

# A COMPREHENSIVE REVIEW ON MACHINE LEARNING BASED HEART DISEASE DIAGNOSIS

Anas Mohammad Saeed Idlibi

Ajman University of Science & Technology

## KeyWords

Cardiovascular Disease, Machine Learning, random forest, feature selection, Non-Invasive Methods, Prediction, Classification

## ABSTRACT

Diseases of the cardiovascular system take the top spot on the list of the world's leading causes of mortality. However, accurate prognostication and prompt diagnosis may assist in lowering the overall death rate and increasing the percentage of individuals who are able to survive their condition. One key problem that slows down the process of early diagnosis is the lack of availability or shortage of radiologists and other medical professionals in various nations for a variety of reasons. In the realm of medical imaging, the use of computational intelligence as a method to identify, prognosticate, and diagnose illness is a burgeoning trend that is one of the many different efforts being made to construct decision support systems. It makes it easier for radiologists and other medical professionals to deal with their excessive workloads and reduces the amount of time lost due to unnecessary delays in the diagnosis of patients. Despite the fact that a significant number of cardiac disorders are avoidable, their prevalence is nevertheless on the increase, mostly as a result of insufficient preventative efforts. In the modern, digital age, a number of clinical decision support systems for predicting cardiac disease have been created by a variety of researchers in order to simplify diagnosis and guarantee that it is carried out effectively. Using data mining and machine learning techniques, the authors of this work analyze the current state of the art regarding a variety of clinical decision support systems that have been presented by different researchers for the prediction of cardiac disease. It has been discovered that Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are superior in the majority of the research in all of the frameworks. Deep neural networks are a relatively newer technique of machine learning that is producing noticeable results in classifying heart sound signals and cardiovascular imaging. These findings have been achieved in recent years. The current study will assist in automating the process of diagnosing heart disease by providing new researchers in the field of machine learning with recommendations and potential new research topics to investigate.

## 1. Introduction

Illnesses that are contagious, such as the swine flu and the corona virus, may spread quickly; nevertheless, fatality rates for communicable diseases are often lower when compared to those of numerous non-communicable diseases, such as diabetes, heart disease, liver cancer, and breast cancer. Researchers in the field of machine learning place a significant emphasis on medical data mining. These researchers are interested in a wide range of diseases, including breast cancer, heart disease, diabetes, Parkinson's disease, hepatitis, liver disorder, lung cancer, pancreatic cancer, leukemia, and brain tumors [1-10]. Out of all of these ailments, heart disease is the most unexpected and potentially fatal one. When compared to other non-communicable diseases, it has the greatest fatality rate. In 2015, various cardiac illnesses claimed the lives of more than 17.7 million individuals worldwide [11]. According to research that was published not too long ago [12], heart disease is responsible for the deaths of 17.9 million individuals each and every year all over the globe. Using this information as a proxy for a raw figure, we may estimate that around 49 000 persons throughout the globe pass away owing to cardiac problems every single day. This is a very large amount. The difficult and expensive diagnostic procedure contributes to the higher death rates caused by heart disease in impoverished and emerging countries.

According to the World Health Organization (WHO), cardiovascular disease (CVD) is one of the most common causes of mortality in the United States and is one of the leading causes of death globally [13], [14]. Cardiovascular disease (CVD) is a disorder that occurs when the heart is not working adequately to pump the needed volume of blood to other regions of the body, which may lead to heart failure [15]. The most common cause of heart failure is the narrowing or blocking of coronary arteries. Early symptoms of cardiovascular disease include an irregular heartbeat, chest pain or discomfort, shortness of breath, swelling feet or ankles, exhaustion, and fainting. Other symptoms include chest pain or discomfort. The earlier a patient's condition can be diagnosed and treated, the longer they are likely to live. There is a severe problem that is a bottleneck in this respect, and that is the lack of resources and the unavailability of physicians and radiologists in developing or low-income nations. This causes sickness to be diagnosed at an advanced stage when it might have been prevented. This is one of the primary reasons why the average length of life for individuals diagnosed with cardiovascular disease is between one and two years [16]. The patient's medical history, age, gender, and lifestyle, among other things, may all be variables in the development of cardiovascular disease (CVD). Alterations in lifestyle, such as increased movement and abstinence from tobacco use, may help lessen the impact of risk factors by bringing cholesterol and blood pressure levels under control. These potential dangers are currently being evaluated by a medical professional. The analysis of early symptoms, adjustments in lifestyle, and the medical examination report all assist medical professionals diagnose the illness. As a result of human error and the fact that the expertise utilized to evaluate patient data might vary depending on the amount of knowledge, accuracy and prognostication cannot be guaranteed [17-18]. A heart attack or heart failure are two common ways to describe issues related to the cardiovascular system. Congenital heart disease, left sided heart failure, right sided heart failure, ischemic heart disease (IHD), myocardial infarction, arrhythmias, systemic and pulmonary hypertensive heart disease, valvular heart disease, infective endocarditis and non-infective vegetation, cardiomyopathies and myocarditis, pericardial disease, and pericardiac tumors [19] are the most common types of heart conditions. Congenital heart disease [20], which is often diagnosed in newborn newborns and has a low death rate, is one of the conditions that researchers that study machine learning pay the least attention to diagnosing. The phrase cardiovascular illnesses encompass any other issues pertaining to the circulatory system (CVDs). The extensive body of research on the use of machine learning to diagnose cardiovascular diseases includes the diagnosis of ischemic heart disease (IHD) [21], arrhythmia [22], and valvular heart disease [23-25]. The fact that conditions related to the heart are more likely to result in death explains why there is such a wealth of published information on these disorders.

Angiograms are one of the potential diagnostic tools for determining the extent of cardiovascular disease (CVD). Angiography may have some unpleasant side effects, and it also requires a high degree of competence from the patient, which is one of the primary reasons why researchers are looking at automated solutions that might assist ease the diagnostic procedure. In the previous several decades, researchers and academics have been looking forward to the development of such automated machine intelligent expert systems to lessen the related risks of medical examination. The machine learning (ML) predictive models or their ensembled versions, such as logistic regression (LR), k-nearest neighbor (kNN), support vector machine (SVM), naive bayes (NB), adaboost (AB), K-means clustering, linear regression, decision trees (DTs), and random forest, can be utilized by the decision support systems (RF). Rule-based fuzzy logic, often known as RBFL, and the opposing firefly with BAT technique are what Reddy and Khare [26] employ in order to make a prediction about heart disease. On the basis of rule-based fuzzy logic (RBFL) and the oppositional BAT firefly (OFBAT) method, a hybrid approach for detecting heart disease known as OFBAT-RBFL has also been developed. In addition, Arabasadi et al. [27] proposed a neural network-based hybrid neural network technique combined with a genetic algorithm for accurately classifying heart disease datasets.

The majority of fatalities caused by all of the various types of cardiovascular diseases are attributable to coronary artery disease (CAD). Myocardial infarction is the medical term for a heart attack, which is the outcome of coronary artery disease (CAD). CAD occurs when a material known as atherosclerotic plaque forms, which limits the blood flow to the heart and ultimately causes a heart attack (MI). Therefore, it is essential that an individual should notice the early creation of atherosclerotic plaque by means of some clinical measurement, and the formation must be prevented by means of some particular medical therapy. The diagnosis of cardiac illness may be carried out in one of two ways: using invasive procedures, or (ii) using non-invasive methods [28]. The non-invasive approach known as coronary angiography (CA), often considered the gold standard for the identification of heart disease, is known as coronary angiography. On the other hand, it is very expensive, difficult to use, and calls for specialized knowledge. The invasive procedure comes with a number of risks, including the possibility of arterial dissection and arrhythmia. In certain people, it may lead to issues with the kidneys as well as paralysis. Under severe conditions, CA has the potential to possibly cause death [28], [29]. In addition, continual imaging and screening are essential in CA, which contributes to the state's high operating expenses. Once again, significant costs associated with diagnosis make this an impractical option, particularly in underdeveloped and emerging countries. Therefore, despite the fact that it is considered the gold standard, it is not universally acceptable, and patients from underdeveloped countries such as India will often avoid it. The automatic diagnosis of cardiovascular diseases is an essential application field for ma-

chine learning, and it eventually helps save many lives. This is because the majority of heart diseases may be treated successfully provided, they are diagnosed in a timely manner. This effort will foster the development of clinical decision support systems, which are decision support systems that are based on models of machine learning and are utilized in diagnosis. These systems are often referred to as CDSS. Not only will heart patients benefit from CDSS, but doctors and even the government might find it to be a valuable tool.

The human heart is considered to be one of the most important organs in the body. It is a kind of muscle organ that is located at the center of the body's circulatory system and is responsible for pumping blood into the body [42]. The cardiovascular system is made up of all of the blood vessels in the body, including arteries, veins, and capillaries, which together create a complicated network that allows blood to circulate throughout the whole body [43]. Any restriction or abnormality in the normal circulation or flow of blood from the heart would result in a number of serious complications of cardiac disease. These problems might range from mild to life-threatening. These conditions are referred to as cardiovascular diseases (CVDs), and they are among the most lethal illnesses that may be found everywhere on the globe. Disorders affecting the heart, blood vessels in the brain, and diseases of the blood vessels themselves are all included in the category of cardiovascular diseases [44]. According to the report "Global Atlas on Cardiovascular Disease Prevention and Control" published by the World Health Organization (WHO), cardiovascular diseases (CVDs) are the main causes of mortality and disability across the world [45]. Despite the fact that CVDs may be avoided by making adjustments to one's way of life and taking other associated preventative steps, all signs point to the fact that these conditions are still on the increase on a daily basis, as mentioned in a number of publications published by the WHO. The World Health Organization (WHO) has published many papers that reveal an alarming increase in the prevalence of cardiovascular diseases around the globe. Around the globe, cardiovascular diseases are the leading cause of death, accounting for an estimated 17.5 million deaths in 2012 [46]. Another research from the WHO estimates that cardiovascular diseases are responsible for 17.9 million deaths annually globally, which accounts for around 31 percent of all fatalities. Eighty-five percent of these occur as a result of heart attacks and strokes [47]. The World Health Organization (WHO) has published a number of reports that indicate an increase in the number of deaths caused by cardiovascular diseases. Despite the increasing prevalence of risk factors, the majority of experts believe this trend can be attributed to a lack of effective preventative measures.

