Comparison between Machine Learning Algorithms for Cardiovascular Disease Prediction

By

Dr. Hanan Khodhary Mahdi Mahmoud

Lecturer of Statistics

Nile Institute for Commercial Sciences and Computer Technology, Mansoura

lecture0100@gmil.com

Scientific Journal for Financial and Commercial Studies and Research (SJFCSR)

Faculty of Commerce – Damietta University

Vol.4, No.1, Part 1., January 2023

APA Citation:

Mahmoud, H. K. M. (2023). Comparison between Machine Learning Algorithms for Cardiovascular Disease Prediction, ***Scientific Journal for Financial and Commercial Studies and Research***, Faculty of Commerce, Damietta University, 4(1)1, 909-926.

Website: https://cfdj.journalsekb.eg

Comparison between Machine Learning Algorithms for Cardiovascular Disease Prediction

Dr. Hanan Khodhary Mahdi Mahmoud

ABSTRACT

Healthcare is the cornerstone of a cohesive society and there are many diseases threaten human life. Cardiovascular diseases are one of them. Cardiovascular diseases (CVDs) are a group of heart and blood vessel problems. They include coronary heart diseases, which affect the blood vessels that supply the heart muscle; cerebrovascular illnesses, which affect the blood vessels that supply the brain; and peripheral arterial diseases, which affect the blood vessels that supply the arms and legs. In this study machine learning algorithms for adaptive perdition of cardiovascular diseases are proposed.

This study aims to significantly reduce the potential failure of machine learning algorithms to predict cardiovascular diseases. This is performed by comparing seven machine learning models: Random Forest, Decision Tree, Support Vector Machine (SVM), Adaptive Boosting (Adaboost), Nave Bayes, K-Nearest Neighbors (KNN), and Logistic Regression (LR). In the proposed research, four Kaggles were selected, and the dataset relied on ten years of historical records with 13 technical features. Furthermore, seven models of machine learning (Decision Tree, Random Forest, Adaboost, SVM, Naive Bayes, KNN, and Logistic Regression) were utilized as predictors. The input values of the methods are also used to produce three different measures for evaluations.

Keywords: Cardiovascular prediction; machine learning Algorithms.

1) INTRODUCTION

Healthcare is essential for the continuance of human life on the planet. But some serious diseases need a lot of awareness to be identified early. Cardiovascular disease is one of these serious diseases.

Unhealthy food, physical inactivity, cigarette use, and hazardous alcohol consumption are the major behavioral risk factors for heart disease and stroke. Individuals may have symptoms such as high blood pressure, high blood glucose, high blood lipids, and overweight or obesity as a result of behavioral risk factors. These "intermediate-risk factors" are detectable in primary care settings and suggest a higher risk of heart attack, stroke, heart failure, and other consequences. Cardiovascular disease (CVD) refers to a group of illnesses affecting the heart and blood vessels [1]. Coronary artery diseases (CAD), such as angina and myocardial infarction, are examples of CVD (commonly known as a heart attack) . Stroke, cardiac failure, hypertension, rheumatic heart disease, cardiomyopathy, irregular heart rhythms, congenital heart disease, valvular heart disease, carditis, aortic aneurysms, coronary artery disease, thromboembolic disease, and venous thrombosis are some of the other CVDs [2].

Although there are a large number of machine algorithms used to predict heart problems, some of them give higher accuracy or poor accuracy. Therefore, studying algorithms has become crucial nowadays to give high accuracy in the early stages [3]. With the confirmation of reports issued by the United Nations, the incidence and mortality of cardiovascular diseases in developing countries rises year by year and becomes one of the most frequent causes of death [4].

Depending on the disorder, the underlying mechanisms differ. Atherosclerosis is present in coronary artery disease, stroke, and peripheral artery disease. Coronary atherosclerotic plaques1 are divided into fatty streaks or intimal xanthomas, intimal thickening, pathological intimal thickening (PIT), fibroatheromas, and fibrocalcific plaques based on descriptive pathology. [5]. High blood pressure, smoking, diabetes mellitus, lack of exercise, obesity, high blood cholesterol, poor nutrition, and excessive alcohol intake, may all contribute to this. High blood pressure is thought to be responsible for about 13% of CVD deaths, with cigarettes accounting for 9%, diabetes for 6%, lack of exercise for 6%, and obesity for 5%. Untreated strep throat can lead to rheumatic heart disease [6].

