Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 1, pp. 187–203, DOI 10.12694/scpe.v25i1.2326

AN INSIGHT INTO VIABLE MACHINE LEARNING MODELS FOR EARLY DIAGNOSIS OF CARDIOVASCULAR DISEASE

MUKKOTI MARUTHI VENKATA CHALAPATHI*, DUDEKULA KHASIM VALI†, YELLAPRAGADA VENKATA PAVAN KUMAR‡, CHALLA PRADEEP REDDY§, AND PURNA PRAKASH KASARANENI¶

Abstract. Cardiovascular diseases (CVD) are a prominent source of death across the globe, and these deaths are taking place in low-to middle-income nations. Due to this, CVD prevention is a pressing issue that has already been the subject of extensive research. Innovative methodologies in machine learning (ML) can have a greater impact on the diagnosis of CVD, yet the research on CVD is more challenging and attracting more research indeed. In this paper, we investigate the differences between four distinct machine learning models, support vector machine (SVM), logistic regression, decision trees (DT), and artificial neural networks (ANN) in their classification accuracy and possible practicality in CVD classification. techniques such as ensemble learning and other model-specific optimizations are not part of this study, but more basic implementations of the various models were used. To implement abovementioned ML models, a subset of 14 features from the original heart disease dataset is considered and deemed relevant for classification where no individual feature data are missing. From the results, it is observed that there is no clear winner in the comparison of models. There is no significant difference in the average accuracy of models. The highest average hit rate is observed in SVM and ANN, however it is slightly lower in ANN. Even though the DT had lower accuracy, the fully trained model can be easily visualized and interpreted by humans. Hence, the DT is possibly the most practical model to use as a complement to doctors in their current methods of diagnosis.

Key words: Average Classification Accuracy, Cardiovascular Disease, Computer Aided Diagnostics, Machine Learning Algorithms, Test Data Split

1. Introduction. After covid-19 pandemic, the number of deaths due to cardiovascular disease (CVD) is increased worldwide. Over 26 million people have died of CVD in 2021 which corresponds to 39% of total deaths across the globe. Every year WHO suggests different methods to be followed to overcome the situation of CVD [1]. As per the WHO statistics, 75% of premature CVD, is avertable, and reducing risk factors can assist patients and healthcare providers to cope up with the observed high of CVD. The Interheart trial explicated the effects of CVD risk factors such as abdominal obesity, smoking, hypertension, diabetes, and dyslipidemia while demonstrating the beneficial advantages of eating fruits and vegetables and engaging in regular physical exercise. All demographics and socioeconomic levels evaluated had the same risk factors, demonstrating the validity of standardized methods for CVD primary prevention globally. Among them is the taxation of tobacco, better diet, and more physical exercise to name a few. However, these prevention methods need to be complemented together with early diagnostics and identification of high-risk patients which would have a substantial impact on public health [2].

Heart disease categories and learning machines have both been used to improve the heart failure analytical procedure. This research intends to investigate various machine learning methods and improve the utilization of healthcare data. The effectiveness of the classifier should increase. The circumstances of each person have an impact on their health problems, particularly the risk of heart failure (HF) rate. Machine learning techniques

*School of Computer Science and Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India (mmv.chalapthi@vitap.ac.in)
†School of Computer Science and Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India (d.khasimvali@vitap.ac.in)
‡School of Electronics Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India (pavankumar.yv@vitap.ac.in)
§School of Computer Science and Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India (pradeep.ch@vitap.ac.in)
¶Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, Andhra Pradesh, India (kpurnaprakash@kluniversity.in)

are more effective in predicting the risk factors and predicting the possibility of high blood pressure [3]. The majority of heart failure situations can be linked to problems with the heart's physiology or anatomy. As a result, HF has been associated with reduced life satisfaction and less effort put into engaging in mental and physical activities. According to estimates, HF affects 10% of the elderly and 1% to 2% of the general population in wealthy nations. As our population ages, heart failure is anticipated to become more common. After leaving the hospital, a readmission rate was present in patients with heart failure of 56.6%. Ignoring high frequency now will result in significant problems later on. Currently, reducing re-admissions is one of the most urgent needs.

After the initial incident, patients with heart failure are frequently kept in the hospital for a long period. Patients regularly get blood drawn to collect various health statistics [4]. It is possible to collect non-hematological data like gender, age, and smoking history. Machine Learning models intend to gain knowledge of the surroundings and forecast upcoming occurrences using user-provided data. People who possess this quality are flexible and may make decisions today based on their perceptions of the past. These models are utilized in the diagnostic procedure for several disorders, including multiple sclerosis. The main objective of this research is to create a learning model for prognosticating the possibility of heart failure. In this research on early diagnosis of heart failure, the data analysis employs standalone machine learning models decision trees, support vector machine (SVM), logistic regression, and neural networks [5, 6, 7]. Socio-demographic factors, like gender, high blood pressure, smoking status, and the existence of chronic medical disorders, have an impact on survival rates as well. It is challenging to establish reliable prognostic forecasts for persons with heart disease because survival rates for them are so unpredictable [8].

With the rise of machine learning and deep learning, new techniques of medical diagnostic methods have been opened up to complement the expertise of physicians. Many studies such as Parkinson's disease [9, 10], covid-19 [11, 12] have already been made on a range of medical diagnostic topics with different levels of accuracy [3]. Ihe internet of things also playing a key role in health care systems [13]. Deep learning concepts also extended their implementation in smart televison for program recommendation based on user's choice of priority [14].

The objective of this present study is to compare machine learning models' accuracy and their viability in cardiovascular disease detection. This paper compared the classification accuracy of the presence of cardiovascular disease on the implementations of four different broad machine learning techniques and their viability on this disease. This was tested on the Cleveland heart disease dataset from the UC Irvine Machine Learning Repository (UCIMLR) [2]. A comparison of classification accuracy and viability in the detection of CVD between the machine learning models. This problem has previously been investigated using different approaches. Models like SVM [5], artificial neural networks (ANN) [6], logistic regression and decision trees among others have been used on the same heart disease dataset from UCIMLR. These studies mainly focus on optimizing the selected model to achieve maximum accuracy. This study instead focused on comparing different machine learning models in their basic form and analyzing their weaknesses and strengths for this particular problem.

The use of machine learning algorithms has improved the prognosis for patients' diagnoses using their medical information. To detect cardiac sickness in its earliest stages, machine learning techniques based on linear, ensemble, and boosting are used. In this paper, we have implemented the early detection of cardiovascular disease on linear machine-learning models. The implemented models are trained and tested on the heart disease dataset. The test data split has been considered into 30% and 20% and the performances are ascertained with the impact of the test data split on the model accuracy.

The study used a subset of 14 features from the dataset which were used in all previous studies. The features include information such as age, sex, blood sugar, cholesterol, blood pressure, and more. The different models were then applied to the dataset by using already existing tools and frameworks such as TensorFlow [8] and sklearn [16]. The paper did not use any image analysis or image-based data which also could be used in the diagnosis since it was out of scope for this paper. The results were measured with a percentage of prediction accuracy for each respective model. The data were divided into training and test subsets where the models tried to predict the presence of CVDs in the previously unseen data of the test set after training on the training set.