A number of the medical disorders described above have treatment options available. However, the potential for optimum benefit can not be realized unless patients have sufficient adherence to their prescription regimens. According to the World Health Organization, medical adherence (MA) is defined as "the degree to which a person's behavior—taking medicine, following a diet, and/or implementing lifestyle changes—corresponds with approved recommendations from a health care professional" [31]. The estimated MA among patients using CVD preventative medications is approximately 57 percent (95 percent confidence interval: 50–64 percent), with reported adherence in the included studies ranging from 32 percent to 68 percent [30]. This estimate is based on a meta-analysis that included 376,162 patients. According to the findings of Naderi et al. [30], it is possible that the kind of drug and the potential adverse effects of pharmaceuticals are not the primary reasons for poor adherence to CVD preventive medications. It would seem that the involvement of variables that are not connected to drugs has a more significant impact instead [32]. A variety of regression approaches have been used in a number of studies [33,34,63] to discover MA related variables. The prediction of future MA has not seen a discernible shift despite the enhancements that have been made to databases that include health care claims and the rise in the amount of patient data. Most of the attention in traditional techniques was concentrated on a limited set of clinical criteria that were defined by investigators as well as demographic data that were accessible when therapy first began. Predictors of multiple sclerosis include a diverse set of factors, such as age, race, socioeconomic status, psychological issues, cognitive problems, and the complexity of the treatment protocol [35-37]. Some of these factors include age, race, socioeconomic status, psychological issues, cognitive problems, and the complexity of the treatment protocol. Even though these characteristics have a correlation with adherence, they are not able to differentiate very well between patients who adhere to their treatment and those who do not [38,39]. As a consequence of this, the most accurate adherence-prediction models, which are those that are based on standard statistical approaches, yield at best a modest accuracy of 0.65 [40]. Because of advancements in the processing power, memory, and storage capacity of computers, it is now feasible to carry out more complex learning tasks using machine learning (ML) and deep learning approaches. The use of variable selection algorithms and the analysis of nonlinear associations between predictors and outcomes, as well as deep interactions among predictors, have both contributed to the improvement in prediction made possible by the new methods. For example, the selection of essential variables does not need to adhere to any particular criterion or undergo any linearity checks, both of which might be helpful for modeling a behavior as complicated as MA [41].

## 2. Related Work

The author of [48] conducted a research that compared the ANN classification algorithm with the SVM classification method, bas-

ing their findings on the Positive Predictive value (PPV) of cardiovascular illnesses. Their data came from a selection of three hospitals that were all associated with the AJA University of Medical Sciences in Iran. There are a total of 1324 occurrences and 25 characteristics included in the sample. The sample consists of medical records of coronary artery disease patients who were admitted to one of the three hospitals indicated between March 2016 and March 2017, and the sample spans the time period from March 2016 to March 2017. The variables described in the Cleveland heart disease data policy in the UCI machine learning repository served as the basis for the collection of the data. Several procedures, including data preparation, integration, cleaning, normalization, and reduction, were used in order to exercise control over the data that were obtained. After the data were imported into SPSS (version 23.0) and Microsoft Excel 2013, statistical computation was carried out using R 3.3.2. The sample was cut such that 70 percent would be used for training the algorithm and 30 percent would be used for testing it. The findings of their studies demonstrated that the SVM method exhibited more accuracy and better performance than the ANN model, and it was distinguished by higher power and sensitivity. [Citation needed] [Experimental results] The author of [49] developed a framework for the prediction of cardiac illness that was based on the RF method in machine learning and used Python. For the training and testing of the algorithm, they used the Cleveland heart disease datasets that were acquired from the machine learning repository at the University of California, Irvine. This sample was comprised of 303 instances and 76 features when it was first collected; however, after preprocessing and the hand selection of attribute features, only 9 features were used. Seventy-five percent of the sample was used for the training of the algorithm, while the remaining twenty-five percent was put to use for testing. Visual Studio Code was used to create a graphical user interface (GUI), which was then applied to the task of visualizing the trials. During the classification process, an RF classifier was used, and as a result, an accuracy of 97.56 percent was achieved. The diagnosis of heart disease was broken down into four (4) phases depending on the percentage of arterial blockage; the existence of heart disease is indicated when the percentage of artery blockage is more than fifty percent. Because fifty percent of artery blockage is still considered normal or absent in patients without heart disease, this model was not able to identify heart disease in its early stages. The author also conducted a comparison study on the prediction of heart disease using support vector machine, decision tree, and k-nearest neighbor algorithms and presented their findings in [50]. For the purposes of both training and testing the system, they made use of the VA Long Beach dataset that they had got from the UCI machine learning repository. This dataset has 270 instances and 12 characteristics. The accuracy, sensitivity, and specificity of the model were examined using a confusion matrix for the purpose of evaluation. Support Vector Machine (SVM) fared better than KNN and DT in categorizing heart disease patients, with an accuracy of 92 percent, sensitivity of 100 percent, and specificity of 83 percent, according to the findings of their experimental work.

Using a Multi-Layer Perceptron Neural network (MLP) with back propagation as the training technique, the author of [51] presented a diagnostic system for predicting cardiac illness. This system would use back propagation. Sensitivity, specificity, precision, and accuracy were the metrics that were used to assess the produced system's overall performance. For the purposes of both model training and testing, the Cleveland data from the UCI machine learning repository was used. This data set includes 303 occurrences and 76 characteristics. A preprocessing operation was done on the data, and it removed the six cases that had missing values. Only 14 of the 76 characteristics were considered to have any significant bearing on heart disease. The MLP-NN suggested model showed a high accuracy of 93.39 percent for 5 neurons in the hidden layer with a running duration of 3.86 seconds in the prediction of heart disease. These results are based on the tests that were carried out. The author of [52] introduced a machine learning-based method for detecting heart disease utilizing sampling methods to handle imbalanced datasets. This method was successful in identifying a significant number of cases. Adaptive Synthetic Sampling Approach, Random Over-Sampling, and Synthetic Minority Over-Sampling (SMOTE) are some of the sampling strategies that were used in this study (ADASYN). For the purposes of both the training and testing of the system, Framingham datasets downloaded from the Kaggle website were used. These datasets include 4239 occurrences and 15 characteristics. The objective was to determine, on the basis of these characteristics, whether or not a patient had a 10-year risk of developing coronary heart disease in the future. LR, KNN, AdaBoost, DT, NB, and RF are some of the machine learning algorithms that were used here. These classification algorithms' performances were analyzed by measuring their precision, recall, and accuracy, and the results were quantified and compared. Each of these factors might change depending on the kind of sampling method that was carried out. According to the findings of their experiments, the SVM classifier combined with the Random OverSampling approach performed the best in the prediction of heart disease, with an accuracy of 99.99 percent. However, RF did better with the SMOTE approach, achieving a level of accuracy of 91.3%, while both DT classifier and RF fared better with the ADASYN technique, achieving a level of accuracy of 90.3%. Therefore, the classification accuracy of this strategy was mainly relied on the sampling procedures, which are not always required in all kinds of datasets. This is because sampling techniques are not always essential.

The author of [53] constructed a model for predicting individuals who will be diagnosed with coronary artery disorders by employing a combination of data mining approaches that were descriptive and predictive (CAD). We utilized datasets received from a medical center that had a total of 282 patients and 58 different characteristics. The data were preprocessed to get rid of any missing values and any extreme values that stood out. The K-means algorithm was selected as the clustering method (descriptive), while several

classification algorithms, such as CHAID, Quest, C5.0, C & RT-DT, and ANN were chosen as the prediction techniques. K-means was selected as the clustering approach (descriptive). When all of the datasets were considered, the results of their experiments showed that the C & RTDT algorithm seemed to be the most accurate in predicting CAD, with an error rate of just 0.074. Nevertheless, the findings that were obtained for each of the three groups were distinct. In clusters 1 and 2, the performance of C & RT-DT was superior, with error rates of 0.022 and 0.023 respectively. During the time spent in cluster 3, the CHAID method seemed to be the best-performing classifier with zero errors. The accuracy of the forecast is dependent on the characteristics of the heart disease as well as other variables; nevertheless, having an excessive number of characteristics, such as the one shown here (58), may lead to incorrect predictions. The author of [54] conducted a comparison analysis on several data mining categorization methods, including some of the most often used ones. They are the K-Nearest neighbor (KNN) algorithm, the Support Vector Machine (SVM), and the Artificial Neural Network (ANN) algorithm utilizing MATLAB multilayered feed forward back propagation. The UCI machine learning repository provided the data for this study, which consisted of 303 instances of Cleveland heart disease data and 76 characteristics. After doing data cleaning to eliminate records that included a large number of missing values, the size of the data set was reduced to 270 instances with just 13 characteristics. The first half of the data was utilized to train the models, while the remaining half was used for evaluating the models. The results of their experiments indicated that SVM performed better than both KNN and ANN in terms of accuracy of classification, with SVM coming out on top with 85 percent, while KNN fared slightly better than ANN with 82 percent, which was about 73 percent.

A Neural Network-based prediction of coronary heart disease risk using feature correlation analysis (NN-FCA) was developed by the author in [55]. This method has two stages: feature selection and feature correlation analysis. The KNHANES-V1 dataset was chosen in the first phase of the system process, and statistical analysis was carried out in the second step to determine the characteristics that are associated with an increased risk of coronary heart disease. The third phase was the identification of predictors of the risk of coronary heart disease using a feature sensitivity-based feature selection. In the fourth phase, Neural Network (NN)-based coronary heart disease risk predictors were developed using feature correlation analysis of features. These predictors were then tested. The fourth phase was the validation of NN-based coronary heart disease risk estimates via the use of feature correlation analysis. Performance metrics were taken. The KNHANES-V1 study, which was carried out by the Korean Center for Disease Control and Prevention, was used to collect the datasets. The total number of cases in the sample is 8108, however ambiguity led to the elimination of 3324 of those instances. And 630 of those cases involved people younger than 30 years old. As a consequence, there were 4146 occurrences of coronary heart disease in the sample that was generated. Age, gender, cholesterol levels, blood pressure, and a variety of other relevant characteristics were included as training variables for the model. High blood pressure, abnormal cholesterol levels, cardiovascular disease, including myocardial infarction, and angina were the output variables. When these five risk factors are absent, the likelihood of coronary heart disease is minimal. However, coronary heart disease is at an increased risk whenever one of these five risk factors is present. The statistical analysis was carried out using the most recent version of IBM SPSS (22.0). The performance of the classifiers was compared using the Confusion Matrix and the Receiver Operating Characteristics (ROC). According to the findings of the experiments, the NN-FCA model was just as accurate in predicting coronary heart disease as the FRS model. When compared to the validation of the FRS for the Korean population, the ROC curve produced by the NN-FCA model was much larger, and the coronary heart disease risk prediction it produced was significantly more accurate. Other relevant indications, such as diabetes, cholesterol level, and others, that could be a sign of high risk of heart disease but are not added into the model may lead to late prediction of the illness's presence. Fuzzy C Means classifier was used in the model that was reported in reference [56] to estimate the risk of patients having a heart attack. The model was given the title "Heart Attack Prediction using Data mining Techniques." The Fuzzy C means is a machine learning clustering approach that is unsupervised and allows for one piece of data to belong to two or more clusters at the same time. The machine learning repository at the University of California, Irvine (UCI) was the source of the datasets that were used in the training and testing of the model. There are a total of 270 cases and 73 heart disease characteristics included in the sample size; however, only 13 of them were used for the purpose of heart attack prediction. A preprocessing step was taken in order to get rid of any missing data values. The outcomes of the classification experiments that were carried out revealed that the suggested classifier, known as Fuzzy C Means, obtained higher levels of accuracy than the majority of the other classification algorithms that are currently in use.

