteristics (ROC) Curve

6 Conclusion:

In this study, integration of evolutionary computation in different ML phases is focused. Most of authors used KNN, Random Forest (RF), Support Vector Machine (SVM), Decision Tree, AdaBoost, GradientBoost and XGBoost for prediction and diagnosis of cardiovascular diseases and reviewed the results of performance of machine learning by calculating Accuracy, Precision, Recall, Specificity, F1 Score and Error rate. One of the authors of reviewed papers witnessed the high performance of Model obtained by applying hybrid approach. Finally, our conclusion reflects that hybrid approach can help to minimize features and ultimately increase the accuracy of machine learning model. In future, through parameter tuning of genetic algorithms, accuracy may be increased and can play pivotal role in prediction and disease diagnosis.

Author Contributions

Syed Akbar Ali Shah: Conceptualization, Methodology, Data curation, Writing- Original draft preparation and Visualization **Imtiaz Ali:** Supervision, Investigation and Software implementation **Gulsher Laghari:** Writing- Reviewing and Editing. **Kamran Ali:** Writing- Reviewing and Editing **Ghulam Ali:** Writing- Reviewing and Editing

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

- [1] S. Kanwal, J. Rashid, M. W. Nisar, J. Kim, and A. Hussain, "An Effective Classification Algorithm for Heart Disease Prediction with Genetic Algorithm for Feature Selection," in *Proceedings of the 2021 Mohammad Ali Jinnah University International Conference on Computing, MAJICC 2021, Institute of Electrical and Electronics Engineers Inc., Jul. 2021*. doi: 10.1109/MAJICC53071.2021.9526242.
- [2] Cleveland Clinic, "Cardiovascular Disease." Accessed: Nov. 21, 2023.
- [3] V. Srinadh, B. Maram, J. Mandala, and R. R. Maram, "A Hybrid Machine Learning Model (NB-SVM) for Cardiovascular Disease Prediction," in *2nd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023*, pp. 7–13. doi: 10.1109/ICSCDS56580.2023.10104808.
- [4] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013, doi: 10.1016/j.protcy.2013.12.340.
- [5] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [6] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 7, pp. 8459–8486, 2023, doi: 10.1007/s12652-021-03612-z.
- [7] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Mar. 01, 2022, MDPI*. doi: 10.3390/healthcare10030541.
- [8] K. Y. Tai and J. Dhaliwal, "Machine learning model for malaria risk prediction based on mutation location of large-scale genetic variation data," *Journal of Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00635-x.