Basic preliminaries and related work are described in the remainder of the paper, which is organized similarly to section 2. Section 3 presents a thorough implementation of machine learning methods for early

cardiovascular disease diagnosis. The empirical analysis is deconstructed in Section 4, which is followed by a discussion in Section 5. Section 6 contains the conclusion.

2. Basic Preliminaries and Related work. This section discusses the basic preliminaries and the literature works.

2.1. Computer-Aided Diagnostics. With the rise of computational power, computer-aided diagnostics (CAD) has become a part of medicine. One example is in the detection of breast cancer [15] on mammograms where CADs are routinely used as a second opinion [17]. One commonly used technique for CAD is machine learning (ML). As data-gathering advances, the effectiveness of ML as a CAD has progressed and is predicted to have a significant effect on medicine [3]. In broad terms, CADs using ML perform analysis on a great amount of patient data with and without a disease. The data consists of attributes that are considered to have a contribution to this disease and are often called training data. Using this training data the ML model is modified slightly by each data sample resulting in a model that can, to some degree, correctly classify the samples in the training data. This model is then tested on test data which is the same kind of data as training data but previously unseen to the model. The caliber of the model is protracted by the classification accuracy of the test data.

2.2. Earlier techniques for diagnosing cardiovascular disease. Most applications rely on common clinical risk factors like diabetes and hypertension. Research has also been conducted in Brazil on locations that are at high risk for cardiovascular disease-related mortality. In addition, the study's analysis revealed that 50% of the participants had several, difficult-to-treat illnesses that raised their mortality risk. This subject hasn't been the subject of enough study [19]. It can be more difficult to set a baseline for normalcy when lab results are assessed more realistically. For calculating risk, using trustworthy survival models is essential [20].

Traditional proportional risk models can uncover more relevant predictors by automatically creating connections between their component values and large data response values using computational methods [1]. The medical industry has used a variety of machine-learning methods, including decision trees [22]. Nonparametric survival is an auxiliary for parametric and semi-parametric models since they are independent of time in characterizing relationships [23]. The ensemble approach and survival trees combine to produce more precise projections. The common CVD diagnostic methods are:

- Physical Examination - Patient medical history and lifestyle Blood pressure, patient weight, listening to heart and lungs, etc.
- Blood tests - Level of different electrolytes, proteins, and other biomarkers.
- Chest X-rays - Image of the chest, revealing enlarged heart, lung congestion, and other abnormalities [21].
- Echocardiography - Ultrasound images of a heart showing heart structure.
- Cardiac catheterization - X-ray images reveal if there are any blockages in the heart.
- Radionuclide Ventriculography - images that demonstrate how well the heart's blood supply and chambers are functioning.

There are multiple indicators or methods that physicians can use and combine in the diagnosis of CVD. Some of them require analysis of images after the test has been made and some do not. The CVD diagnostic methods contain different methods of diagnostics together with what the method result type is and if this paper uses any of the data the respective method generates in the training of the models. In our proposed research work, we have used physical examination, and blood tests to examine the dataset. This indicates what types of tests medical personnel must do to be able to use the models described in this paper. Note that Used does not mean every data point from the test is used. The detailed view of CVD data the models are trained in is presented in Table 3.1.

2.3. Machine Learning Models. Machine Learning models can be used as CAD tools. However, there exist many different ML models that can be used to perform CAD. The models vary in complexity and how well they perform on different kinds of data. There are various applications of machine learning techniques such as electric vehicle [24], smart home [25], and microgrid [26]. This section will describe three frequently used models that were also used in this study.

2.3.1. Support Vector Machine. The technique of using so-called support vectors in the data to find the widest margin between two classes [20]. With linearly separable data consisting of data points from two distinct classes, to find a hyperplane separating the classes with a considerable margin to the data points of each class, Support Vector Machine could be utilized [19]. Thus creating an optimal linear decision boundary given the data. The problem was that most data is not perfectly linearly separable which made the technique hard to use. The solution to this came in the 1990s when the so-called kernel trick was discovered. This made it possible to use Support Vector Machine for non-linearly separable data by applying a non-linear transformation on the data and finding a linear decision boundary on the transformed data [19]. Another development was the concept of soft margins. This made it possible to find solutions in data that is not linearly separable even after applying a kernel transformation by allowing a few points to lie on the other side of the edge. Soft margins also had the added benefit of allowing a few points to lie within the margin even though the data is linearly separable to gain a larger margin.

2.3.2. Decision Tree. Decision trees (DT) are used for predicting what target class or value an observation belongs to. When building a decision tree given some data, all possible values for the features are divided into several regions. Each region has a specific target class assigned to it. An observation, with specific values for all features, is then given its target class or value depending on what region it falls in, so it can be used as Regression and Classification [19]. Depending on the number of features the Decision Tree will have a different amount of region layers. The first feature splits the Decision Tree into regions J1 to Jn, then depending on the values of another feature each region Jk is split into regions Rk to Rm and so on. Because of this attribute, Decision Trees can easily be represented as trees. Here the observations have two features X1 and X2 and are split into five final regions. This entails that Decision Trees often can easily be interpreted by a human.

2.3.3. Artificial Neural Networks. The thought behind ANN is to create a model that resembles a nervous system that, in humans and animals, can solve complex problems with the help of a network of many interconnected cells [22]. ANN consist of an input layer where the values of the observations are fed forward to one or several hidden layers of neurons. Depending on the values from the input layers, weights, and thresholds of each neuron, different neurons will fire and result in a specific output value which ends up in the output layer. The process of generating an output value given observation is called forward propagation [22]. One way of training ANN is to modify the weights of the neurons to minimize the error of ANN that have given training data. The minimizing is done numerically by using gradient descent which aids in finding a local minimum of the error function. This process is called backpropagation [22].

2.3.4. Logistic Regression. Logistic regression can be employed for binary classification by shaping the probability of data points' affinity to a certain class, e.g. providing an anticipation of a person being healthy. This modeling is broadly done by regressing a sigmoid function (or an alternative function with a bound of 0 to 1) to a dataset [32]. This regression can be performed both with single variables and collective variables and the prompted function will need an input vector x with dimension n the count of features in the dataset. The regression treasure trove optimal values for the parameter vector β in the sigmoid function. Logistic regression is an elementary approach to classification that is speedy. The classifier deteriorates however when regressing composite relationships [3].

2.3.5. Cross-validation. A cross-validation is an approach where a model is trained and validated on the same data, proving a validation dataset unnecessary. This is achieved by splitting the data toward K different folds, the model is formally trained on K-1 of the folds and computed on the remaining fold, this has recurred K times for each fold and the median of the results is reported [36]. For implementing the proposed work, ten folds are considered. Cross-validation can be calculational expensive but erases the necessity of a validation dataset and more data can then be employed for training and testing.

2.4. Related work. The area of machine learning in health care is commonly split into physical and virtual implementations, where virtual experimentation mainly concentrates on analyzing patient data. Contemporary articles that reveal virtual implementations for analyzing data to predict heart failure will be illustrated in this section.

Machine learning has from an early stage been used for computer-aided medical diagnostics [3]. In this

field, CVD classification has been a prominent problem that has been thoroughly researched with a considerable amount of time being spent on optimizing different models.