Author studied the case of predicting the risk of cardiovascular diseases (CVDs) by comparing Auto Machine Learning techniques against a graduate student using several important metrics, including the total amount of time required for building machine learning models and the final classification accuracies on unseen test datasets. In [57], author studied the case of predicting the risk of CVDs by comparing Auto Machine Learning techniques against a graduate student. They presented a model known as Auto-SKlearn, which was inspired by Scikit-Learn, one of the most well-known generic machine learning toolboxes. Their model makes extensive use of the Scikit-learn toolbox's machine learning classifiers and the preprocessing procedures it provides. LR, SVM, RF, Boosting, NN, and KNN are some of the classifiers that are used here. Auto-SKlearn begins by selecting a suitable set of data preparation processes,

such as the imputation of missing values, based on the training data. After that, it transfers the processed data to a feature processing block, which further normalizes the data or decreases the dimensions of the data using common procedures, principal component analysis. At last, the datasets are sent on to the estimator block, which chooses and trains machine learning algorithms to predict desired outputs from input data samples. This is done by passing the data samples on to the block. Training, testing, and assessment were carried out on two separate datasets each pertaining to cardiovascular disease. The datasets on Cleveland heart disease were obtained from the machine learning repository at the University of California, Irvine. These datasets had 303 cases and 76 attributes, but researchers only utilized 13 of them. The other datasets were CVD data, and they had 70,000 cases and 11 characteristics. The source of the data was not identified. For the sake of the studies, each category of the dataset was divided into three parts: the training set, the validation set, and the testing set. One hundred of the 303 instances of the UCI data were used for testing, while the remaining 203 were used for training and validation. The Kfold cross validation technique was selected as the one to use. 14,000 of the 70,000 records of the CVD data were used for testing, while the remaining 56,000 records were used for training and cross validation. It took the graduate student a significant amount of time (432 hours for UCI data and 360 hours for CVD datasets) to develop comparable classifiers, but the Auto-ML model only required 30 minutes to build a competitive classifier for each dataset. This is in contrast to the amount of time it took the graduate student to develop similar classifiers. It takes the Auto-ML model an unacceptable amount of time, which is thirty minutes, to construct a classifier. Using NB and DT algorithms written in Python, the author of [58] suggested a method for predicting cardiac disease that was based on machine learning approaches. The datasets that included the 13 heart disease characteristics were downloaded from the Kaggle website and utilized for the training and testing of the model. For the simulation, an additional dataset was taken from the machine learning repository at UCI. The environment of Sci-py was used to test the suggested model when it was developed. When it came to the prediction of cardiac diseases, the results of their studies indicated that the DT algorithm performed much better than the NB. Their research suffered from a number of flaws, the most notable of which were insufficiently described datasets, a lack of access to actual tests, hazy results, and an incorrect method for selecting features.

A framework for the prediction of cardiac disease was provided by the author in [59]. This framework was given the name "Hybridization," and it incorporated multiple different machine learning algorithms into a single model. During the procedures of model training and testing, the Cleveland datasets that can be found in the online machine learning repository of the UCI were used. These datasets are comprised of 303 occurrences and 14 characteristics. Preprocessing of the data was done in order to minimize the number of characteristics from 14 to 12. In the process of predicting heart disease, a wide variety of classification algorithms were used, including NB, SVM, KNN, NN, J48, RF, and GA. These algorithms' levels of accuracy, sensitivities, and specificities were taken into consideration. They were executed in succession on the same dataset and set of characteristics that were being used. The findings of the studies showed that NB and SVM performed better in the prediction of heart disease with an accuracy of 89.2 percent. The author of the paper [60] built a machine learning-based hybrid intelligent system framework for the identification of heart disease patients by using seven of the most common classification algorithms and the programming language Python. KNN, ANN, DT, SVM, NB, LR, and MLP are all part of this group. For the purposes of both model training and testing, the Cleveland dataset was used. This dataset has 303 occurrences and 76 characteristics. On the data, they used a cross validation method with 10 different iterations. The best heart disease connected characteristics were chosen with the use of feature selection algorithms such as Relief, MinimalRedundancy-Maximal-Relevance (mRMR), and Least Absolute Shrinkage and Selection Operator (LASSO). The data were preprocessed to get rid of the occurrences that had significant amounts of data missing. This brought the total number of instances down to 297, and there were only 14 features used. After using the feature selection algorithms, the characteristics were narrowed down to only six that were connected to heart disease. In order to find the model that had the highest overall performance, they put each of the classifiers through their paces using any of the available feature selection strategies. The results of their experiments indicated that the SVM combined with the LASSO feature selection technique seemed to be the combination that performed the best when compared to the various other feature selection algorithms and classifiers. If the characteristics of heart disease were reduced to only six, the resulting classification accuracy would be incorrect since many important characteristics would be left out.

### 3. Methodology

Literature suggests three workflows for machine learning-based heart disease detection. Machine Learning Framework-A, Framework-B, and Framework-C [61] are the 3 workflows. Categorization depending on ML data types.

#### 3.1 Different ways to diagnose heart disease and frameworks for machine learning

Figure 1 displays the different ways that heart diseases can be diagnosed. In contrast to CA, NI has a number of ways to diagnose heart disease.

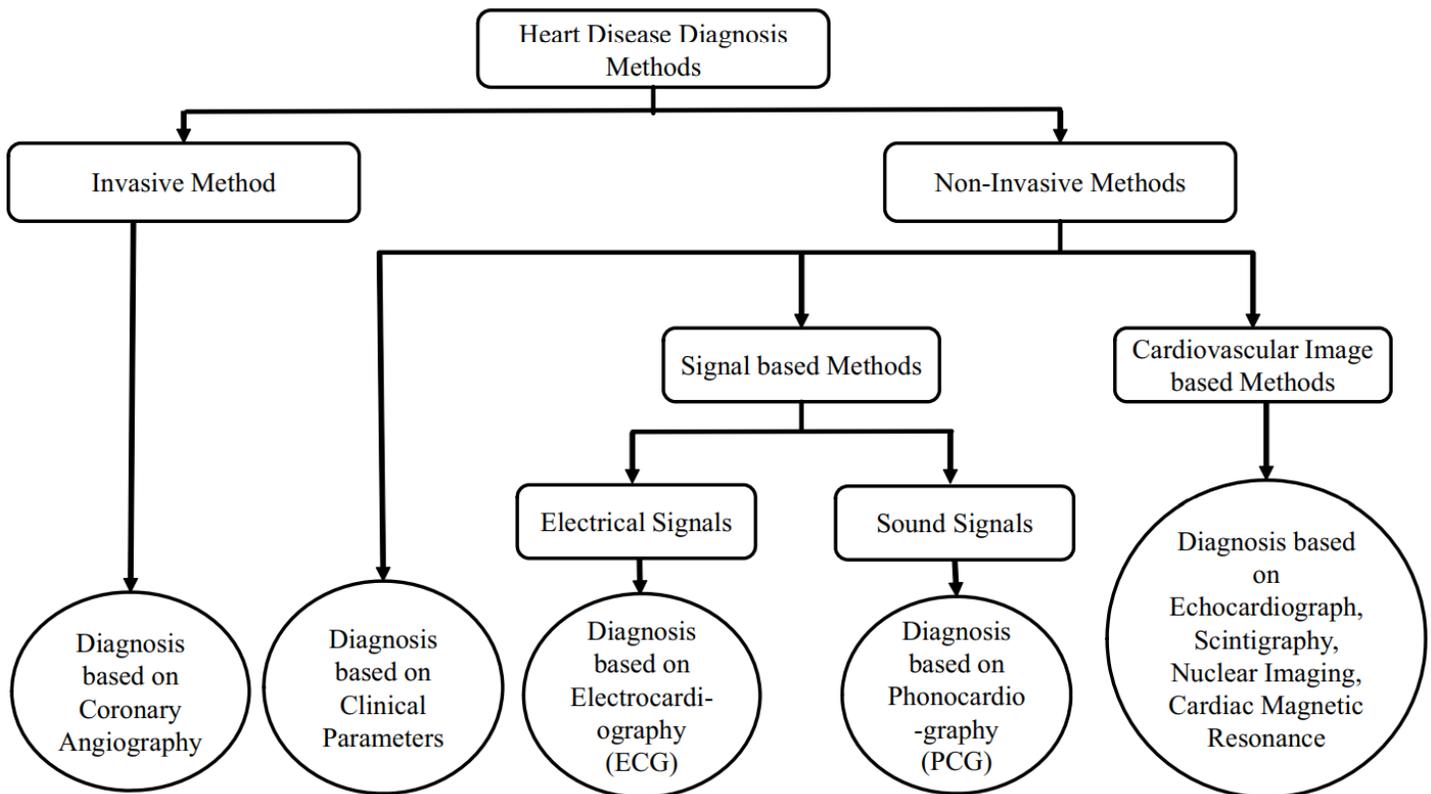


Figure 1. Methods for detecting heart problems [61]

Non-invasive (NI) approaches may create a variety of datasets that can be used to construct machine learning (ML) frameworks for the prediction of cardiac disease. This kind of framework may be classed as:

The ML Framework-A: A wide range of clinical factors, including demographic data, symptoms and examinations, laboratory tests, and many more, are used. A relational database stores these records. This framework includes datasets with numeric, category, or binary domain attributes.

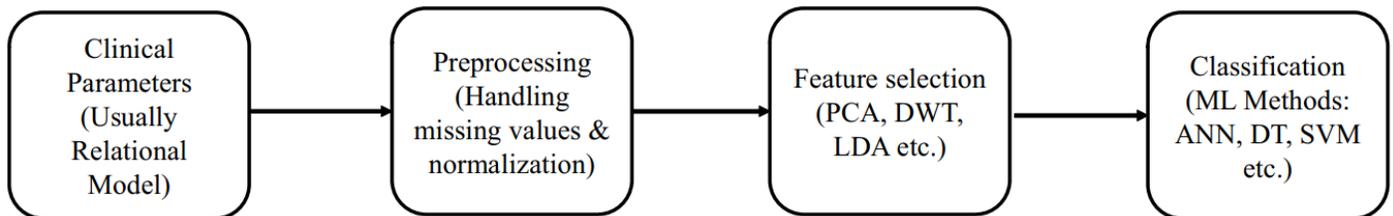


Figure 2. Machine Learning Framework-A [61]

ML Framework-B detects heart illness using cardiac signals. Heart signals are electrical and audible. This framework employs Electrocardiograms (ECG) or Phonocardiograms as raw cardiac signals (PCG). MIT-BIH (for ECG) and 2016 Physio Net/Computing in Cardiology (CinC) Challenge are examples (for PCG).