Up to 90% of cardiovascular disease is thought to be preventable. CVD can be avoided by increasing risk factors such as healthy eating, exercise, avoid smoking, and limiting alcohol consumption. It is also useful in the treatment of risk factors such as hypertension, blood lipids, and diabetes. Antibiotics can reduce the risk of rheumatic heart disease in people who have strep throat. The importance of taking aspirin in otherwise stable people is debatable [7].

In contrast, cardiovascular diseases are the primary causes of death in Africa. After a combination of communicable, maternal, neonatal, and nutritional disorders, 2.6 million people died, accounting for nearly 35% of all mortality (CMNNDs) [8]. In 2015, 17.9 million people died from cardiovascular disease (CVD), up from 12.3 million (25.8%) in 1990. CVD deaths are more frequent and it rises even more in the developing world, while death rates in the developed world have been decreasing since the 1970s. Cardiovascular disease and stroke account for 80% of CVD deaths in men and 75% of CVD deaths in women. Total CVD cases nearly doubled from 271 million in 1990 to 523 million in 2019, and CVD deaths increased steadily from 12.1 million in 1990 to 18.6 million in 2019. The majority of cardiovascular disease affects people in their fifties and sixties. In the United States, 11% of people between the ages of 20 and 40 have CVD, while 37% of people between the ages of 40 and 60, 71% of people between the ages of 60 and 80, and 85% of people over the age of 80 have CVD. In the developed world, the average age of death from coronary artery disease is around 80, while in the developing world, it is around 68 [9].

Machine learning is a field of artificial intelligence (AI) and computer science that concentrates on using data and algorithms to mimic the way people learn and improve accuracy over time. Machine learning is an essential part of the rapidly expanding area of data science [10].
This study aims to decrease the failure rate of machine learning algorithms to predict cardiovascular diseases through machine learning algorithms.

To achieve this aim, a comparison is conducted between seven machine learning models: Decision Tree, Random Forest, Adaptive Boosting (Adaboost), and extreme, Support Vector Classifier (SVC), Nave Bayes, K-Nearest Neighbors (KNN), and Logistic Regression (LR). In the following sections, each of the seven machine learning models is discussed separately.

2) PREDICTION MODELS

Predictive algorithms use one of two things: machine learning or deep learning and both of them are subsets of artificial intelligence (AI). Machine learning (ML) involves structured data, such as spreadsheets or machine data. Deep learning (DL) deals with unstructured data such as video, audio, text, social media posts, and images—essentially the things that humans can communicate with such as things that are not numbers or metric reads. Predictive analytics, in brief, saves time, effort, and money when predicting business outcomes. Environmental considerations, competition information, regulatory changes, and market situations may all be integrated into the mathematical formula to produce more comprehensive perspectives at a lower cost [11].

There are some prediction forms that may help organizations such as demand forecasting, personnel planning, churn analysis, external variables, competitive analysis, fleet and IT hardware maintenance, and financial hazards are examples of particular. In this study, seven machine learning methods (Decision Tree, Random Forest, Adaboost, SVC, Naive Bayes, KNN, and Logistic Regression) are used.

A. Decision Tree:

Decision tree is a popular supervised learning technique used for regression and classification problems. This technique is used to predict the goal by applying simple decision-making rules based on the dataset and related features. Being easy to understand or being able to solve problems with different outputs are two benefits of using this model. Construction of over-complex trees, on the contrary, that trigger over-Ting is a common downside. A schematic representation of the decision tree is shown in Figure ١ adapted from [12].