K-nearest neighbor (KNN) and SVM, supervised learning algorithms are pragmatic to predict heart disease [33, 35]. The researchers contend that machine learning models can be cast-off to forecast cardiovascular diseases [35]. Numerous other algorithms have been investigated, such as random forest and logistic model tree (J48). The Cleveland Heartland Registry is used by UCIMLR researchers to validate heart disease in patients and screen them for it. The following categories of information are included in this dataset. Next, a recommended large-scale classification algorithm will be made. To provide patients with more precise diagnoses, machine learning can be utilized to discover connections and predict risk factors for occurring heart disease.

For patients with heart failure [34], machine learning appears to be able to predict survival duration with accuracy. Ejection fraction and blood creatinine readings are both excellent predictors of a patient's life expectancy, according to patient charts. With the aid of a novel technique, survival rates in heart failure patients might be predicted [35].

2.4.1. Results with Optimization of Classifiers on Heart Disease Data. A new measure of classification reliability guaranteed 100% accuracy on all patient data points it can classify, with the drawback that possibly only a subset of all the data points could be classified [5]. To test this proposed measure of classification reliability they used a Support Vector Machine on the UCIMLR heart disease dataset [19]. This resulted in an average classification accuracy of 73% with radial basis function (RBF) kernel and the polynomial of degree is 73.7%.

Optimizing accuracy is often the utmost sought-after end goal but there has been researched done trying to generate a greater accuracy while still keeping the model simple enough to be able to interpret it. Freund and Mason [17] propose a solution to this that they call an alternating decision tree (ADT) which achieves similar or better accuracy than C5 and is easy to interpret. Whereas the C5 decision tree implementation produces good accuracy at the cost of interpretability. The ADT achieves a classification accuracy of 83.0% which is better than the C5 classification accuracy of 79.8% on the same dataset.

The usage of learning models for heart disease prediction is discussed in this article [31]. Data analytics were employed in this study to look at heart disease. The researchers carried out a study to assess the reliability and accuracy of three different data analysis methods such as KNN, ANN, and SVM. The neural networks produce better results and enable faster model learning with 93% accuracy. The researchers demonstrated the use of machine learning to solve the problem of heart disease prediction [29]. Researchers from Stanford University have created a revolutionary algorithm that predicts a patient's likelihood of acquiring heart disease based on their medical history. Heart disease patients' outcomes were categorized and predicted using machine learning techniques like KNN and logistic regression. These actions led to the creation of a model that can predict cardiac events in a large population more accurately. The main advantage of this approach is the time and effort savings in assessing whether a classifier can accurately diagnose heart illness. You can save time and money by using the provided method for heart disease prediction. It is feasible to conclude the long-term health of people with heart disease. In this study, machine learning is utilized to identify key components for the detection of cardiovascular disease. A prediction model is built using several features and classification techniques. The authors assert that by uniting a random forest and a linear model, a 92% accuracy rate can be reached.

One of the highest accuracies on the heart disease dataset was achieved by [6] using an ensemble of independent multi-layered feed-forward neural networks. The accuracy achieved was 89.01% using three neural networks. Using an ensemble of more than three neural networks did not show any improvements. This paper uses models based on the same principles as the models mentioned above, however, instead of attempting to optimize a single model it investigates how they compare to each other in terms of classification accuracy and viability for cardiovascular disease.

3. Early diagnosis of cardiovascular disease using Machine Learning Methods. This section explains cardiovascular disease prediction and the process of constructing the feature dataset for the numerical experiments.

3.1. Dataset. The scope of the research primarily relies on a single dataset and it has been used extensively in research of computer-aided diagnostics [6, 7, 30]. Containing only 297 patients with 14 features makes it hard to draw any wide-stretching conclusions on which of the compared techniques is best suited for CVD diagnostics. The original dataset obtained from UCIMLR contains 76 features where the occurrence of heart disease is an integer-valued feature from 0, no presence of disease, to 4. All values greater than 0 indicate the presence of heart disease. All previous ML research on this particular dataset has not been done on all 76 features, but a subset of 14 features deemed relevant for classification and where no individual feature data was missing. The prediction values have then also been binary instead of integers valued up to 4 since it is out of scope to diagnose different kinds of CVD. The binary value will indicate angiographic disease status where value 0 corresponds to <50% diameter narrowing and value 1 corresponds to >50% diameter narrowing. This is called the "processed dataset", henceforth in this paper, it will be referred to as "dataset" and is the dataset that was used for this research work.

The different sources have varying degrees of missing features from some data samples. For example, almost all data samples from the Hungarian source are missing data from three features, among them the number of major vessels colored by fluoroscopy. The data samples from the other data source are missing even more features like blood sugar. Because of this, in this research, we have only used the data donated from the Cleveland source since that data contains very few data points with missing features.

The Cleveland source contains 303 samples in total with all features listed in Table 3.1. From these 303 samples, 6 samples have been removed because of missing values. The choice of using the same dataset and not cherry-picking data points from other datasets was made so the comparison between previous research would be fairer. The dataset was not altered or preprocessed in any other way and all models used the same data. This was done so the comparison is as fair as possible even though some models might benefit from further preprocessing. For example, artificial neural networks might benefit from bundling up the ages into ranges of ages (35-45, 45-60).

3.2. Implementation. This section will present the implementation of the proposed early diagnosis of cardiovascular disease. The different machine learning models used the same method for splitting the dataset into training data and test data. The data were randomly shuffled and then split into two categories. The training data was used to train/fit the different models to the problem and the test data was then used to assess the performance of the model to see how well it performs on new unseen data. The test size is a parameter that represents the proportion of the data that is used as test data compared to training. For example, a test size of 0.3 means training data is 70% and testing data is 30%. Because of the limited size of the dataset, only 30% and 20% of test data splits were tested, which represents the training sets of 70% and 80% respectively. A larger split would lead to even fewer training points and a lower one would lead to testing with less than 60 patients which increases the risk that a model that performs well was only "lucky".

The Python library is employed to implement each specific model. Logistic regression was executed using the class LogisticRegression from the linear_model package, neural network utilizing the class MLPClassifier from neural_network, decision tree applying the class DecisionTreeClassifier from the tree, and support vector machine harnessing the class support vector classification (SVC) from the SVM package. The linear regression, decision tree, and neural network models were all trained on the entire training dataset. The SVM model however was trained on a chunk of the training data due to poor performance, more on this in Discussion.

3.2.1. Decision Tree. The DT implementation was made using sci-kit learn, a machine learning library that has pre-made ML models [16]. Specifically, the DT classifier [23] from sci-kit learn was used for generating DT. The DT classifier uses an optimized version of the CART algorithm which is a high-performing general-purpose algorithm for building decision trees [27]. When building a Decision Tree many different parameters can be set which decide how the resulting tree will look and perform. The maximum depth of the tree, and the number of samples to consider when splitting a node, are used when split among others. In this study, only the max depth of the tree and the criterion for splitting were taken into consideration. These parameters were chosen as they have to most impact on the complexity and structure of the generated decision tree.