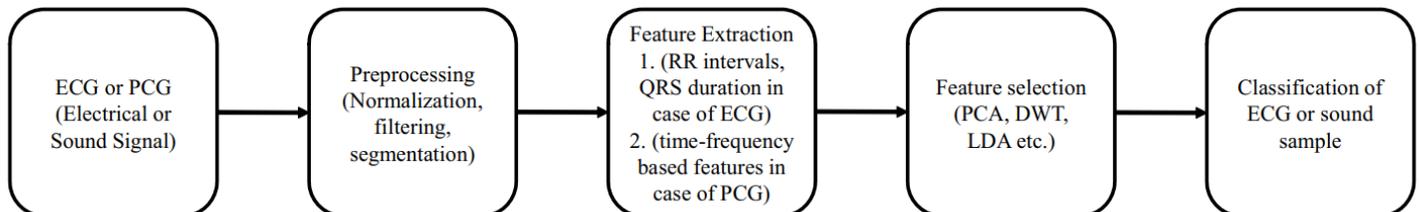


Figure 3. Machine Learning Framework-B [61]

ML Framework-C employs pictures of hearts in its documentation. Echocardiography, X-ray imaging (Computed tomography (CT)), cardiac computed tomography (CCT), nuclear imaging, single photon emission computed tomography (SPECT), scintigraphy, positron

emission tomogr-aphy (PET), etc. are used to produce these pictures [65].

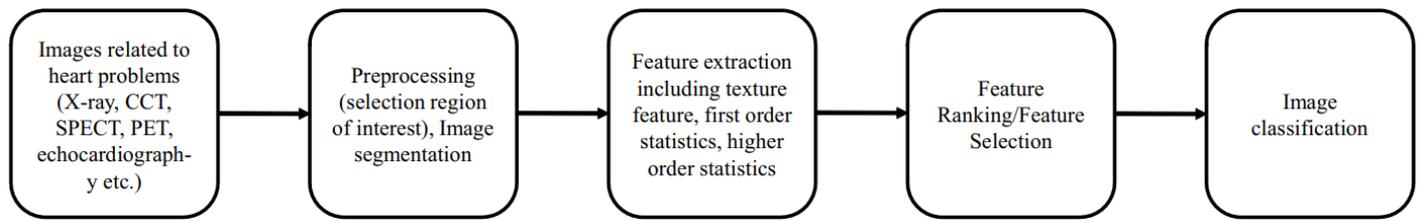


Figure 4. Machine Learning Framework -C [61]

### 3.2 Machine-learning datasets for heart disease detection

Data pertaining to patients who have heart disease are made accessible in a variety of formats inside the electronic health record [66]. In general, both clinical and physiological indicators are included in the cardiac data collection. Symptoms, demographic information, electrocardiograms (ECGs), characteristics of echocardiograms, and the results of lab examinations are some of the components that might make up a typical cardiac data set [67]. UCI is a fantastic place to look for data like these [68–71]. Perfusion imaging is performed before coronary angiography (CA), and it produces a significant amount of data for cardiologic imaging using single photon emission based computed tomography (SPECT) [72]. At the University of California, Irvine, a data collection pertaining to SPECT is accessible under the names SPECT and SPECTF. In SPECTF, there are 44 different variables, however in SPECT, the number of variables has been cut down to 22. In the image dataset known as SPECTF, all of the variables have been preprocessed to have a continuous integer domain with values ranging from 0 to 100, while in the image dataset known as SPECT, the variables have been preprocessed to be binary [73]. The electrocardiogram is the most crucial piece of data that may be affected by arrhythmia. The identification of a cardiovascular disease (CVD) may be accomplished by the categorization of an electrocardiogram (ECG). [74] Approaches such as threshold-based methods [75], methods based on digital filters, and discrete wavelet transforms [28] are often used in the process of heartbeat detection. The MIT-BIH arrhythmia dataset is the most popular source of ECG data that is utilized in the process of medical data mining [22]. The diagnosis of cardiac illness may be accomplished quickly and accurately by employing machine learning algorithms in conjunction with heart sound data [28]. Phonocardiograms are used to display information on heart sounds (PCG). PCG data are stored in a well-known repository called the 2016 PhysioNet/Computing in Cardiology (CinC) Challenge [76]. The UCI data collection contains eight different data sets that are connected to cardiac conditions. R. Alizadehsani et al. (2019) give detailed survey datasets linked to CVDs in a study that encompasses 68 databases [11]. This research was conducted in the United States. Recent research, notably those using deep learning [77–79], has been able to successfully segment and classify acoustic data. The data sets that are accessible for cardiac disorders range in size from 20 samples to 24 thousand samples, and the number of features may be anywhere from 9 to 55. The sample size that is considered to be the norm, however, is 350, and there are 9 numbers of features. The Cleveland data set, which can be found in the machine learning repository at UCI and has a sample size of 303 and a features size of 13, is by far the most popular and widely used data set [11]. The Cardiovascular Health Study, often known as CHS, is a vast repository of heart data that has 5888 samples and a significant number of features. It was produced for the prediction of heart stroke. Different numbers of samples and attributes were utilized by the researchers that examined this dataset. In a study, the CHS was used as a data set with 4988 examples and 796 features after preprocessing by Khosla et al. (2010) [80], while Tay et al. (2014) [81] used it with 5888 instances and 355 features before preprocessing the data. Following preprocessing, the number of examples was cut down to 4612, and the following analysis only used 272 characteristics.

### 3.3 Heart disease diagnosis using machine learning

In this part, a variety of machine learning frameworks, along with the models and techniques that belong to each of those frameworks, are addressed. Techniques and procedures that are considered cutting edge are dissected and contrasted across all machine learning frameworks. First, a broad explanation of machine learning will be offered, and then we will go on to discussing the specifics of the various machine learning techniques. Machine learning may be broken down into many major categories, including supervised learning, unsupervised learning, learning by reinforcement, and active learning [82]. Both supervised and unsupervised machine learning have a significant body of research in the medical area. The present investigation focuses on categorization, which is a technique to machine learning that requires human supervision. Classification is a process that involves developing or learning a mapping or connection via the use of experience gained from training data, followed by categorizing testing data using the same model or mapping or relation. A machine learning model that makes use of classification is defined in a way that is both succinct and comprehensive by Equation 1. Let it be assumed that a data collection, denoted by the letter "D," is characterized by a feature set, denoted by the notation "A=A1, A2...Am," and a sample set, denoted by the notation "S=S1, S2...Sn," with m features and s samples respectively. Let us further suppose that a machine learning framework may be characterized by a model L that accurately predicts the class C. Let us proceed with this assumption (Q). When evaluating the accuracy of the prediction, performance measure (P) is used in such

a manner as to minimize the amount of error (Err) across all training samples while in the training phase and across all testing samples while in the testing phase [83].

$$P \rightarrow \min [Err\{c - Q[L(\sum_{j=1}^n \sum_{i=1}^m D(Ai, Sj))]\}]$$

There are a variety of metrics that may be used to evaluate performance, including accuracy, specificity, sensitivity, recall, precision, and area under the curve (AUC). These performance measurements serve as a comparison operator between the many machine learning models that are available, such as decision trees (DT), support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbor (k-NN), Bayesian classifiers, and so on.

### 3.4 Recent Methodologies Applied to the ML Framework-A

Preprocessing seems to be an abstract stage shared by all types of frameworks with distinct concrete capabilities. Structure A is shown in Figure 2. In preprocessing, missing values are filled in and the data is normalized. Min-max normalization is a typical approach for filling missing values using instance mean during normalizing [84]. The second key phase in the framework is feature selection, which speeds the process of developing a machine learning (ML) model, resulting in enhanced efficiency and accuracy [85]. Principal component analysis (PCA) has been employed in several recent studies of heart disease prediction [12] for feature selection. A very recent and comprehensive study has been given for data pre-processing during cardiac disease categorization [86]. The subsequent step is classification, which is carried out using a number of classifiers, including decision trees, Bayesian classifiers, artificial neural networks, and support vector machines. Zhou et al. (2021) [85] used feature weight-based feature selection and classification through decision tree. In preprocessing, several approaches, like kmean and ReliefF, are used. Karaolis et al. (2010) [87] examined the risk of coronary heart disease using a decision tree, and classification rules were retrieved by selecting the most significant risk prediction component. Alizadehsani et al. (2013) [67] presented a very informative dataset for the prediction of cardiac disease. Using 10-fold cross validation, Sequential Minimum Optimization (SMO) Bagging, a training method of SVM, ANN, and Nave Bayes algorithms, is trained and evaluated, and it is discovered that SMO bagging has the greatest accuracy.

Table 1. Some Recent Applied to the ML Framework-A [61]

Reference	Data Set	# of (features, samples)	Classifiers	Accuracy
[85]	SPECT(heart)	(23, 267)	DT	67.44
	SPECTF	(44, 267)		71.14
	Statlog	(13, 270)		72.74
[83]	Cleveland heart	(303, 14)	FAMD-LR	91.8
			FAMD-kNN	90.16
			FAMD-SVM	91.8
			FAMD-DT	81.96
			FAMD-RF	93.44
[87]	Dept of Card.PGH, Cyprus	(14, 1500)	DT	75
[90]	Statlog(heart)	(13, 270)	ANN-CAPSO	81.85
[67]	Z-Alizadeh Sani	(54, 303)	Bagging SMO	94.08

It has been noted throughout the course of the previous decade that hybridizing machine learning methods with a metaheuristics approach for the purpose of medical diagnosis, particularly in the prediction of heart illnesses, led to improvements [91]. When a single hidden layer Feed forward Neural Network was trained using Competitive Swarm Optimization and Extreme Machine Learning [88], these kinds of gains were seen. In an ensemble fuzzy boosting strategy for classification, the Particle Swarm Optimization algorithm was used to improve accuracy in heart disease prediction [89]. In Table 1, we included some recent noteworthy research that adhere to the approach of the ML Framework-A. The reference to the study, the name of the dataset that was used, the number of features, the number of samples, the name of the classifier, and its performance metrics are all included in the table, which is orga-

nized according to Equation 1.

### 3.5 Recent Methodologies applied to the machine learning framework-B

Electrocardiographs (ECG) and phonocardiograms (PCG) are diagnostic tools that use electrical and sound signals, respectively, to depict the physiological state of the heart [79]. Classification of an electrocardiogram (ECG) is a multi-step procedure that begins with preprocessing and continues through segmentation, feature extraction, feature selection, and finally classification [92]. Preprocessing, which includes filtering and segmentation, is the first phase in the PCG signals classification process [77]. The next step is feature extraction, and then the last step is classification. Figure 3 provides a combined workflow of the process that is utilized for the categorization of signal data. Framework B is illustrated in this figure (ECG and PCG). In Table 2, an approach known as ML Framework-B is used to provide a summary of some more recent research that summarize ECG and PCG data. The reference number for the study effort, the name of the data set, the primary preprocessing tasks that were completed, and the names of the classifiers with the best accuracy, sensitivity, and specificity are all included in the table. In the next two subsections, a short explanation is provided for recent research that used ECG and PCG categorization.