- [9] R. Ahmed, M. Bibi, and S. Syed, "Improving Heart Disease Prediction Accuracy Using a Hybrid Machine Learning Approach: A Comparative study of SVM and KNN Algorithms," *International Journal of Computations, Information and Manufacturing (IJCIM)*, vol. 3, no. 1, 2023, doi: 10.54489/ijcim.v3i1.223.
- [10] J. Azmi, M. Arif, M. T. Nafis, M. A. Alam, S. Tanweer, and G. Wang, "A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data," *Medical Engineering and Physics*, vol. 105, Jul. 2022, doi: 10.1016/j.medengphy.2022.103825.
- [11] L. Dwarakanath, A. Kamsin, R. A. Rasheed, A. Anandhan, and L. Shuib, "Automated Machine Learning Approaches for Emergency Response and Coordination via Social Media in the Aftermath of a Disaster: A Review," 2021, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2021.3074819.
- [12] A. Pandey, M. P. Gupta, M. Diwakar, S. Dangi, P. Madan, and P. Singh, "An Effective Machine Learning based Heart Disease Diagnosis Analysis," in *2023 International Conference on Computational Intelligence, Communication Technology and Networking, CICTN 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 536–540. doi: 10.1109/CICTN57981.2023.10141082.
- [13] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 1329–1333. doi: 10.1109/ICICT50816.2021.9358597.
- [14] S. M. Ganie and M. B. Malik, "An ensemble Machine Learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators," *Healthcare Analytics*, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100092.
- [15] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. doi: 10.1007/978-3-030-51517-1_26.
- [16] A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, "Evolutionary Machine Learning: A Survey," 2022. doi: 10.1145/3467477.
- [17] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [18] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES," *Malaysian Journal of Computer Science*, vol. 2022, no. Special Issue 1, pp. 132–148, 2022, doi: 10.22452/mjcs.sp2022no1.10.
- [19] S. M. A. Rahman, S. Ibtisum, E. Bazgir, and T. Barai, "The Significance of Machine Learning in Clinical Disease Diagnosis: A Review," *International Journal of Computer Applications*, vol. 185, no. 36, pp. 10–17, 2023, doi: 10.5120/ijca2023923147.
- [20] S. F. Da Silva, M. X. Ribeiro, J. D. E. S. Batista Neto, C. Traina-Jr., and A. J. M. Traina, "Improving the ranking quality of medical image retrieval using a genetic feature selection method," *Decision Support Systems*, vol. 51, no. 4, 2011, doi: 10.1016/j.dss.2011.01.015.
- [21] P. A. Castillo, M. G. Arenas, J. J. Merelo, V. M. Rivas, and G. Romero, "Multiobjective optimization of ensembles of multi-layer perceptrons for pattern classification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006. doi: 10.1007/11844297_46.
- [22] H. Kwaśnicka and K. Świtalski, "Discovery of association rules from medical data-classical and evolutionary approaches," 2006.

- [23] P. C. H. Ma, K. C. C. Chan, X. Yao, and D. K. Y. Chiu, "An evolutionary clustering algorithm for gene expression microarray data analysis," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 296–314, Jun. 2006, doi: 10.1109/TEVC.2005.859371.
- [24] V. Mangat and R. Vig, "Novel associative classifier based on dynamic adaptive PSO: Application to determining candidates for thoracic surgery," *Expert Systems with Applications*, vol. 41, no. 18, 2014, doi: 10.1016/j.eswa.2014.06.046.
- [25] Q. Song, Y. J. Zheng, Y. Xue, W. G. Sheng, and M. R. Zhao, "An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination," *Neurocomputing*, vol. 226, pp. 16–22, Feb. 2017, doi: 10.1016/j.neucom.2016.11.018.
- [26] C. H. Wu, T. C. Chen, Y. C. Hsieh, and H. L. Tsao, "A hybrid rule mining approach for cardiovascular disease detection in traditional Chinese medicine," in *Journal of Intelligent and Fuzzy Systems*, 2019. doi: 10.3233/JIFS-169864.
- [27] C. To and T. D. Pham, "Analysis of Cardiac Imaging Data using Decision Tree based Parallel Genetic Programming."
- [28] G. Chugh, S. Kumar, and N. Singh, "Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis," *Cognitive Computation*, vol. 13, no. 6, 2021, doi: 10.1007/s12559-020-09813-6.
- [29] "Predicting Heart Disease Risk Factors using Hawkins Dataset and Decision Tree Algorithms," *International Journal of Mechanical Engineering*, 2023, doi: 10.56452/6-3-674.
- [30] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Mar. 01, 2022, MDPI*. doi: 10.3390/healthcare10030541.
- [31] SVS College of Engineering and Institute of Electrical and Electronics Engineers, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms,"
- [32] P. Singh, G. K. Pal, and S. Gangwar, "Prediction of Cardiovascular Disease Using Feature Selection Techniques," *International Journal of Computer Theory and Engineering*, vol. 14, no. 3, pp. 97–103, Aug. 2022, doi: 10.7763/IJCTE.2022.V14.1316.
- [33] T. H. Choudhury and B. Choudhury, "Automated Cardiovascular Disease Prediction Models: A Comparative Analysis," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 9, no. 1, 2023, doi: 10.4108/eetpht.8.3402.
- [34] S. M. Ganie, P. K. D. Pramanik, M. B. Malik, A. Nayyar, and K. S. Kwak, "An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms," *Computer Systems Science and Engineering*, vol. 46, no. 3, pp. 3993–4006, 2023, doi: 10.32604/csse.2023.035244.
- [35] N. Louridi, S. Douzi, and B. El Ouahidi, "Machine learning-based identification of patients with a cardiovascular defect," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00524-9.
- [36] A. García-Domínguez et al., "Diabetes Detection Models in Mexican Patients by Combining Machine Learning Algorithms and Feature Selection Techniques for Clinical and Paraclinical Attributes: A Comparative Evaluation," *Journal of Diabetes Research*, vol. 2023, 2023, doi: 10.1155/2023/9713905.
- [37] J. Yang and J. Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Information (Switzerland)*, vol. 13, no. 10, Oct. 2022, doi: 10.3390/info13100475.
- [38] T. Chen, ... C. G. sigkdd international conference on knowledge, and undefined 2016, "Xgboost: A scalable tree boosting system," *dl.acm.org*T Chen, C GuestrinProceedings of the 22nd acm sigkdd international conference on knowledge, 2016•dl.acm.org, vol. 13-17-Aug, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.
- [39] S. Yakovlev, C. M. Bhatt, P. Patel, T. Ghetia, and P. Luigi Mazzeo, "Effective heart disease prediction using machine learning techniques," *mdpi.com*CM Bhatt, P Patel, T Ghetia, PL MazzeoAlgorithms, 2023•mdpi.com, 2023, doi: 10.3390/a16020088.