3.2.2. Support Vector Machines. The SVM was implemented using sci-kit learn, the same library used for decision trees. As described, SVMs use different kernels to transform the data to find optimal support

Table 3.1: CVD Dataset Features

S. No.	Feature	Description
1	Age: age in years	Integer value
2	Sex	1 = male, 0 = female
3	CP: Chest pain type	0: typical angina 1: atypical angina 2: non-anginal pain 3: asymptomatic
4	trestbps: Resting blood pressure	mm Hg
5	chol: Serum cholesterol	mg/dl
6	fbs: Fasting blood sugar >120 mg/dl	1 = true, 0 = false
7	restecg: Resting electrocardiographic results	0: Normal 1: having ST-T wave abnormality 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8	thalach: maximum heart rate achieved	Integer value
9	exang: Exercise induced angina	1 = yes, 0 = no
10	oldpeak = ST depression induced by exercise relative to rest	Floating value
11	slope: the slope of the peak exercise ST segment	0: upsloping 1: flat 2: downsloping
12	ca: number of major vessels (0-3) colored by flourosopy	Integer Value 0-3
13	thal: Thalassemia	0: Normal 1: Fixed defect 2: Reversible defect
14	Condition	0 = no disease 1 = disease

vectors. The implementation of SVM include the usage of three different kernels, a linear, a polynomial, and a RBF [28]. Both the polynomial and radial basis function kernel can be modified by input parameters. When using the polynomial kernel, the degree of the polynomial, which determines the complexity of the model, needs to be specified. This study used polynomials of degree two and degree three. The RBF kernel requires a parameter gamma which specifies how much a single training point influences the kernel. Here, the default value of Nl. Lastly, there is C, which regulates how wide the margin between the different classes is allowed to be. A high value of C results in a strict margin that does not allow data points of different classes to be on the wrong side of the margin. As data from the dataset used most likely is not linearly separable a lower value of C is preferred for the classifier to be able to create a margin. Therefore C=1 was used in this study.

3.2.3. Artificial Neural Networks. All experimentation with ANN was conducted using the TensorFlow library [8]. TensorFlow is one of the most widely used libraries for building machine learning models in production. The library makes it easy to build and test different types of networks and network structures with different hyperparameters. The optimizer used for network training in this work was the optimizer with a batch size of 32 and a cross-entropy loss function.

Many different types of hyperparameters can be hard to set for artificial neural networks even when the type of network is known. These hyperparameters include the number of layers, the number of nodes in the hidden layers, the type of activation functions the number of training epochs in the training of the model.

This paper used a sequentially built feed-forward network. The input layer has the same size as the 13 features in the dataset and the rectified linear (ReLU) activation function used for the input layer. The ReLU activation function which has been proven to give better results than the sigmoid activation function [29]. The

output layer is a single node that is activated with the sigmoid function since it is a single binary output the network produces. There is no exact formula for selecting the number of hidden layers. In general, if the data are less complex 1 or 2 hidden layers are used, if the data are complex 3 or 5 hidden layers are used. When the number of hidden layers are increased it increases the capacity of the model, simultaneously it requires more samples. Every hidden layer within the network also uses the ReLU activation function. Each network was trained for 350 epochs. The structure of the network varied with different tests to examine which structure fits the problem best. The tested structures were selected in an ad-hoc way where rough testing first was done before selecting some of the best-performing structures. Note that the network depth includes the input and output layers in the results window.

3.3. Evaluation. The machine learning models were assessed based on the mean and standard deviation of average classification accuracy they have on the test data. After the model has been trained it is presented with previously unseen data samples from the test dataset and it then tries to forecast the presence of CVDs and then checks if the prediction was correct. The accuracy then becomes the ratio of correct predictions with the test dataset. As mentioned in previous sections the different methods depend on varying hyperparameters that can be tweaked and tuned and yield different results. Each model was implemented with multiple different hyperparameters to get a better overview of the model potential.

There are also elements of randomness in which data points the models get to train on depending on the train/test data split. The SVM and ANNs also depend on randomly initialized weights for their models which might affect the outcome of the training. To remedy this the training and evaluation of each implementation were done in 100 independent runs and the classification accuracy is assumed to follow a Gaussian distribution. This makes it possible to compare not only the mean (μ) but also the standard deviation (σ) of each implementation to see which implementation had the best consistency. Metrics employed in this study to correlate machine learning models are illustrated in this section. Multiple metrics are accustomed to evaluating the accomplishment of a network [28]. Exploring different metrics for a variety of tasks is meant to represent the network's ability to clarify a given problem. To define all metrics the listed terms will be explored: The evaluation metrics utilize true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

The accuracy [28] ratio of cases that were accurately identified as a percentage of all instances can be calculated as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Precision [28] ascertains with what precision the model fits in the positive instance category. Precision is calculated as follows.

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

Recall [28] stipulates how many positive occurrences the model recorded. The recall is calculated as follows.

$$Recall/Sensitivity = \frac{TP}{TP + FN} \tag{3.3}$$

Precision expresses the number of positive guesses that were true. Sensitivity expresses the proportion of all positive illustrations that were apparently labeled. Specificity expresses the proportion of all negative examples that were labeled.

$$Specificity = \frac{TN}{TN + FP} \tag{3.4}$$

F1 score [28] is a metric that combines precision and sensitivity to a single metric. It is defined as follows.

$$F1Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{3.5}$$

Table 4.1: Average Classification Accuracy of Different Classifiers

Classifier	**Test Data Split in %**	**μ on Classification Accuracy %**	**σ on Classification Accuracy**	**Classification Accuracy (Max) in %**	**Classification Accuracy (Min) in %**
Decision Tree	20	76.53	0.053	88.33	61.66
Decision Tree	30	76.18	0.040	85.45	65.55
Logistic Regression	20	92.85	0.050	92.44	71.11
Logistic Regression	30	87.86	0.051	89.44	68.11
Support Vector Machine	20	83.18	0.044	91.66	70.00
Support Vector Machine	30	83.01	0.035	92.22	71.11
Artificial Neural Networks	20	82.16	0.042	91.66	66.66
Artificial Neural Networks	30	80.64	0.040	92.22	67.7

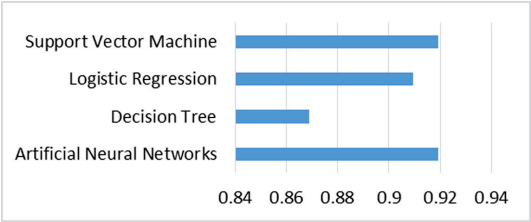

Fig. 4.1: Average of classification accuracy (maximum) by classifier

A receiver operating characteristic (ROC) curve is a graph that plots sensitivity and 1-specificity [18]. A good model has high sensitivity and specificity. A procedure to measure how well the model does this is to consider the area under the curve (AUC). A lofty AUC is better than a squat AUC. The top right of the graph used to gauge performance is where a model attempts to achieve high precision and high recall.

4. Experimental Results Analysis. The computer employed for the experimentation has an i5 processor at 1.8 GHz, with 8 GB of RAM, and a Windows 8.1 OS. The original data set [2] was exclusively used for classification experiments without any preceding feature selection technique. The investigation of these phases is described in full, along with a description of the results, in the immediate sections. In our research, two test data splits 30% and 20% are considered to test the accuracy of the machine learning models on test data.

Table 4.1 shows a general overview of the best average classification accuracy results of the different classifiers. The logistic regression has the highest mean (μ) classification accuracy 92.85 in both 30% and 20% of test data split together with the lowest $\sigma = 0.051$. The logistic regression and SVM also shared the highest accuracy 92.44% and 92.22% with ANN and the logistic regression and SVM had the highest lowest value 71.11%. The DT had the highest standard deviation of measured accuracy $\sigma = 0.053$. Figure 4.1 depicts the average of classification accuracy (Maximum) by the classifier.