Table 2. Some Recent applied on ML Framework-B [61]

Reference	Data Set	Pre-processing	Classifiers	Accuracy
[75]	MIT-BIH	Shot Time Fourier Transform (STFT)	CNN	99
[93]	MIT-BIH	Welsh method, Discrete Fourier Transform(DFT)	GA-SVM	98.85
[94]	MIMIC-II	Gaussian curve fitting, Singular Value Decomposition (SVD)	SVM	*
[95]	Long Term ST Database	Cut of Freq 0.3 Hz (LPF) 20Hz(HPF)	SVM	99.2
[96]	PhysioNet 2016 challenge	HSMM	RF	85

#### 3.5.1 Methodologies involving ECG

When compared to ML Framework-A, ML Framework-B consists of a greater number of individual stages. Preprocessing, segmentation, feature extraction, feature selection, and classification are all a part of it. In this particular research endeavor, the processes of preprocessing and segmentation were grouped together as one step. During preprocessing of an electrocardiogram (ECG), one of the most important tasks is to identify and reduce frequencies that are linked with artifacts. Because indiscriminate and adaptive filters may sometimes warp the true shape of data, wavelet transformations have been more popular as a preprocessing technique in recent years [99]. In the preprocessing of ECG data, some recent research [97, 98] used the use of high pass filters, low pass filters, band rejection filters, base line wander filters, and notch filters. Before beginning the segmentation process, normalization and augmentation of the QRS complex are carried out [50]. A signal is broken up into its component parts, or segments, so that it may be studied more effectively [100]. Segmentation is the process that does this.

The part of the ECG classification process known as feature extraction is highly critical. Common characteristics of an electrocardiogram include the length of the P wave, the PQ/PR/QT interval, the width of the QRS, the amplitudes of the P/T/QRS, and the ST level. However, the RR interval is the characteristic that is used most often in machine learning [99]. The Pan-Tompkins technique is of

ten used to the task of analyzing R peaks, which is eventually necessary for the identification of RR intervals and QRS [101]. Both temporal characteristics and morphological features were offered by Das and Ari (2014) as two primary kinds of extracted features. Temporal characteristics are calculated using RR intervals, whereas other features are calculated using the S-transform or the wavelet transform. The process of feature extraction was carried out by Pawiak (2018) [93] using the power spectral density technique, the Welch method, the periodogram, the Fourier discrete transform method, the Hamming window, and a series of logarithms of signals. The next step in framework-B is the feature selection stage, which aims to identify the characteristics that are the most relevant. The characteristics that were chosen to be included in the categorization helped to increase its accuracy while also accelerating the process. In a research conducted by Song et al. (2006) [102], the authors found that the performance of an SVM classifier could be improved using linear discriminant analysis (LDA) by categorizing ECG signals. As a result, they were able to reach a maximum accuracy of 99.88 percent. The genetic algorithm was employed as a feature selection method by encoding genes with a value of 0 to indicate a refusal to accept a particular feature and a value of 1 to indicate acceptance of the feature. After the feature selection is complete, the ECG signals are classified using support vector machines, k-neural networks, probabilistic neural networks, and radial basis function neural networks [93]. In a separate piece of research conducted by Dalal and Vishwakarma (2020), [103] the evolutionary algorithm is used to optimize the kernel extreme machine. These days, deep learning techniques are also demonstrating their efficacy in the categorization of electrocardiograms (ECGs). In contrast with 1-D CNN, Huang et al. (2019) [75] introduced an electrocardiogram (ECG) classification system that was built on a 2-dimensional convolution neural network (2-D CNN), and they obtained a considerable classification accuracy of 99.00 percent (90.93 percent). The need for preprocessing has been drastically cut down thanks to deep neural networks.

### 3.5.2 Methodologies involving PCG

Phonocardiography is a very fundamental approach for identifying the sounds that the heart makes. Phonocardiograms, often known as PCGs, are the results of recordings made using phonocardiography. An ancient but well-known non-invasive approach for diagnosing cardiac conditions is called auscultation, and it involves listening to the sounds that the heart makes [79]. The auscultation of the heart sound calls for a significant amount of experience and skill [104]. Recent years have seen the emergence of automated categorization of heart sound waves as a prospective study subject in artificial intelligence [105]. The segmentation of sounds, the extraction of features, and the categorization of sounds that come from the heart are the three steps that make up the classification process. An electrocardiogram (ECG) recording is segmented into a series of basic heart sound signals during the heart sound segmentation process (FHSs). The first and second fundamental heart sounds, or FHSs, occur at the beginning of systole and the beginning of diastole, respectively [78]. Sound segmentation is performed with the assistance of an ECG reference [106] in order to locate any additional sounds or murmurs that may be present between the periods S1-S2 (systole) or S2-S1 (diastole).

In research that was conducted by Chen et al. (2017) [28], the identification of first heart sound, second heart sound, third heart sound, and fourth sound segmentations are discussed with references to different time segments of ECG. By giving characteristics for the envelope, one may prevent the interference caused by noise. These are referred to as the harmonic envelope, the Hilbert envelope, the wavelet envelope, and the Power Spectral Density (PSD) envelope, respectively. After that, the duration parameters eHR and eSys are determined by using the envelope feature in conjunction with some autocorrelation analysis of PCG [78]. Feldman and Braun and Liang and colleagues have contributed some preliminary work to the field of sound segmentation [107], [108]. Maglogiannis et al. and Schmidt et al. have recently carried out this research [109], [110]. The process of extracting features requires the collection of distinguishing characteristics. The classifier will ultimately make use of the characteristics in order to assign categories to the sound signals. The PCG classification uses either time or frequency as the basis for the characteristics that are retrieved [111, 112]. There have been recent developments in feature extraction approaches that include both the domain and the time-frequency domain [113–115]. Following the steps of sound segmentation and feature extraction, respectively, the following step (which is also the last step) in the sound classification approach is training using the features that were extracted in order to predict the label of the sound signals. Various classifiers have been put up as potential solutions for the classification [96, 116]. For PCG classification, the deep learning technique employing convolutional neural networks was shown to perform better than decision trees, Bayesian classifiers, artificial neural networks, and support vector machines.

### 3.6 Recent Methodologies applied on machine learning framework-C

X-ray computed tomography (CT), echocardiography, cardiac computed tomography (CCT), cardiac magnetic resonance (CMR), nuclear imaging, single photon emission computed tomography (SPECT), scintigraphy, and positron emission tomography (PET) are all part of the multimodal nature of cardiovascular imaging [117]. Radiologists are seeing a paradigm shift thanks to automated cardiovascular image processing, which also helps speed up the identification of heart issues [65]. The procedures involved in image categorization using ML framework-C are broken down and summarized in Figure 4. Echocardiography is primarily a cost-effective imaging method that may be performed. The picture obtained from echocardiogram undergoes preprocessing, which consists of

reducing the noise in order to provide a smoother appearance. During the preprocessing stage, any extraneous artifacts are eliminated, and the area of the picture that includes the heart is chosen. The subsequent step is the extraction of characteristics such as the mean, standard deviation, entropy, and texture features, and the last step is the performance of picture classification [118]. In scintigraphy and SPECT, perfusion abnormalities may be identified by the use of some kind of correlation computation in both resting and stressful pictures. In most cases, perfusion may be identified with the use of some color thresholding. Following that, feature extraction and segmentation are carried out, and this is done just before feature ranking. Feature ranking is a process that is similar to feature selection in that it first ranks various features according to their discriminative property and then selects the best compact extracted feature subset [119]. Feature ranking is a process that is used to select the best compact extracted feature subset. The existence of cardiac disease may be determined by abnormal imaging [72].

Pictures obtained using echocardiography primarily focus on quantifying chambers, measuring ejection fraction and strain, capturing images of valves, and assessing the heart's overall function [117]. When doing echocardiography, ultrasound imaging may be used to get a picture of arteries that have been damaged by atherosclerosis, which is a condition that can eventually lead to coronary heart disease [120]. In the fields of AI and ML, ultrasound imaging is an extremely widespread, low-cost, highly reliable, and non-invasive approach [121]. In their 2015 study, Sudarshan et al. examined and gave a short explanation of commonly utilized clinical and non-clinical aspects, as well as current investigations of the features that were used to diagnose myocardial infarction [119]. CMR is utilized to measure the function of the heart as well as the volumes of the heart in a more accurate manner. Tan et al. (2018) employed an artificial neural network (ANN) based entirely automated short axis and long axis information to segment the left ventricle picture [122]. This led to the enhanced diagnosis of cardiovascular disease, which eventually increased the study's effectiveness. The jaccard index, a tool that measures the degree to which two picture subjects are same, was chosen as the performance metric for this research. AI-based CCT techniques are used in the detection of quantification of the arterial plaques, flow of blood, and coronary artery calcium scoring [117]. When it came to building an ML model to identify obstructive coronary artery stenoses, researchers found that AdaBoost performed much better than Naive Bayes and Random Forest in terms of accuracy, sensitivity, and specificity as well as the ROC curve [123].

Both PET and SPECT make use of the same approaches in order to identify cardiovascular problems via the use of cardiovascular imaging. A cardiac image data set was created for public use like that of UCI based on the results of a fairly recent research that employed stress and rest pictures of 192 patients. SPECT imaging was used in conjunction with both a knowledge-based classification model and a deep learning-based model for the purpose of identifying perfusion anomalies. Both models have been shown to have a sensitivity of one hundred percent, although deep learning performed better in terms of accuracy. The shallow features, on the other hand, take precedence over the deep features.

Table 3. Some Recent Studies applied on ML Framework-C [61]

Reference	Techniques	Imaging	Accuracy
[72]	Deep Learning SVM	SPECT	93
[120]	Random Forest	Echocardiography	*
[122]	Neural network regression	CMR	*
[118]	BPNN	Echocardiography	87.5
[123]	AdaBoost, PCA	CCT	70

The SVM was responsible for the creation of the shallow features. In a nutshell, the research was able to attain greater accuracy by combining the traditional machine learning model with the deep learning model [72]. A comprehensive analysis of the many classifications used for cardiovascular images is beyond the purview of this paper. On the other hand, an overview of current research on cardiovascular imaging and the prediction of heart disease is provided in Table 3. The reference number of the study article, metho-

dologies used for classification, the kind of imaging, accuracy, sensitivity, specificity, and some other performance indicators are included in the table. The p value, area under the curve (AUC), amount of time required to develop the model, and jaccard index are these measurements. If the value of p is less than or equal to 0.05, then the findings may be considered to be statistically significant. If the value of the area under the curve (AUC) is close to 1, then the classification is good, but it is deemed to be bad if it is less than 0.5. The jaccard coefficient is used in [122] to assess how similar the cardiac images are to one another.