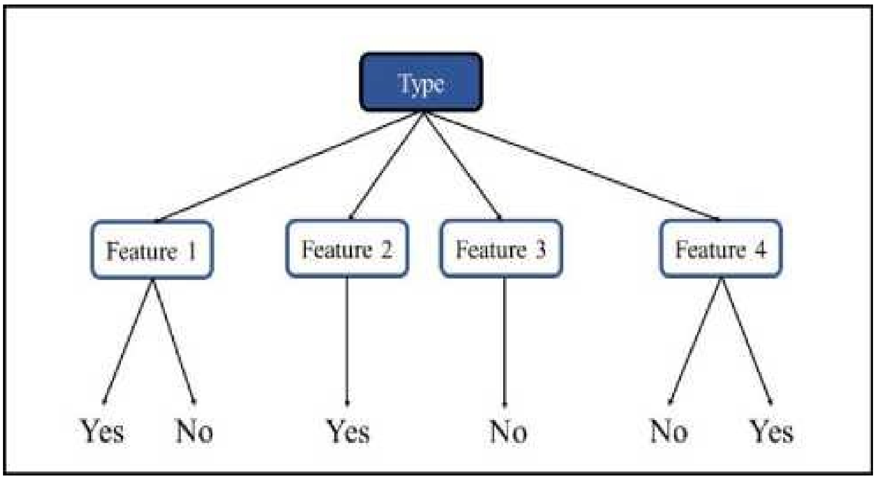

FIGURE 1. Schematic illustration of Decision tree

B. Random Forest:

A large number of decision-making trees are making a random forest model. Essentially, the model averages the predicted outcome of the trees called the forest. Also, the algorithm contains three random concepts, selecting training data randomly when creating trees, randomly selecting certain variable subsets when dividing nodes, and deeming only a subset of all variables to divide each node in each simple decision tree. Every basic tree learns from a random sample of the data set during a random forest training process. A schematic illustration of the random forest, adapted from [12], is indicated in Figure 2