4.1. Performance Evaluation on Decision Tree. Decision trees produce in all cases an average classification accuracy above 70% with the best mean result being an accuracy of 81.11%. As Table 4.2 indicates the depth of the tree has a greater impact on the classification accuracy than the splitting criterion. Setting a limit on the depth to 6 produces improvement in the classification accuracy compared to using the same test data split and criterion with no limit on the tree depth. Figure 4.2 is a graph that represents the average classification accuracy. Figure 4.3 shows what the decision tree looks like for the test with criterion entropy

Table 4.2: Average Classification Accuracy of Decision Tree Classifier

Max Depth	**Split Criterion**	**Test Data Split**	**μ on Classification Accuracy in %**	**σ standard Deviation on Classification Accuracy**	**Classification Accuracy (Max) in %**	**Classification Accuracy (Min) in %**
None	Gini index	30	74.56	0.043	83.33	61.11
None	Entropy	30	73.49	0.039	84.44	63.33
None	Gini index	20	72.23	0.049	86.66	60.00
None	Entropy	20	72.41	0.051	83.33	60.00
4	Gini index	30	80.73	0.045	85.55	64.44
4	Entropy	30	81.11	0.040	85.55	65.55
4	Gini index	20	74.62	0.054	86.66	56.55
4	Entropy	20	79.99	0.053	88.33	61.66

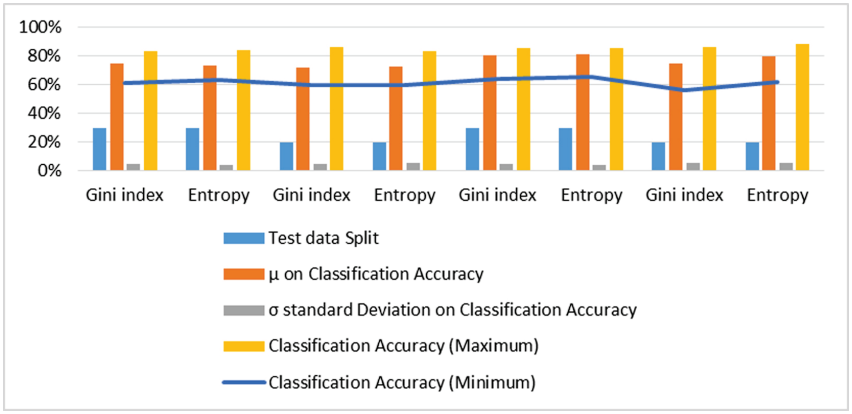

Fig. 4.2: Graph representation of average classification accuracy

and max depth 6 to classify a patient the Decision Tree first splits on the attribute thalassemia. This indicates that thalassemia might be one of the most influential factors of heart disease.

4.2. Performance Evaluation on Logistic Regression. Table 4.3 presents classification accuracy achieved with the different Regularization used for the Logistic Regression ranges from around 75% to around 91%. Table 4.3 shows that the regularization of Lasso performs well and has the highest μ accuracy of 93.18%, which is the highest overall mean classification accuracy achieved. At the same time the Ridge results in the lowest recorded accuracy, at 87.33%, across the different classifiers. Figure 4.4 shows the average classification accuracy of the logistic regression classifier.

4.3. Performance Evaluation on Artificial Neural Networks. Table 4.4, presents the shallower network had, in general, a higher μ classification accuracy than the deeper one for both 30% and 20% test data split and all hidden layer sizes. The networks had a higher μ together with a lower standard deviation σ for the test data split of 30% compared to the same type of network with a split of 20%. In general, the less complex networks with either smaller hidden layer sizes and/or shallower networks also had a higher μ and lower σ. The

Table 4.3: Average Classification Accuracy of the Logistic Regression Classifier

Regularization Type	**Test Data Split in %**	**μ on Classification Accuracy in %**	**σ standard Deviation on Classification Accuracy**	**Classification Accuracy (Max) in %**	**Classification Accuracy (Min) in %**
Lasso (L1)	20	93.18	0.054	91.66	70.00
Lasso (L1)	30	92.31	0.047	93.33	72.22
Ridge (L2)	20	88.96	0.054	87.77	67.77
Ridge(L2)	30	87.33	0.049	91.66	68.33

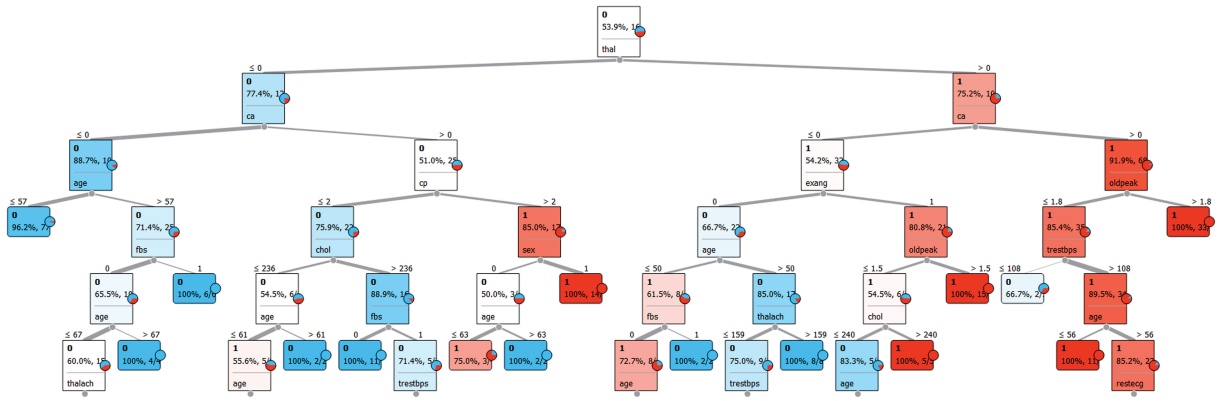

Fig. 4.3: The first six levels of the trained decision tree

network depth of 3 with 20% test data split and hidden layer size 14 had the highest μ. The corresponding network with a split of 30% had the most consistent results ($\sigma = 0.039$) and the highest accuracy (92.22%) came from the network with 4 layers with layer size 14 and 30% test data split.

4.4. Performance Evaluation on Support Vector Machine. The degree of classification accuracy attained using various kernels used for the SVM ranges from around 65% to around 83%. Table 4.5 shows that the polynomial degree of 2 kernels performs well and has the highest μ accuracy of 83.18%, which is the highest overall mean classification accuracy achieved. At the same time the radial basis function kernel results in the lowest recorded accuracy, at 65.61%, across the different classifiers. The kernels linear and RBF are having no degree information and is represented as not available (NA). This degree sometimes may not be available when there is no instance of the graph crossing the x-axis.

4.5. Cardiovascular Disease Detection using Machine Learning Models. The ML algorithms are trained and tested on two different test data split 30% and 20%. Cardiovascular disease detection on the four machine learning models is retrospect values measured by computing the sensitivity, precision, and the F1-score of each model. A ROC curve was depicted for all models. Figure 4.5 constitutes the Precision, specificity, sensitivity, and F1-score, achieved on the test data, for each model.