### 3.7 MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis

The purpose of developing a machine intelligence framework for the diagnosis of heart disease is to increase the survival rate of patients through the accurate, precise, and early detection of disease in order to potentialize the system in predicting heart disease in order to increase the survival rate of patients. The purpose of this project is to provide an automated solution that will aid medical professionals, namely radiologists, in making diagnoses and choices with more accuracy and certainty, all while reducing the amount of time spent on analysis. Due to the highly variable nature of medical data, pre-processing of medical datasets is an important step that should not be skipped. The framework includes both the extraction of features and the analysis of the effect that machine learning prediction models have. The selection and reduction of features, as well as the creation of brand new features, are all included in the process of "extracting" features. The goal is to develop a machine intelligence framework for the identification of heart illness, with the primary emphasis being placed on the process of label prediction for the UCI Cleveland heart disease dataset. Because the Cleveland dataset is of mixed type and comprises both numeric and categorical variables, factor analysis of mixed data (FAMD) is an appropriate method for obtaining those features. During the training of the models, both the extracted features and the generated features are used in order to classify normal participants and cardiac patients [124–131]. Figure 5 presents a schematic depiction of the process, which is presented in detail in order to improve comprehension of the proposed framework MIFH.

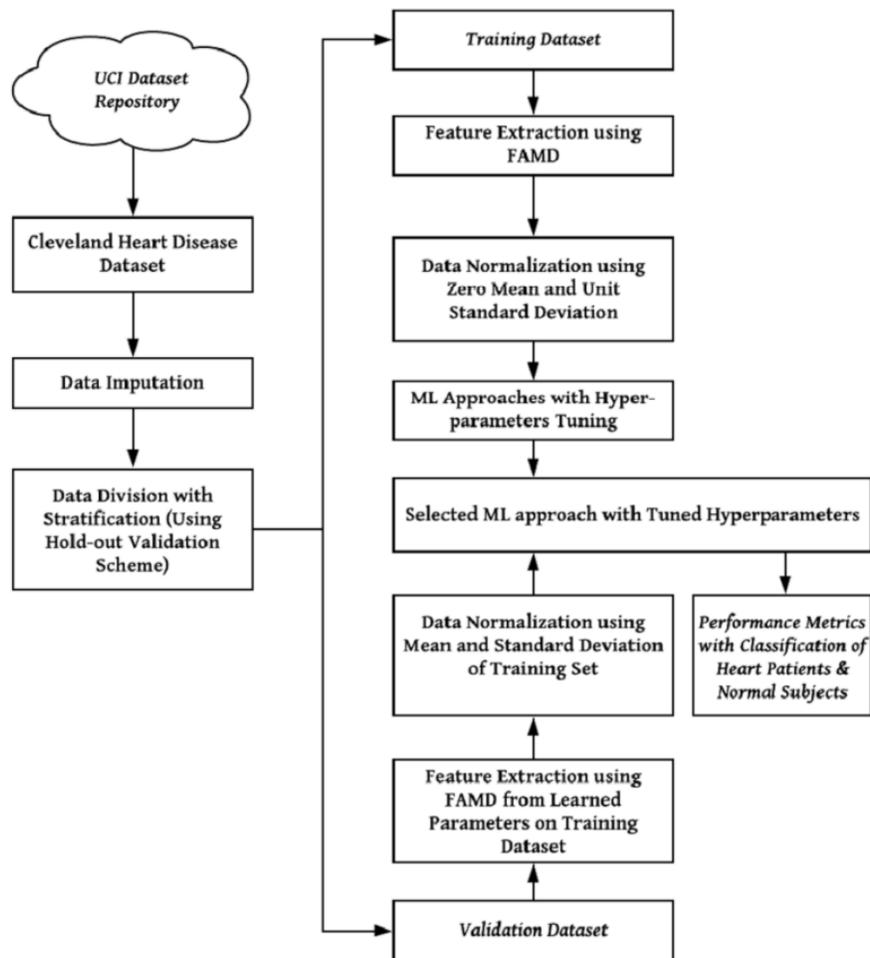


Figure 5. MIFH: A machine intelligence framework for Cleveland heart disease dataset [62]

The MIFH framework is comprised of a number of processes, the most important of which are the following: data imputation and partitioning; feature extraction via FAMD; feature normalization; machine learning approach; and performance metric evaluation.

#### 4. Discussion

As can be observed from the recent works that are reviewed and emphasized in this paper, the use of machine learning techniques has made a significant contribution to the development of automated methods for the detection of cardiac disease. There are three different types of data that can be used to construct a machine learning model: (i) data based on clinical parameters and physiological parameters, (ii) data based on signals, and (iii) data based on images of the heart. Each of these three types of data can be used to build the model. This review research presents and discusses three different machine learning frameworks, each of which corresponds to one of these kinds. The preprocessing involved in framework A is far less complicated than that involved in frameworks B and C. Machine learning algorithms such as ANN, SVM, KNN, DT, and Bayesian classifiers operate well when given physiological and clinical data as well as ECG signals. Table 1 displays and provides a description of the state-of-the-art machine learning approaches (ML Framework-A) utilized for predicting heart disorders using contemporary comparison operators such as accuracies, sensitivities, specificities, AUC, F-measures, precision, recall, and so on [83]. If machine learning methods are hybridized with metaheuristic techniques, such as feature selection algorithms or optimization techniques, the accuracies of diagnosing cardiac diseases are shown to improve [63], [88-90], [103].

When employing ML Framework-A, most of the attention paid by researchers is directed toward feature selection and classification, but in Framework-B and C, feature extraction is the stage that receives the most attention. If we do not compute the relevant characteristics and parameters, the raw ECG or PCG that was acquired is of little value. Bio-signals may take on a variety of convoluted forms, and unearthing their hidden characteristics can be a difficult endeavor. ECG and PCG are regarded to be part of the same ML Framework despite certain differences in the feature extraction process. This is due to the fact that signals processing is a problem shared by both. Some of the common features of an electrocardiogram (ECG) include the RR intervals, QRS duration, mean of RR intervals, and standard deviation of successive RR intervals. In contrast, the analysis of PCG signals can take place in either the time domain, the frequency domain, or the time-frequency domain. In recent research, 'envelope features' and 'duration features' have been used for heart sound segmentation, which has finally led to improvements in the prediction of heart illness [78], [64]. It has been shown that SVM and ANN classification are superior than ECG classification. According to the most recent developments, PCG classification is best handled by deep neural networks. The detection of heart illness by cardiovascular imaging has likewise been rendered obsolete by deep neural networks. PCG classification and cardiovascular picture categorization are two areas where deep neural networks perform very well. Although there is now a large number of data repositories and machine learning approaches accessible for diagnosing heart disease, there is still room for development in terms of accuracies and other performance indicators.

#### 5. Conclusion

All the different kinds of data that are utilized for the prognosis of heart disease are subjected to a concise yet educational evaluation. The modern supervised machine learning approaches that are employed in the detection of heart disease have been gathered, examined, and it was discovered that there are three primary processes. The huge body of research that was examined revealed that the majority of the studies relied on the Cleveland heart disease dataset. This dataset consists of just 303 occurrences and 14 characteristics and comprises only 303 cases. The number of people in the sample that are intended to represent a certain region is very low and constrained. Only a small number of research that relied on other sources of data also utilized a single dataset with restricted heart disease characteristics. According to the findings of the research, standard prediction models are neither effective nor purposefully developed to deal with class-based data imbalances. In addition, class imbalanced datasets may be investigated in order to cope with real-life events that occur in healthcare facilities such as hospitals and clinics. Because of this, it was impossible to generalize the numerous classification accuracies that were gained in the process of predicting heart disease. Other multiple heart disease datasets from geographically diverse sources with more features should be explored for the purpose of developing more efficient machine learning models. This is the fundamental intent of our future research, which is currently in progress. Obtaining a more generalized classification and prediction accuracy will allow for the development of more accurate diagnostic tools. This would allow for a more accurate categorization of heart disorders as well as earlier prediction of their onset, which would, in turn, reduce the alarmingly high rates of morbidity and death caused by cardiovascular diseases (CVDs).

#### 6. References

- [1] H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, "Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence," *J. Med. Syst.*, vol. 36, no. 4, pp. 2505-2519, Aug. 2012, doi: 10.1007/s10916-011-9723-0.
- [2] K. Mala, V. Sadasivam, and S. Alagappan, "Neural network based texture analysis of CT images for fatty and cirrhosis liver classification," *Appl. Soft Comput.*, vol. 32, pp. 80-86, Jul. 2015, doi: 10.1016/j.asoc.2015.02.034.
- [3] N. Nai-arun, "Comparison of Classifiers for the Risk of Diabetes Prediction," *Procedia Comput. Sci.*, p. 11, 2015.
- [4] H.-L. Chen, D.-Y. Liu, B. Yang, J. Liu, and G. Wang, "A new hybrid method based on local fisher discriminant analysis and support vector ma-