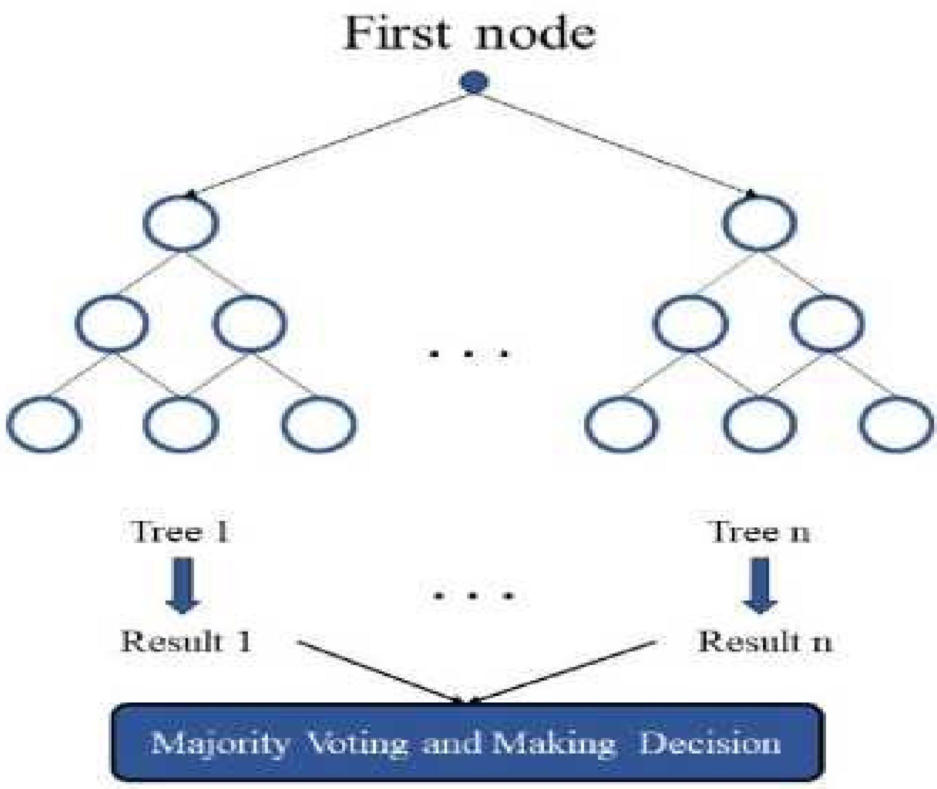

FIGURE 2. Schematic illustration of Random Forest

C. Adaboost:

The process of transforming some poor learners into strong ones is called the Boosting Form. Adaboost is a special form of boosting that uses an ensemble model to advance the predictions of each learning technique. The aim of boosting is to train poor learners to change their previous predictions sequentially. This model is a meta-predictor, which begins by setting the model to the basic dataset before applying additional copies to the same dataset. During the training process, sample weights are modified based on the current forecast error. Thus, the resulting model focuses on hard products.

D. Support Vector Machines (SVM's):

Support Vector Machines (SVMs) are a set of supervised learning methods that can be used for classification and regression problems. The classier variant is called SVM. The aim of the method is to create a decision boundary between two vector groups. The boundary must be far from every point in the data set, and the support vectors are the observation coordinates with a distance called margin Figure 3 shows the schematic illustration of SVM method adapted from [13].

$$F(X) = \text{sgn}\left(\sum_{i=1}^{n} \propto_i y_i . K(x,x_i) + b\right) \quad (1)$$

SVMs can perform linear or non-linear classifications effectively, but they must use a kernel trick to map inputs to high-dimensional spaces for non-linear classifications. SVMs transform non-separable groups into separable groups through kernel functions such as linear, non-linear, sigmoid, radial base function (RBF), and polynomial. The formula of the kernel functions is shown in Equations 2-4, where the radial base function is constant and the degree of the polynomial function is d. The sigmoid function has two adjustable parameters: the slope and the intercepted constant c.

$$\text{RBF: } K(x_i,x_j)= \exp(-\gamma \parallel x_i,-x_j \parallel^2) \quad (2)$$

$$\text{Polynomial: } K(x_i,x_j)=((x_i.x_j)+1)^d \quad (3)$$

$$\text{Sigmoid: } K(x_i,x_j)= \tanh(\propto_i^T y + c) \quad (4)$$

SVCs are also useful in high dimensional spaces. In cases where the number of dimensions is greater than the number of samples, but to prevent over-the-counter collection and kernel functions, the number of features should be much greater than the number of samples.