As shown in Table 4.6, the DT has achieved the highest precision and F1 score. Apparently, neither algorithm labeled a no disease instance incorrectly nor the logistic regression correctly predicted 85.8% of the heart disease cases while the artificial neural network truly predicted 84.5%.

Figure 4.6 represents the participation of each feature of the cardiovascular disease database on four different

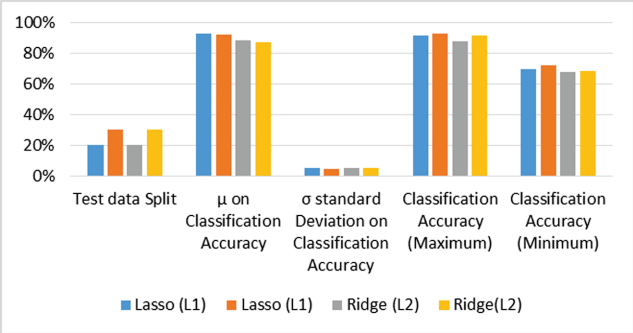

Fig. 4.4: Average Classification Accuracy of the Logistic Regression Classifier

Table 4.4: Average Classification Accuracy of the Artificial Neural Networks Classifier

Depth	**Hidden Layer Size**	**Test Data Split in %**	**μ on Classification Accuracy in %**	**σ standard Deviation on Classification Accuracy**	**Classification Accuracy (Max) in %**	**Classification Accuracy (Min) in %**
4	128	30	77.33	0.045	90.00	66.66
3	128	30	78.29	0.053	86.66	64.44
4	128	20	78.38	0.051	91.66	66.66
3	128	20	79.75	0.054	91.66	65.00
4	14	30	80.64	0.040	92.22	67.70
3	14	30	80.63	0.039	91.11	68.88
4	14	20	80.48	0.044	90.00	66.66
3	14	20	82.16	0.042	91.66	70.00

Table 4.5: Average Classification Accuracy of the Classifier Support Vector Machine

Kernel	**Degree**	**Test Data Split in %**	**μ on Classification Accuracy in %**	**σ standard Deviation on Classification Accuracy**	**Classification Accuracy (Max) in %**	**Classification Accuracy (Min) in %**
Linear	NA	20	82.56	0.041	93.33	75.00
Linear	NA	30	83.01	0.035	92.22	71.11
Polynomial	2	20	83.18	0.044	91.66	70.00
Polynomial	2	30	82.31	0.037	93.33	72.22
Polynomial	3	20	78.96	0.044	91.66	68.33
Polynomial	3	30	77.33	0.039	87.77	67.77
Radial Basis Function	NA	20	65.61	0.049	75.00	51.66
Radial Basis Function	NA	30	65.72	0.042	76.66	56.66

Table 4.6: Precision, Recall, and F1-score, achieved on the test data, for each model

Model	Precision	Recall	F1 Score
Logistic Regression	85.9	85.9	85.8
Artificial Neural Networks	84.5	84.5	84.5
Decision Tree	95.3	95.3	95.3
Support Vector Machine	69.2	67.0	66.7

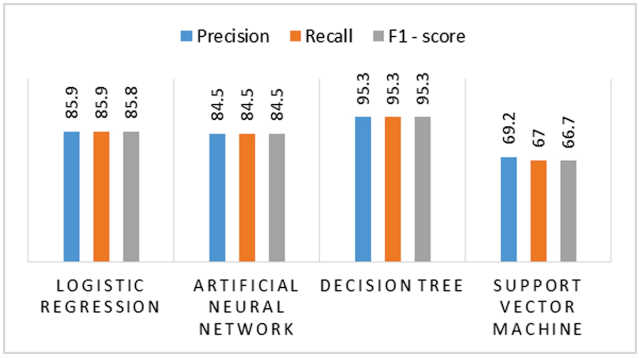

Fig. 4.5: Precision, Recall, and F1-score, achieved on the test data, for each model

ML algorithms. The values presented in the graph are mean, range, error rate, and hit rate on accuracy in the presence of disease across the attributes.

The DT earned a precision of 95.3% and a sensitivity of 95.3%. Truly predicting 95.3% of all cardiovascular occurrences but a false prediction of 5% of non-cardiovascular instances. The SVM realized a precision of 69.2% and 30% of predicted cardiovascular instances were not diagnosed.

The AUC is 0.923 for the logistic regression, 0.925 for the ANN, 0.989 for the DT, and 0.231 for the SVM. The erected value for AUC was achieved by the DT, followed by the ANN. Discover that the ROC curve for the DT classifier is a raw estimate due to the discrete essence of the classifier. The classifier only spawns a single false positive rate (FPR) and a single true positive rate (TPR) resulting in an atomic point on the graph, the ROC curve is estimated by combining this point with the points (0,0) and (1,1). The other algorithms rely on a continuous value and a threshold to classify data, this threshold can be altered to generate an endless relationship betwixt the FPR and TPR, which is depicted by the graph. The following PR curves were produced by the models on the test data.

4.6. Performance Differentiation of Accuracy on Machine Learning Algorithms. When applied to the whole dataset, DT can be observed to achieve 95.3% lofty accuracy than additional classification methods, while SVM achieves a modest accuracy of 67%. Figure 4.7 depicts the Accuracy of ML algorithms on cardiovascular disease data.

5. Discussion. The biggest benefit of the DT model is the interpretability of the resulting trained DT. The complete model can be visualized. This means that the results can be used by physicians as a complement to the current methods of diagnostics and patient examination. The tree in Figure 4.3 suggests that thalassemia has the biggest impact on the prediction of the presence of CVDs in the patient. This information can for example be used to make sure that all physical examinations of patients when checking for potential CVDs

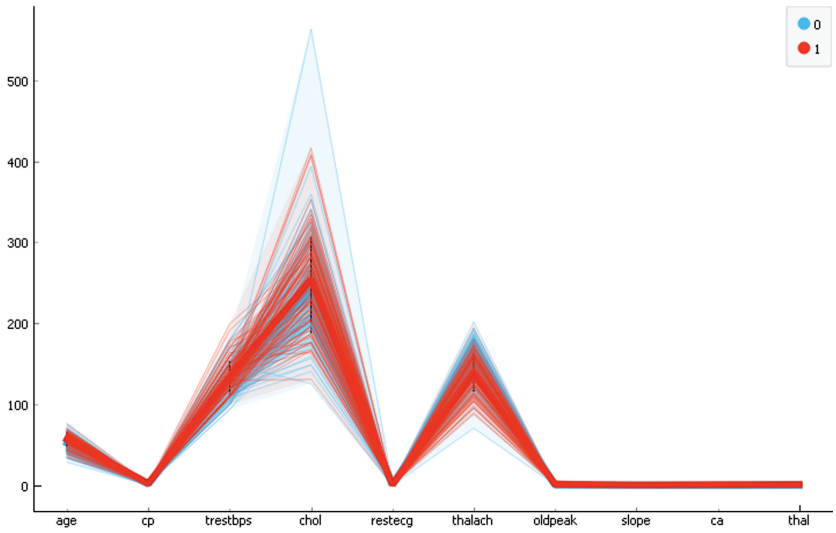

Fig. 4.6: Representing the mean, range, Error rate, and hit rate on features of the cardiovascular disease database