- chines for hepatitis disease diagnosis," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11796–11803, Sep. 2011, doi: 10.1016/j.eswa.2011.03.066.
- [5] F. Özyurt, E. Sert, and D. Avci, "An expert system for brain tumor detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine," *Med. Hypotheses*, vol. 134, p. 109433, Jan. 2020, doi: 10.1016/j.mehy.2019.109433.
- [6] H. Yang and Y.-P. P. Chen, "Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information," *Expert Syst. Appl.*, vol. 42, no. 15–16, pp. 6168–6176, Sep. 2015, doi: 10.1016/j.eswa.2015.03.019.
- [7] B. Alizadeh Savareh et al., "A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures," *Pancreatology*, vol. 20, no. 6, pp. 1195–1204, Sep. 2020, doi: 10.1016/j.pan.2020.07.399.
- [8] S. F. Weng, J. Repts, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLOS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.
- [9] J. Laosai and K. Chamnongthai, "Classification of acute leukemia using medical-knowledge-based morphology and CD marker," *Biomed. Signal Process. Control*, vol. 44, pp. 127–137, Jul. 2018, doi: 10.1016/j.bspc.2018.01.020.
- [10] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 904–913, Mar. 2014, doi: 10.1016/j.cmpb.2014.01.004.
- [11] R. Alizadehsani et al., "A database for using machine learning and data mining techniques for coronary artery disease diagnosis," *Sci. Data*, vol. 6, no. 1, p. 227, Dec. 2019, doi: 10.1038/s41597-019-0206-3.
- [12] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Inform. Med. Unlocked*, vol. 19, p. 100330, 2020, doi: 10.1016/j.imu.2020.100330.
- [13] P. A. Heidenreich et al., "Forecasting the future of cardiovascular disease in the United States," *Circulation*, vol. 123, no. 8, pp. 933–944, 2011.
- [14] World health organization: Who. Accessed: Jul. 25, 2019. [Online]. Available: <https://www.who.int>
- [15] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," *Nature Rev. Cardiol.*, vol. 8, no. 1, pp. 30–41, Jan. 2011.
- [16] J. López-Sendón, "The heart failure epidemic," *Medicographia*, vol. 33, no. 4, pp. 363–369, 2011.
- [17] L. A. Allen, "Decision-making in advanced heart failure," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.
- [18] T. Tirkes, M. A. Hollar, M. Tann, M. D. Kohli, F. Akisik, and K. Sandrasegaran, "Response criteria in oncologic imaging: Review of traditional and new criteria," *Radio Graph.*, vol. 33, no. 5, pp. 1323–1341, Sep. 2013.
- [19] "Robbins Basic Pathology - 10th Edition," 2017. <https://www.elsevier.com/books/robbins-basic-pathology/kumar/978-0-323-35317-5> (accessed Dec. 15, 2020).
- [20] B. Xiao et al., "Follow the Sound of Children's Heart: A Deep-Learning-Based Computer-Aided Pediatric CHDs Diagnosis System," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1994–2004, Mar. 2020, doi: 10.1109/JIOT.2019.2961132.
- [21] R. Tao et al., "Magnetocardiography-Based Ischemic Heart Disease Detection and Localization Using Machine Learning Methods," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1658–1667, Jun. 2019, doi: 10.1109/TBME.2018.2877649.
- [22] Z. F. M. Apandi, R. Ikeura, and S. Hayakawa, "Arrhythmia Detection Using MIT-BIH Dataset: A Review," in *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, Kuching, Aug. 2018, pp. 1–5, doi: 10.1109/ICASSDA.2018.8477620.
- [23] D. Hanbay, "An expert system based on least square support vector machines for diagnosis of the valvular heart disease," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4232–4238, Apr. 2009, doi: 10.1016/j.eswa.2008.04.010.
- [24] R. Das, I. Turkoglu, and A. Sengur, "Diagnosis of valvular heart disease through neural networks ensembles," *Comput. Methods Programs Biomed.*, vol. 93, no. 2, pp. 185–191, Feb. 2009, doi: 10.1016/j.cmpb.2008.09.005.
- [25] E. Avci, "A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier," *Expert Syst. Appl.*, p. 9, 2009.
- [26] G. T. Reddy and N. Khare, "An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model," *J. Circuits, Syst. Comput.*, vol. 26, no. 4, Apr. 2017, Art. no. 1750061.
- [27] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, Apr. 2017.
- [28] T. Chen, S. Zhao, S. Shao, and S. Zheng, "Non-invasive diagnosis methods of coronary disease based on wavelet denoising and sound analyzing," *Saudi J. Biol. Sci.*, vol. 24, no. 3, pp. 526–536, Mar. 2017, doi: 10.1016/j.sjbs.2017.01.023.
- [29] R. Alizadehsani et al., "Machine learning-based coronary artery disease diagnosis: A comprehensive review," *Comput. Biol. Med.*, vol. 111, p. 103346, Aug. 2019, doi: 10.1016/j.combiomed.2019.103346.
- [30] Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. *Am J Med* 2012;125:882-7.e1.
- [31] Yach D. Adherence to long-term therapies evidence for action, 2003. Available online: <https://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf>
- [32] Leslie KH, McCowan C, Pell JP. Adherence to cardiovascular medication: a review of systematic reviews. *J Public Health (Oxf)* 2019;41:e84-94.
- [33] Fernandez-Lazaro CI, Adams DP, Fernandez-Lazaro D, et al. Medication adherence and barriers among low-income, uninsured patients with multiple chronic conditions. *Res Social Adm Pharm* 2019;15:744-53.
- [34] Al-Ganmi AHA, Alotaibi A, Gholizadeh L, et al. Medication adherence and predictive factors in patients with cardiovascular disease: A cross-sectional study. *Nurs Health Sci* 2020;22:454-63.
- [35] Mann DM, Woodward M, Muntner P, et al. Predictors of nonadherence to statins: a systematic review and metaanalysis. *Ann Pharmacother* 2010;44:1410-21.
- [36] Choudhry NK, Fischer MA, Avorn J, et al. The implications of therapeutic complexity on adherence to cardiovascular medications. *Arch Intern Med* 2011;171:814-22.

- [37] Gellad WF, Grenard JL, Marcum ZA. A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. *Am J Geriatr Pharmacother* 2011;9:11-23.
- [38] Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005;353:487-97.
- [39] Steiner JF. Can we identify clinical predictors of medication adherence... and should we? *Med Care* 2010;48:193-5.
- [40] Lauffenburger JC, Franklin JM, Krumme AA, et al. Predicting Adherence to Chronic Disease Medications in Patients with Long-term Initial Medication Fills Using Indicators of Clinical Events and Health Behaviors. *J Manag Care Spec Pharm* 2018;24:469-77.
- [41] Deo RC. Machine Learning in Medicine. *Circulation* 2015;132:1920-30.
- [42] Nashif, S., Raïban, M., Islam, M., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World Journal of Engineering and Technology*, 6, 854-873.
- [43] Hussein, M. U. (2017, October 29). Physics and the Cardiovascular System. Retrieved from ResearchGate: <https://www.researchgate.net>.
- [44] WHO. (2011). Global Atlas on cardiovascular disease prevention and control. Geneva: WHO Library Cataloguing.
- [45] Nagendra, K. V., & Ussenaiah, M. (2018). A study on various data mining techniques used for heart diseases. *International Journal of Recent Scientific Research*, 24350- 24354.
- [46] WHO. (2016). Technical package for cardiovascular disease management in primary health care. Geneva: WHO Library Cataloguing.
- [47] WHO. (2017). Global action plan for the prevention and control of noncommunicable diseases. Geneva: WHO Library Cataloguing.
- [48] Ayatollahi, H., Gholamhosseini, L., & Salehi, M. (2019). Predicting coronary artery disease: a comparison between two data mining algorithms. *BMC Public Health*. doi: 10.1186/S12889-019-6721-5.
- [49] Annepu, D., & Gowtham, G. (2019). Cardiovascular disease prediction using machine learning techniques. *International Research Journal of Engineering and Technology*, 6 (4), 3963-3971.
- [50] Hariharan, K., Vigneshwar, W. S., Sivaramakrishnan, N., & Subramaniaswamy, V. (2018). A comparative study on heart disease analysis using classification techniques. *International Journal of Pure and Applied Mathematics*, 119 (12), 13357-13366.
- [51] Subhadra, K., & Vikas, B. (2019). Neural network based intelligent system for predicting heart disease. *International Journal of Innovative Technology and Exploring Engineering*, 8 (5), 484-487.
- [52] Lakshmanarao, A., Swathi, Y., Sri, P., & Sundareswar, S. (2019). Machine learning techniques for heart disease prediction. *International Journal of Science and Technology Research*, 8 (11), 374-377.
- [53] Shamsollahi, M., Badiie, A., & Ghazanfari, M. (2019). Using combined descriptive and predictive methods of data mining for coronary artery disease prediction: a case study approach. *Journal of Artificial Intelligence and Data Mining*, 7 (1), 47-58.
- [54] Rabbi, M. F., Uddin, M. P., Ali, M. A., & Kibria, M. F. (2018). Performance evaluation of data mining classification techniques for heart disease prediction. *American Journal of Engineering Research*, 7 (2), 278-283.
- [55] Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. *Hindawi Journal of Healthcare Engineering*. doi: 10.1155/2017/2780501.
- [56] Banu, G. R., & Jamala, J. H. (2015). Heart attack prediction using data mining technique. *International Journal of Modern Trends in Engineering and Research*, 2 (5), 428-432.
- [57] Padmanabhan, M., Yuan, P., Chada, G., & Nguyen, H. V. (2019). Physician-friendly machine learning: a case study with cardiovascular disease risk prediction. *Journal of Clinical Medicine*. doi: 10.3390/jcm8071050.
- [58] Sridhar, A., & Kapardhi, A. (2018). Predicting heart disease using machine learning algorithm. *International Research Journal of Engineering and technology*, 6 (4), 36-38.
- [59] Tarawneh, M., & Embarak, O. (2019). Hybrid approach for heart disease prediction using data mining techniques. *Acta Scientific Nutritional Health*, 3 (7), 147-151.
- [60] Haq, A. U., Li, J.-P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Hindawi Mobile Information System*. doi: 10.1155/2018/3860146.
- [61] Kumar, N., & Kumar, D. (2021, August). Machine learning based heart disease diagnosis using non-invasive methods: a review. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012081). IOP Publishing.
- [62] Gupta, A., Kumar, R., Arora, H. S., & Raman, B. (2019). MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE access*, 8, 14659-14674.
- [63] Jalil, N., Smith, S. C., & Green, R. (2020). Performance optimization of rotation-tolerant Viola-Jones-based blackbird detection. *Journal of Real-Time Image Processing*, 17(3), 471-478.
- [64] Jalil, N. (2021). Introduction to Intelligent User Interfaces (IUIs). In *Software Usability*. IntechOpen.
- [65] G. Litjens et al., "State-of-the-Art Deep Learning in Cardiovascular Image Analysis," *JACC Cardiovasc. Imaging*, vol. 12, no. 8, pp. 1549-1565, Aug. 2019, doi: 10.1016/j.jcmg.2019.06.009.
- [66] B. Wang et al., "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," *IEEE Access*, vol. 7, pp. 178392-178400, 2019, doi: 10.1109/ACCESS.2019.2956859.
- [67] R. Alizadehsani et al., "A data mining approach for diagnosis of coronary artery disease," *Comput. Methods Programs Biomed.*, vol. 111, no. 1, pp. 52-61, Jul. 2013, doi: 10.1016/j.cmpb.2013.03.004.
- [68] G. Kan, C. A. Visser, J. J. Koolen, and A. J. Dunning, "Short and long term predictive value of admission wall motion score in acute myocardial infarction. A cross sectional echocardiographic study of 345 patients," *Br. Heart J.*, vol. 56, no. 5, pp. 422-427, Nov. 1986.
- [69] H. A. Guvenir, B. Acar, G. Demiroz, and A. Cekin, "A supervised machine learning algorithm for arrhythmia analysis," in *Computers in Cardiology* 1997, Sep. 1997, pp. 433-436, doi: 10.1109/CIC.1997.647926.
- [70] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de-Sá, and L. Pereira-Leite, "SisPorto 2.0: a program for automated analysis of cardiocograms," *J. Matern. Fetal Med.*, vol. 9, no. 5, pp. 311-318, Oct. 2000, doi: 10.1002/1520-6661(200009/10)9:53.0.CO;2-9.