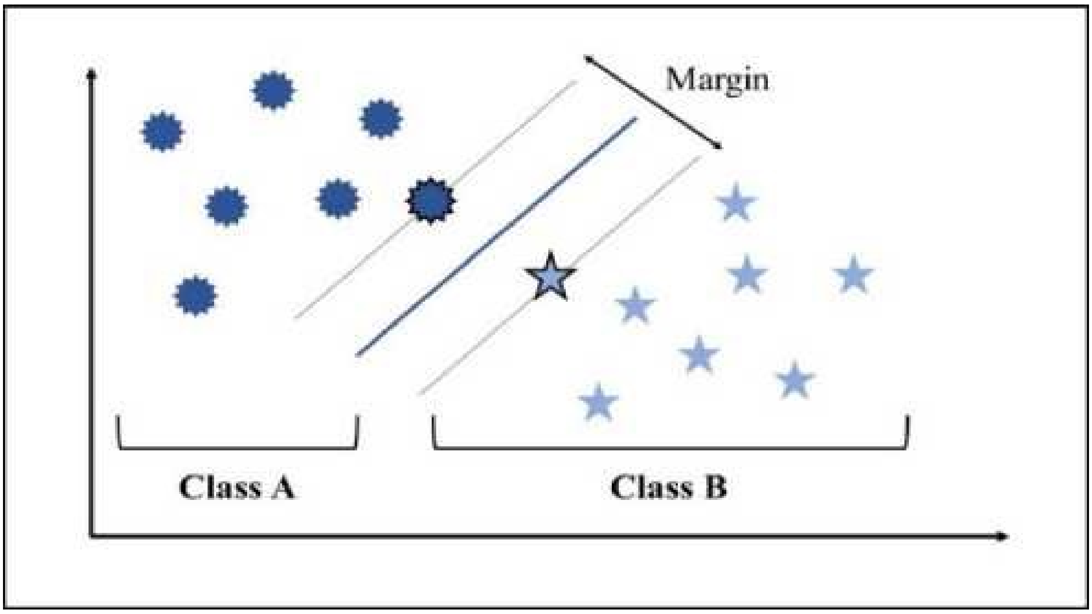

FIGURE 3. Schematic illustration of SVM.

E. Naive Bayes:

The Naive Bayes Classifier is a member of the probabilistic classifiers based on Bayes' theorem, with a clear independent inference between the characteristics given the value of the class variable. This is a compilation of supervised learning algorithms. The following relationship is stated in Equation 5 by Bayes' theorem, where y is a variable class, and x1 through xn are dependent vector features.

$$P(y/x_{1,\ldots\ldots}x_n)\ p(y)\prod_{i=1}^{n} p(x_i/y)/p(x_{1,\ldots\ldots}x_n) \qquad (5)$$

The Naive Bayes classifier can be very fast compared to more sophisticated algorithms. Separating the class distributions ensures that each can be measured independently as a one-dimensional distribution. Therefore, it can help ease the problems caused by the curse of dimensionality.

F. The k-nearest neighbors (KNN):

Two properties are proposed for KNN; lazy learning and non-parametric algorithms, as there is no assumption for the underlying distribution of data by KNN. The approach follows a variety of steps to find targets: dividing the data set into training and test data, selecting the value of K, deciding which distance function should be used, selecting the sample from the test data (as a new sample), and calculating the distance to its training samples, sorting distances obtained, taking the nearest k-data samples, and finally assigning the test class to the sample based on the majority vote of its k neighbors. Figure 4 displays the schematic illustration for the KNN process adapted from [13].

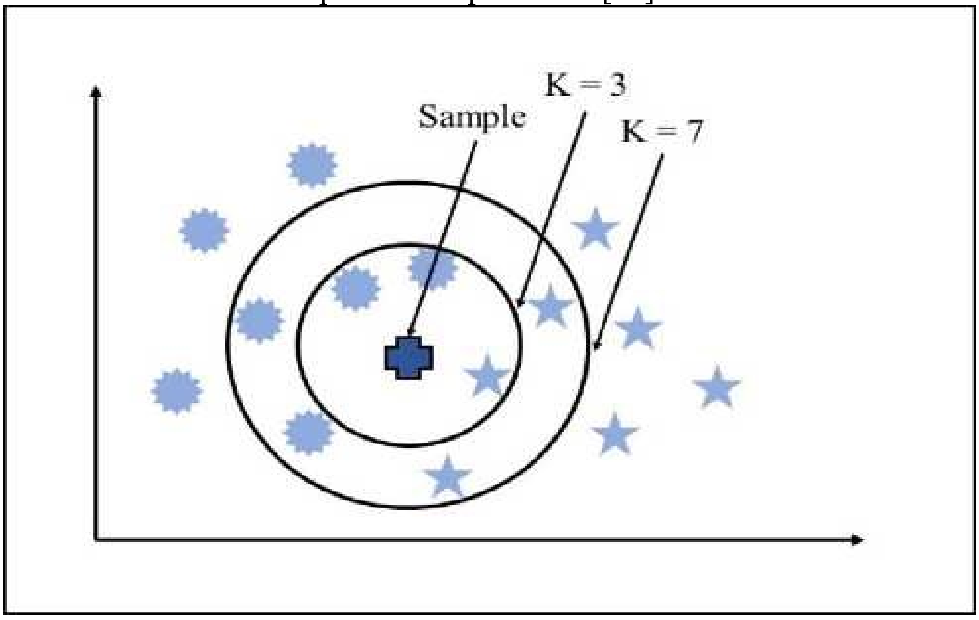

FIGURE 4. Schematic illustration of KNN.