- [40] A. E. Eiben and J. E. Smith, "Introduction," 2003, pp. 1–14. doi: 10.1007/978-3-662-05094-1_1.
- [41] M. Govindaraj, V. Asha, B. Saju, M. Sagar, and Rahul, "Machine Learning Algorithms for Disease Prediction Analysis," in Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, 2023. doi: 10.1109/ICSSIT55814.2023.10060987.
- [42] S. Osama, H. Shaban, and A. A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review," 2023. doi: 10.1016/j.eswa.2022.118946.
- [43] P. C. H. Ma, K. C. C. Chan, X. Yao, and D. K. Y. Chiu, "An evolutionary clustering algorithm for gene expression microarray data analysis," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 296–314, Jun. 2006, doi: 10.1109/TEVC.2005.859371.
- [44] A. Zafar et al., "Machine learning-based risk factor analysis and prevalence prediction of intestinal parasitic infections using epidemiological survey data," *PLoS Neglected Tropical Diseases*, vol. 16, no. 6, 2022, doi: 10.1371/JOURNAL.PNTD.0010517.
- [45] N. Louridi, S. Douzi, and B. El Ouahidi, "Machine learning-based identification of patients with a cardiovascular defect," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00524-9.
- [46] K. P. Kresoja, M. Unterhuber, R. Wachter, H. Thiele, and P. Lurz, "A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction," Dec. 01, 2023, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s00395-023-00982-7.
- [47] N. Gupta, D. Gupta, A. Khanna, P. P. Rebouças Filho, and V. H. C. de Albuquerque, "Evolutionary algorithms for automatic lung disease detection," *Measurement: Journal of the International Measurement Confederation*, vol. 140, pp. 590–608, Jul. 2019, doi: 10.1016/j.measurement.2019.02.042.
- [48] A. E. Eiben and J. E. Smith, "Introduction," 2003, pp. 1–14. doi: 10.1007/978-3-662-05094-1_1.
- [49] S. Kanwal, J. Rashid, M. W. Nisar, J. Kim, and A. Hussain, "An Effective Classification Algorithm for Heart Disease Prediction with Genetic Algorithm for Feature Selection," in Proceedings of the 2021 Mohammad Ali Jinnah University International Conference on Computing, MAJICC 2021, Institute of Electrical and Electronics Engineers Inc., Jul. 2021. doi: 10.1109/MAJICC53071.2021.9526242.
- [50] A. A. Moiz et al., "A Machine Learning-Genetic Algorithm (ML-GA) Approach for Rapid Optimization Using High-Performance Computing," *SAE International Journal of Commercial Vehicles*, vol. 11, no. 5, 2018, doi: 10.4271/2018-01-0190.
- [51] K. Kangra and J. Singh, "Comparative analysis of predictive machine learning algorithms for diabetes mellitus," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1728–1737, Jun. 2023, doi: 10.11591/eei.v12i3.4412.
- [52] N. S. Nagaraj M. Lutimath, H. V. Ramachandra, S. Raghav, "Prediction of Heart Disease Using Genetic Algorithm," in Proceedings of Second Doctoral Symposium on Computational Intelligence, P. Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Ed., Springer, 2021, pp. 66–75.
- [53] P. R. L. S. V. Jinny, and Y. V. Mate, "Early prediction model for coronary heart disease using genetic algorithms, hyperparameter optimization and machine learning techniques," *Health and Technology*, vol. 11, no. 1, pp. 63–73, 2021, doi: 10.1007/s12553-020-00508-4.
- [54] A. Durga and D. M. Phil, "Enhanced Prediction of Heart Disease by Genetic Algorithm and RBF Network," 2015.
- [55] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020. doi: 10.1007/978-3-030-51517-1_26.