- [71] "UCI Machine Learning Repository: SPECT Heart Data Set," 2001. <https://archive.ics.uci.edu/ml/datasets/SPECT+Heart> (accessed Dec. 15, 2020).
- [72] S. Kaplan Berkaya, I. Ak Sivriköz, and S. Gunal, "Classification models for SPECT myocardial perfusion imaging," *Comput. Biol. Med.*, vol. 123, p. 103893, Aug. 2020, doi: 10.1016/j.combiomed.2020.103893.
- [73] L. A. Kurgan, K. J. Cios, R. Tadeusiewicz, M. Ogiela, and L. S. Goodenday, "Knowledge discovery approach to automated cardiac SPECT diagnosis," *Artif. Intell. Med.*, vol. 23, no. 2, pp. 149–169, Oct. 2001, doi: 10.1016/S0933-3657(01)00082-3.
- [74] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," *J. Adv. Res.*, vol. 4, no. 4, pp. 331–344, Jul. 2013, doi: 10.1016/j.jare.2012.05.007.
- [75] J. Huang, B. Chen, B. Yao, and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019, doi: 10.1109/ACCESS.2019.2928017.
- [76] G. D. Clifford et al., "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, no. 8, pp. E10–E25, Aug. 2017, doi: 10.1088/1361-6579/aa7ec8.
- [77] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, Oct. 2020, doi: 10.1016/j.neunet.2020.06.015.
- [78] Y. Chen, J. Lv, Y. Sun, and B. Jia, "Heart sound segmentation via Duration Long-Short Term Memory neural network," *Appl. Soft Comput.*, vol. 95, p. 106540, Oct. 2020, doi: 10.1016/j.asoc.2020.106540.
- [79] [36] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, Jun. 2020, doi: 10.1016/j.neucom.2018.09.101.
- [80] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, Washington, DC, USA, 2010, p. 183, doi: 10.1145/1835804.1835830.
- [81] D. Tay, C. L. Poh, C. Goh, and R. I. Kitney, "A biological continuum based approach for efficient clinical classification," *J. Biomed. Inform.*, vol. 47, pp. 28–38, Feb. 2014, doi: 10.1016/j.jbi.2013.09.002.
- [82] J. Han, "Data Mining: Concepts and Techniques," p. 129, 2013.
- [83] A. Gupta, R. Kumar, H. Singh Arora, and B. Raman, "MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis," *IEEE Access*, vol. 8, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.
- [84] S. Das, S. Mishra, and M. R. Senapati, "New Approaches in Metaheuristic to Classify Medical Data Using Artificial Neural Network," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 2459–2471, Apr. 2020, doi: 10.1007/s13369-019-04026-y.
- [85] H. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, "A feature selection algorithm of decision tree based on feature weight," *Expert Syst. Appl.*, vol. 164, p. 113842, Feb. 2021, doi: 10.1016/j.eswa.2020.113842.
- [86] H. Benhar, A. Idri, and J. L. Fernández-Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Comput. Methods Programs Biomed.*, vol. 195, p. 105635, Oct. 2020, doi: 10.1016/j.cmpb.2020.105635.
- [87] M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi, and C. S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 559–566, May 2010, doi: 10.1109/TITB.2009.2038906.
- [88] M. Eshtay, H. Faris, and N. Obeid, "Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems," *Expert Syst. Appl.*, vol. 104, pp. 134–152, Aug. 2018, doi: 10.1016/j.eswa.2018.03.024.
- [89] N. Ghadiri Hedeshi and M. Saniee Abadeh, "Coronary artery disease detection using a fuzzy-boosting PSO approach," *Comput. Intell. Neurosci.*, vol. 2014, p. 783734, 2014, doi: 10.1155/2014/783734.
- [90] Z. Beheshti, S. M. H. Shamsuddin, E. Beheshti, and S. S. Yuhani, "Enhancement of artificial neural network learning using centripetal accelerated particle swarm optimization for medical diseases diagnosis," *Soft Comput.*, vol. 18, no. 11, pp. 2253–2270, Nov. 2014, doi: 10.1007/s00500-013-1198-0.
- [91] H. Turabieh, "A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease," *Am. J. Oper. Res.*, vol. 06, no. 02, pp. 136–146, 2016, doi: 10.4236/ajor.2016.62016.
- [92] E. JosÁ, "ECG arrhythmia classification based on optimum-path forest," *Expert Syst. Appl.*, p. 13, 2013.
- [93] P. Plawiak, "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system," *Expert Syst. Appl.*, vol. 92, pp. 334–349, Feb. 2018, doi: 10.1016/j.eswa.2017.09.022.
- [94] N. Paradkar and S. Roy Chowdhury, "Coronary artery disease detection using photoplethysmography," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, Jul. 2017, pp. 100–103, doi: 10.1109/EMBC.2017.8036772.
- [95] A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, Jan. 2017, doi: 10.1016/j.cmpb.2016.10.011.
- [96] R. Banerjee et al., "Time:Frequency Analysis of Phonocardiogram for Classifying Heart Disease," presented at the 2016 Computing in Cardiology Conference, Sep. 2016, doi: 10.22489/CinC.2016.165-189.
- [97] M. Kumar, R. B. Pachori, and U. Rajendra Acharya, "An efficient automated technique for CAD diagnosis using flexible analytic wavelet transform and entropy features extracted from HRV signals," *Expert Syst. Appl.*, vol. 63, pp. 165–172, Nov. 2016, doi: 10.1016/j.eswa.2016.06.038.
- [98] S. Patidar, R. B. Pachori, and U. Rajendra Acharya, "Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals," *Knowl.-Based Syst.*, vol. 82, pp. 1–10, Jul. 2015, doi: 10.1016/j.knsys.2015.02.011.
- [99] E. J. S. Luz, "ECG-based Heartbeat Classification for Arrhythmia Detection: A Survey," p. 23, 2016.
- [100] T. Li and M. Zhou, "ECG Classification Using Wavelet Packet Entropy and Random Forests," *Entropy*, vol. 18, no. 8, p. 285, Aug. 2016, doi: 10.3390/e18080285.
- [101] Q. Mastoi, T. Y. Wah, R. Gopal Raj, and U. Iqbal, "Automated Diagnosis of Coronary Artery Disease: A Review and Workflow," *Cardiol. Res. Pract.*, vol. 2018, pp. 1–9, 2018, doi: 10.1155/2018/2016282.

- [102]M. H. Song, J. Lee, S. P. Cho, K. J. Lee, and S. K. Yoo, "Support Vector Machine Based Arrhythmia Classification Using Reduced Features," p. 9, 2006.
- [103]S. Dalal and V. P. Vishwakarma, "GA based KELM Optimization for ECG Classification," *Procedia Comput. Sci.*, vol. 167, pp. 580–588, 2020, doi: 10.1016/j.procs.2020.03.322.
- [104]Z. Jiang and S. Choi, "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 286–298, Aug. 2006, doi: 10.1016/j.eswa.2005.09.025.
- [105]J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1169–1178, Apr. 2015, doi: 10.1109/TBME.2014.2377695.
- [106]A. Gavrovskaa, G. Zajić, V. Bogdanović, I. Reljin, and B. Reljin, "Identification of S1 and S2 Heart Sound Patterns Based on Fractal Theory and Shape Context," *Complexity*, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/1580414.
- [107]M. Feldman and S. Braun, "Description of Free Responses of SDOF Systems via the Phase Plane and Hilbert Transform: The Concepts of Envelope and Instantaneous Frequency," vol. 3089, p. 973, 1997.
- [108]H. Liang, S. Lukkarinen, and I. Hartimo, Heart Sound Segmentation Algorithm Based on Heart Sound Envelopogram, vol. 24. 1997, p. 108.
- [109]I. Maglogiannis, E. Loukis, E. Zafiroopoulos, and A. Stasis, "Support Vectors Machine-based identification of heart valve diseases using heart sounds," *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 47–61, Jul. 2009, doi: 10.1016/j.cmpb.2009.01.003.
- [110]S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a durationdependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010, doi: 10.1088/0967-3334/31/4/004.
- [111]S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, Dec. 2010, doi: 10.1016/j.eswa.2010.05.088.
- [112]F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. Abdullah Ramaiah, "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1407–1414, Oct. 2013, doi: 10.1016/j.compbimed.2013.06.016.
- [113]H. Her and H. Chiu, "Using time-frequency features to recognize abnormal heart sounds," in *2016 Computing in Cardiology Conference (CinC)*, Sep. 2016, pp. 1145–1147.
- [114]W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, Feb. 2017, doi: 10.1016/j.bspc.2016.10.004.
- [115]W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," *Expert Syst. Appl.*, vol. 84, pp. 220–231, Oct. 2017, doi: 10.1016/j.eswa.2017.05.014.
- [116]Z. Abduh, E. A. Nehary, M. Abdel Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers," *Biomed. Signal Process. Control*, vol. 57, p. 101788, Mar. 2020, doi: 10.1016/j.bspc.2019.101788.
- [117]B. Xu, D. Kocyigit, R. Grimm, B. P. Griffin, and F. Cheng, "Applications of artificial intelligence in multimodality cardiovascular imaging: A state-of-the-art review," *Prog. Cardiovasc. Dis.*, vol. 63, no. 3, pp. 367–376, May 2020, doi: 10.1016/j.pcad.2020.03.003.
- [118]G. N. Balaji, "Automatic Classification of Cardiac Views in Echocardiogram Using Histogram and Statistical Features," *Procedia Comput. Sci.*, p. 8, 2015.
- [119]V. Sudarshan, U. R. Acharya, E. Y. Ng, C. S. Meng, R. S. Tan, and D. N. Ghista, "Automated Identification of Infarcted Myocardium Tissue Characterization Using Ultrasound Images: A Review," *IEEE Rev. Biomed. Eng.*, vol. 8, pp. 86–97, 2015, doi: 10.1109/RBME.2014.2319854.
- [120]A. Jamthikar, D. Gupta, N. N. Khanna, L. Saba, J. R. Laird, and J. S. Suri, "Cardiovascular/stroke risk prevention: A new machine learning framework integrating carotid ultrasound image-based phenotypes and its harmonics with conventional risk factors," *Indian Heart J.*, vol. 72, no. 4, pp. 258–264, Jul. 2020, doi: 10.1016/j.ihj.2020.06.004.
- [121]Z. Akkus et al., "A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence-Powered Ultrasound for Improving Clinical Workflow," *J. Am. Coll. Radiol.*, vol. 16, no. 9, Part B, pp. 1318–1328, Sep. 2019, doi: 10.1016/j.jacr.2019.06.004.
- [122]L. K. Tan, R. A. McLaughlin, E. Lim, Y. F. Abdul Aziz, and Y. M. Liew, "Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression," *J. Magn. Reson. Imaging JMRI*, vol. 48, no. 1, pp. 140–152, Jul. 2018, doi: 10.1002/jmri.25932.
- [123]G. Xiong, D. Kola, R. Heo, K. Elmore, I. Cho, and J. K. Min, "Myocardial perfusion analysis in cardiac computed tomography angiographic images at rest," *Med. Image Anal.*, vol. 24, no. 1, pp. 77–89, Aug. 2015, doi: 10.1016/j.media.2015.05.010.
- [124]N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8221–8231, Nov. 2015.
- [125]H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
- [126]E. Avci, "A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10618–10626, Sep. 2009.
- [127]W. Guan, A. Gray, and S. Leyffer, "Mixed-integer support vector machine," in *Proc. NIPS Workshop Optim. Mach. Learn.*, Dec. 2009.
- [128]A. Khemphila and V. Boonjing, "Heart disease classification using neural network and feature selection," in *Proc. 21st Int. Conf. Syst. Eng.*, Aug. 2011.
- [129]J. Nahar, T. Imam, K. S. Tickle, and Y.-P.-P. Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 96–104, Jan. 2013.
- [130]J. A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, "Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system," *Appl. Soft Comput.*, vol. 20, pp. 103–111, Jul. 2014.
- [131]S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis: Evaluation for cardiovascular diseases," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4146–4153, Aug. 2013.