G. Logistic Regression:

Logistic regression is used to assign observations to a different group of classes as a classifier. The algorithm transforms its output to return the probability value to the logistic value Sigmoid function, predicting the target by the principle of probability. Logistic regression is similar to the linear regression model, except that logistic regression uses a sigmoid function rather than a logistic one, with more complexity. The Logistic Regression Hypothesis is intended to restrict the cost function between 0 and 1.

3) MODELS PARAMETERS

All models (except Naïve Bayes) have one or several parameters known as hyper-parameters which should be adjusted to obtain optimal results presented in Tables 1

TABLE 1. Models parameters.

Model	Parameters	Value(s)
Decision Tree	Max Depth	10
Random Forest	Max Depth	10
	Number of Trees	50,100,150,..., 500
Adaboost	Max Depth	10
	Estimator	Decision Tree
	Number of Trees	50,100,150,..., 500
	Learning Rate	0.1
SVM	Kernels	Linear, Poly (degree = 3), RBF, Sigmoid
Naive Bayes	C	1.0
	Gamma	1 /(nunifX variancef) f: features
	Algorithm	Gaussian
KNN Classifier	Number of Neighbors	1,2, 3,..., 100
	Algorithm	K-dimensional Tree
Logistic Regression	Weights	Uniform
	Leaf Size	30
	Metric	Euclidean Distance (L2)
	Tolerance	10^4

4) CLASSIFICATION METRICS

F1-Score, Accuracy, and Receiver Operating Characteristics - Area Under the Curve (ROC-AUC) metrics are employed to evaluate the performance of our models. For Computing F1-score and Accuracy, Precision and Recall must be evaluated by True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values are indicated in Equations 6 and 7.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

By calculating the above equations, F1-Score and Accuracy are set out in Equations 8 and 9.

F1 - Score = [2 (Precision) (Recall)] / [(Precision) + (Recall)] (8)

$$\text{Accuracy} = \frac{TP + TN}{TP+FP+TN+FN} \quad (9)$$

Accuracy is a good metric among classification metrics, but it is not adequate for all classification problems. It is also important to look at certain other metrics to make sure the model is accurate. F1 – The score could be a better metric to use if the results need to be balanced between recall and precision, particularly when there is an unequal distribution of the class. ROC-AUC is another important metric for classical problems and is determined based on the region of the ROC-AUC curve based on the prediction score.

5) EXPERIMENTAL RESULTS

A.DATA Description

This real-world dataset was discovered on Kaggle and includes information on 7000 patients from (1) The Hungarian Cardiology Institute, (2) University Hospital Zurich, (3) University Hospital Basel, (4) V.A. Medical Center Long Beach, and (5) The Cleveland Clinic Foundation. In 1988, this dataset was shared with the rest of the scientific community, and it has been included since that time in hundreds of academic publications as a kind of laboratory for testing new machine learning concepts. There are 14 attributes in the data. Some traits, such as age and cholesterol levels, are constant. Some of them are categorical, such as sex and chest pain sort. The aim is to estimate the presence of heart disease in patients using the first 13 attributes. Since the training set is so small, the task is complicated. There are several technical indicators for forecasting heart disease, each of which has a unique ability to predict future trends in heart disease; however, based on previous studies [14] [15], we have chosen seven technical indicators for this paper. Table 1 (in the Appendix section) lists the technical metrics and formulas, while Table 2 (also in the Appendix section) lists the summary statistics.

B. RESULTS

For training machine learning models, we implement the following steps: randomly splitting the main dataset into train data and test data (30 percent of the dataset was allocated to the test part), fusing the models and testing them with validation data (and 'early stop') to avoid over fitting, and using test data for final evaluation. The entire coding process in this study is implemented by Orange 3 Program.

Based on comprehensive experimental work by considering methods, the following findings are obtained:

The results of this approach are shown in Table 2. For each model, the predictive performance of the three metrics is evaluated. The best tuning parameter for all models (with the exception of Naive Bayes and Logistic Regression) is also stated. In order to achieve a better representation of experimental works, Figure 5 displays the average F1-score based on the average running time of Heart Disease.