- [56] L. R. Guarneros-Nolasco, N. A. Cruz-Ramos, G. Alor-Hernández, L. Rodríguez-Mazahua, and J. L. Sánchez-Cervantes, "Identifying the main risk factors for cardiovascular diseases prediction using machine learning algorithms," *Mathematics*, vol. 9, no. 20, Oct. 2021, doi: 10.3390/math9202537.
- [57] Navita et al., "Advanced Hybrid Machine Learning Model for Accurate Detection of Cardiovascular Disease," *International Journal of Computational Intelligence Systems*, vol. 18, no. 1, Dec. 2025, doi: 10.1007/s44196-025-00771-1.
- [58] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [59] R. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989, doi: 10.1016/0002-9149(89)90524-9.
- [60] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES," *Malaysian Journal of Computer Science*, vol. 2022, no. Special Issue 1, 2022, doi: 10.22452/mjcs.sp2022no1.10.
- [61] S. Bilgaiyan, T. I. Ayon, A. A. Khan, F. T. Johora, M. Parvin, and M. J. Alam, "Heart disease Prediction Using Machine Learning," in *2023 International Conference on Computer Communication and Informatics, ICCCI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICCCI56745.2023.10128378.
- [62] SVS College of Engineering and Institute of Electrical and Electronics Engineers, *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies: 01-03, March 2018*.
- [63] L. Yahaya, N. David Oye, and E. Joshua Garba, "A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques," *American Journal of Artificial Intelligence*, vol. 4, no. 1, p. 20, 2020, doi: 10.11648/j.ajai.20200401.12.
- [64] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "UCI Machine Learning Repository: Heart Disease Data Set," *Uci*, 1988.
- [65] A. Pandey, M. P. Gupta, M. Diwakar, S. Dangi, P. Madan, and P. Singh, "An Effective Machine Learning based Heart Disease Diagnosis Analysis," in *2023 International Conference on Computational Intelligence, Communication Technology and Networking, CICTN 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 536–540. doi: 10.1109/CICTN57981.2023.10141082.
- [66] "What is a Confusion Matrix in Machine Learning - MachineLearningMastery.com." Accessed: Mar. 29, 2024. [Online]. Available: <https://machinelearningmastery.com/confusion-matrix-machine-learning/>
- [67] K. Kangra and J. Singh, "Comparative analysis of predictive machine learning algorithms for diabetes mellitus," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1728–1737, Jun. 2023, doi: 10.11591/eei.v12i3.4412