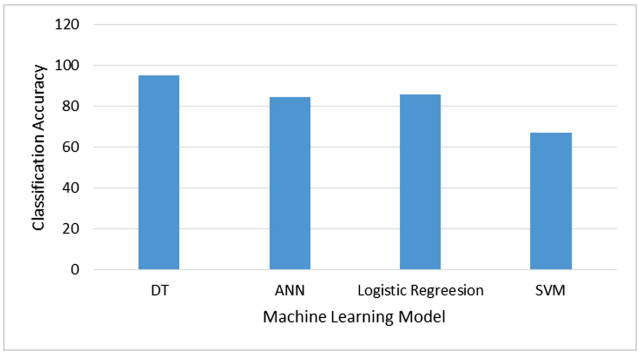

Fig. 4.7: Accuracy of machine learning algorithms on cardiovascular disease data

should include a test of patient thalassemia. This makes Decision Tree a very viable initial technique for CAD of CVDs since the risks are low and the final responsibility of diagnosis still firmly lies in the hands of the physicians and not in a yet unproven autonomous method of diagnosing CVDs. Not limiting the tree depth of a Decision Tree generates deep and complex trees that result in high variance and lower classification accuracy because of model overfitting. Allowing deeper or shallower trees than 4 levels did not improve the results as can be seen in table 3. As for the splitting criterion gini index performs slightly better than entropy however the difference is small and could depend on other factors.

ANN performed well where some individual runs gave a model with up to 92.22% accuracy. The best-averaged performance came from the 3-layered version with a hidden layer size of 14 and 20% test data split. This suggests that the higher complexity models might suffer from overfitting since the dataset only consists of 13 features with binary output. A single hidden layer that is not much larger than the input layer seems to suffice for this type of dataset. The higher test data split can also be an indicator of overfitting since a higher split gives the model fewer data points to train on which will keep the bias higher. Preliminary testing with the number of epochs also showed that more epochs than 350 almost always produced a worse result also indicating problems with overfitting. The standard deviation was generally in line with the SVM implementation indicating a more reliable result than the Decision Tree. The reliability is necessary for actual use on real patients but the dataset is too limited for it to translate to the general reliability of classification accuracy. More data points are needed with more features and with more diversity before the Artificial Neural Networks are truly ready as an autonomous diagnostics tool since the models themselves are not interpretable. How many data points and how diverse the data needs to be is hard to answer but probably in the size of millions of patients from around the globe and where medical experts can be a part of the feature selection.

Regarding SVMs, the two most complex kernels, with polynomial degree 3 and RBF, are the worst performing while the other two simpler kernels perform the best. This indicates that the worst-performing kernels are too complex for the data and that the kernels overfit the training data resulting in a lower classification accuracy just like the deeper ANN models. The highest mean accuracy by a small margin came from the 2nd degree polynomial kernel with a 20% test data split. However, it should be noted that the polynomial kernel had remarkably longer training time than the other kernels, especially compared to the linear kernels. This in combination that the 2nd degree polynomial kernel with 20% had a lower accuracy than the best linear kernel (with test data split at 30%) and a higher standard deviation suggesting that a linear kernel is the best choice for this particular dataset. Both in actual results previously mentioned and in practicality because of the much faster training time. The SVM suffers from the same problem of interpretability as ANN and a lack of data for autonomous use.

In general, the more complex models seem to have fared worse than the lower-complexity models. The deeper neural networks with larger hidden layers and higher complexity kernels in SVM all have lower mean accuracy than the shallower neural network and linear or low-degree polynomial kernels of the SVM. This does not necessarily mean that the problem of diagnosing CVDs can truly be described with a low-complexity model, a high-complexity model might very well outperform the lower-complexity ones with a large enough dataset. The SVM with a linear kernel and test data split of 30% is one of the best-performing models and fast to train however due to the limitations of the dataset it is very hard to know how well the model generalizes for other unseen real-world data. It is possible that the patients that have CVDs in the dataset used to share a hidden feature value, for example, smoking habits as previously mentioned, and that the SVM model instead finds the connecting relationship in another irrelevant feature. The range of classification accuracy for the various regularizations employed in the logistic regression is approximately 75% to approximately 91%. The regularization of Lasso works effectively and achieves the highest overall mean classification accuracy of 93.18%, as shown in Table 4.3. Ridge had the lowest documented accuracy of all the classifiers, at 87.33%, during the same period.

Because of the limitations in the dataset, the focus on mean accuracy lessens, especially since the different models also had very similar mean accuracy. The comparison then instead focuses on the interpretability of the model and viability as a tool for physicians. There are too many uncertainties for the resulting models in this paper to be used autonomously without a pathologist's supervision. A model that can be interpreted by a physician and used as a complement can be used even with its limitations previously described. This makes DT viable as a tool today. The SVM and ANN should be used with caution. They can of course also be used as complementary verification that the physicians can use but they should not be treated as equally accurate as the pathologists' judgment.

5.1. Comparison with earlier research. The previous studies mentioned in the literature have achieved a classification accuracy of around 70% - 90%. The classification accuracy attained in this paper is coherent with earlier reported algorithms as all the models are within this range. This may be a sign that consistently higher accuracy than this might be hard to get on a small dataset, even with specially adapted methods like

the ones used in the previous studies. One thing to note is that the SVM implemented in this paper has a 9.48 percentage point increase over the one implemented by [5]. This is most likely a consequence of [5] not focusing on optimizing the SVM but instead testing their new proposed measure of classification reliability.

The ANNs used in this paper reached a best mean accuracy of 82.16% which is worse than the previous study done by [6] which achieved an average accuracy of 89.01%. One of the main implementation differences is the use of ensemble methods which can potentially lower both bias and variance in the models and thus achieve higher overall accuracy. Their study used 14 nodes in the hidden layer which lowers the overall complexity of the model and is consistent with this paper's best ANN results. Overall slightly lower classification accuracy was achieved by the models in this paper which was expected as none of the models were optimized with a more advanced ensemble learning method which often results in higher accuracy and more reliable results.

5.2. Further Research. One of the biggest limiting factors in this paper, which has been mentioned before in previous sections, is the small size of the used dataset. With only around 300 samples it is hard to draw any general conclusions about the results as the models only get a small sample size to train on. Another factor, also related to the dataset, is the features the dataset contains. These 14 features are most likely not the only contributing factors to CVD and some of them may have very little or no contribution at all. With that said further research which uses a dataset containing a larger sample size and considers more features, such as whether the patient smokes or family history of CVD could achieve a more generalized result. All of the different image-based diagnostic methods could also benefit from machine learning which should be explored further.

Beyond a dataset with a larger sample size and more features, further research could also consider performing feature selection on the dataset that is used. With complex problems like diagnosing heart disease in patients taking into account many different features and then performing feature selection to only use, relevant features could yield better results.