TABLE 2. Models with best parameters

Prediction Model of heart diseases							
Decision Tree				Random Forest			
Fl- score	Accuracy	AUC	ntrees	Fl- score	Accuracy	AUC	ntrees
0.939	0.939	0.991	5	0.742	0.743	0.814	13
Adaboost							
Fl- score	Accuracy	AUC	ntrees				
0.977	0.987	0.998	50				
SVC				Naive Bayes			
Fl- score	Accuracy	AUC	Kernel	Fl- score	Accuracy	AUC	
0.334	0.499	0.500	Poly	0.719	0.719	0.782	
KNN				Logistic Regression			
Fl- score	Accuracy	AUC	Neighbors	Fl- score	Accuracy	AUC	
0.730	0.730	0.802	5	0.695	0.696	0.756	

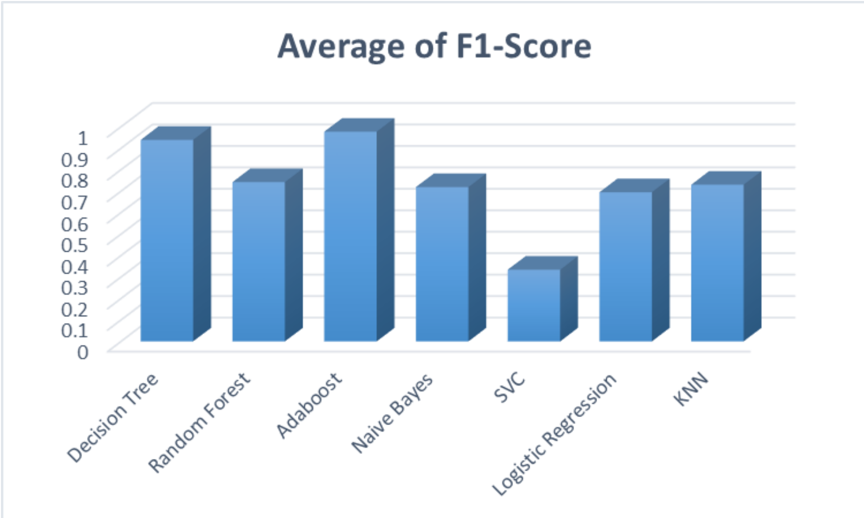

FIGURE 5. Average of F1-Score based on average logarithmic running per sample

6) CONCLUSIONS

Cardiovascular diseases (CVDs) refer to a group of heart and blood vessel issues. Coronary heart disease affects the blood vessels supplying the heart muscle; cerebrovascular disease affects the blood vessels supplying the brain, and peripheral arterial disease affects the blood vessels supplying the arms and legs. Adaptive perdition of cardiovascular disease using machine learning methods. The aim of this research is to compares and contrasts seven machine learning models: Decision Tree, Random Forest, Adaptive Boosting (Adaboost), Support Vector Classifier (SVC), Nave Bayes, K-Nearest Neighbors (KNN), and Logistic Regression (LR). The objective of this paper is to use machine learning algorithms to forecast cardiovascular illnesses. Data from Kaggles were chosen, and the dataset was based on 10 years of historical records with 13 technical characteristics. Future work includes different ensemble methods of these algorithms which can advance to better performance with more parameter settings for these algorithms.

Appendix

Data description

There are 3 types of input features:

- *Objective*: factual information;
- *Examination*: results of medical examination;
- *Subjective*: information given by the patient

TABLE 2 (in the Appendix section) displays the technical metrics and their formulas

1	Age | Objective Feature | age | int (days)
2	Height | Objective Feature | height | int (cm) |
3	Weight | Objective Feature | weight | float (kg) |
4	Gender | Objective Feature | gender | categorical code |
5	Systolic blood pressure | Examination Feature | ap_hi | int |
6	Diastolic blood pressure | Examination Feature | ap_lo | int |
7	Cholesterol | Examination Feature | cholesterol | 1: normal, 2: above normal, 3: well above normal |
8	Glucose | Examination Feature | gluc | 1: normal, 2: above normal, 3: well above normal |
9	Smoking | Subjective Feature | smoke | binary |
10	Alcohol intake | Subjective Feature | alco | binary |
11	Physical activity | Subjective Feature | active | binary |
12	Presence or absence of cardiovascular disease | Target Variable | cardio | binary |

All of the dataset values were collected at the moment of medical examination.