6. Conclusion. The implementation of various machine learning models showed that there is no clear winner in the comparison of models. From the implementation, it is noticed that the support vector machines have achieved the highest average hit rate, while artificial neural networks achieved a similar highest hit rate. The support vector machine is best with regard to mean accuracy, highest accuracy, and lowest accuracy. Artificial neural networks along with the logistic regression gave the most reliable result with the lowest standard deviation. Although the decision tree achieved lower accuracy it can be visualized and interpreted easily by humans. These features led us to make the conclusion that the decision tree is the most practical model and also it is useful to doctors in their current methods of diagnosis. The scope of this work is primarily limited by the size of the dataset which contained a few patients, few features, and not enough diverse data. This must be taken into account while reviewing the relatively high mean accuracy of the models.

REFERENCES

[1] World Health Organization. Prevention of Cardiovascular Disease: Guidelines for Assessment and Management of Cardiovascular Risk. World Health Organization, 2022. https://www.cdc.gov/heartdisease/facts.htm
[2] Daniel Duprez. "Early detection of cardiovascular disease-the future of cardiology". In: EJ ESC Counc Cardiol Pract 4.19 (2006), p. 1.
[3] A. Ishaq, S. Sadiq, M. Umer et al., "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," IEEE Access, vol. 9, pp. 39707–39716, 2021.
[4] Bunyamin Ozaydin et al." Appropriate use of machine learning in healthcare Intelligence-Based Medicine", 2021.
[5] A. Jayakrishnan, R. Visakh, and K. T. Ratheesh, "Computational approach for heart disease prediction using machine learning," in 2021 International Conference on Communication, Control and Information Sciences (ICCISc), pp. 1–5, Idukki, India, 2021.
[6] S. Kathare and S. Gaikwad, "Practicability of heart attack prediction using machine learning," International Journal of Research Publication and Reviews, vol. 2, no. 7, pp. 1473– 1477, 2021.
[7] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system," World Journal of Engineering and Technology, vol. 6, no. 4, pp. 854–873, 2018.
[8] TensorFlow. https://www.tensorflow.org/
[9] C. Lal Chowdhary, and R. Srivatsan, "Non-invasive detection of Parkinson's disease using deep learning," IJIGSP, vol. 14, no. 2, pp. 38–46, Apr. 2022.

[10] S. Padinjappurathu Gopalan, C. L. Chowdhary, C. Iwendi, M. A. Farid, and L. K. Ramasamy, "An efficient and privacy-preserving scheme for disease prediction in modern healthcare systems," Sensors, vol. 22, no. 15, pp. 5574, Jul. 2022.
[11] S. Bhattacharya et al., "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey," Sustainable Cities and Society, vol. 65, pp. 102589, Feb. 2021.
[12] B. Kumar Swain, M. Zubair Khan, C. Lal Chowdhary, and A. Alsaeedi, "SRC: Superior robustness of COVID-19 detection from noisy cough data using GFCC," Computer Systems Science and Engineering, vol. 46, no. 2, pp. 2337–2349, 2023.
[13] B. Singh, S. Bhattacharya, C. L. Chowdhary, and D. S. Jat, "A review on internet of things and its applications in healthcare," vol. 10, no. 1, 2017.
[14] K. V. Dudekula et al., "Convolutional neural network-based personalized program recommendation system for smart television users," Sustainability, vol. 15, no. 3, pp. 2206, Jan. 2023.
[15] C. L. Chowdhary, N. Khare, H. Patel, S. Koppu, R. Kaluri, and D. S. Rajput, "Past, present and future of gene feature selection for breast cancer classification – a survey," IJESMS, vol. 13, no. 2, p. 140, 2022.
[16] Scikit-learn. https://scikit-learn.org/stable/
[17] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," IOP Conference Series: Materials Science and Engineering, Materials Science and Engineering, vol. 1022, no. 1, 2021.
[18] Junji Shiraishi et al. "Computer-aided diagnosis and artificial intelligence in clinical imaging". In: Seminars in nuclear medicine. Vol. 41, no. 6, pp. 449–462, 2011.
[19] M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra, "A review on heart disease prediction using machine learning and data analytics approach," International Journal of Computers and Applications, vol. 181, no. 18, pp. 20–25, 2018.
[20] V. Kumar, G. S. Lalotra, P. Sasikala, et al., "Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques," Healthcare, vol. 10, no. 7, 2022.
[21] T. K. Das, C. L. Chowdhary, and X. Z. Gao, "Chest X-Ray investigation: A convolutional neural network approach," JBBBE, vol. 45, pp. 57–70, May 2020.
[22] A. Jamthikar et al, Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. Cardiovasc Diagn Therapy, vol. 10, no. 4, pp. 919–938, 2020.
[23] DecisionTreeClassifier. https://scikit- learn.org/stable/modules/generated /sklearn. tree.DecisionTreeClassifier.html
[24] B. Prasanth et al., "Maximizing regenerative braking energy harnessing in electric vehicles using machine learning techniques," Electronics, vol. 12, no. 5, pp. 1119, Feb. 2023.
[25] P. P. Kasaraneni, Y. Venkata Pavan Kumar, G. L. K. Moganti, and R. Kannan, "Machine learning-based ensemble classifiers for anomaly handling in smart home energy consumption data," Sensors, vol. 22, no. 23, pp. 9323, Nov. 2022.
[26] S. N. V. B. Rao et al., "Day-ahead load demand forecasting in urban community cluster microgrids using machine learning methods," Energies, vol. 15, no. 17, pp. 6124, Aug. 2022.
[27] Roger J Lewis. "An introduction to classification and regression tree (CART) analysis". In: Annual meeting of the society for academic emergency medicine in San Francisco, California. Vol. 14, 2000.
[28] Y. F. Khan, B. Kaushik, C. L. Chowdhary, and G. Srivastava, "Ensemble model for diagnostic classification of Alzheimer's disease based on brain anatomical magnetic resonance imaging," Diagnostics, vol. 12, no. 12, pp. 3193, Dec. 2022.
[29] Yoshua Bengio Xavier Glorot Antoine Bordes. "Rectifier and soft plus activation functions. The second one is a smooth version of the first." In: Deep Sparse Rectifier Neural Networks. 2011.
[30] Jesmin Nahar et al. "Association rule mining to detect factors which contribute to heart disease in males and females". In: Expert Systems with Applications, vol. 40, no. 4, pp. 1086–1093, 2013.
[31] M. W. Segar, M. Vaduganathan, K. V. Patel, et al., "Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score," Diabetes Care, vol. 42, no. 12, pp. 2298–2306, 2019.
[32] B. K. Turkmenoglu and O. Yildiz, "Predicting the survival of heart failure patients in unbalanced data sets," in 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, Istanbul, Turkey, 2021.
[33] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Medical Informatics and Decision Making, vol. 20, no. 1, 2020.
[34] P. A. Moreno-Sanchez, "Improvement of a prediction model for heart failure survival through explainable artificial intelligence," 2021.
[35] I. A. Marbaniang, N. A. Choudhury, and S. Moulik, "Cardiovascular disease (CVD) prediction using machine learning algorithms," in 2020 IEEE 17th India Council International Conference (INDICON), pp. 491–495, New Delhi, India, 2020.
[36] I. U. Haq, I. Haq, and B. Xu, "Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging," Cardiovascular Diagnosis and Therapy, vol. 11, no. 3, pp. 911–923, 2021.

Edited by: Chiranji Lal Chowdhary
Special issue on: Scalable Machine Learning for Health Care: Innovations and Applications
Received: Jun 2, 2023
Accepted: Sep 10, 2023