TABLE 3. Summary statistics of indicators.

Feature	Min	Max	Mean	Standard Deviation
age	10798.00	23713.00	19468.8658	2467.25167
sex	1.00	2.00	1.3496	.47684
height	55.00	250.00	164.3592	8.21013
weight	10.00	200.00	74.2057	14.39576
ap_hi	1.00	16020.00	128.8422	153.99880
ap_lo	.00	11000.00	96.6328	188.47282
cholesterol	1.00	3.00	1.3669	.68025
gluc	1.00	3.00	1.2265	.57227
smoke	.00	1.00	.0881	.28348
alco	.00	1.00	.0538	.22557
active	.00	1.00	.8037	.39718
cardio	.00	1.00	.4997	.50000

REFERENCES

[1] Mendis, S., Puska, P., Norrving, B. e., & Organization, W. H. (2011). *Global atlas on cardiovascular disease prevention and control*: World Health Organizationpp. 153.

[2] Centers for Disease Control Prevention. (2019). National center for chronic disease prevention and health promotion, division of population health. BRFSS prevalence & trends data.

[3] Arghandabi, H. (2020). A Comparative Study of Machine Learning Algorithms for the Prediction of Heart Disease. *International Journal for Research in Applied Science and Engineering Technology, 8*(12),pp. 677-683.

[4] Yuvalı, M., Yaman, B., & Tosun, Ö. (2022). Classification Comparison of Machine Learning Algorithms Using Two Independent CAD Datasets. *Mathematics, 10*(3),pp. 311.

[5] Narula, N., Olin, J. W., & Narula, N. (2020). Pathologic Disparities Between Peripheral Artery Disease and Coronary Artery Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology, 40*(9),pp. 1982-1989.

[6] WHO. (2020). Rheumatic Heart Disease, *World Health Organization* Retrieved from https://www.who.int/news-room/fact-sheets/detail/rheumatic-heart-disease

[7] WHO. (2021). Cardiovascular diseases (CVDs), *World Health Organization*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[8] Yuyun, M. F., Sliwa, K., Kengne, A. P., Mocumbi, A. O., & Bukhman, G. (2020). Cardiovascular Diseases in Sub-Saharan Africa Compared to High-Income Countries: An Epidemiological Perspective. *Global heart, 15*(1),pp. 15-15.

[9] Reddy, K. S. (2016). Global Burden of Disease Study 2015 provides GPS for global health 2030. *The Lancet, 388*(10053),pp. 1448-1449.

[10] Rekha, G., Bhanu Sravanthi, D., Ramasubbareddy, S., & Govinda, K. (2019). Prediction of Stock Market Using Neural Network Strategies. *Journal of Computational and Theoretical Nanoscience, 16*(5),pp. 2333-2336.

[11] Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems, 164*,pp. 163-173.

[12] Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & S, S. (2020). Deep Learning for Stock Market Prediction. *Entropy (Basel, Switzerland), 22*(8),pp. 840.

[13] Nabipour, M., Nayyeri, P., Jabani, H., S, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access, 8*,pp. 150199-150212.

[14] Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications, 38*(5),pp. 5311-5319.

[15] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications, 42*(1),pp. 259-268.

مقارنة بين خوارزميات التعلم الآلي للتنبؤ بأمراض القلب والأوعية الدموية

د/ حنان خضاري مهدي محمود

مدرس الإحصاء التطبيقي بمعهد النيل العالي للعلوم التجارية وتكنولوجيا الحاسب

الملخص:

الرعاية الصحية هي حجر الزاوية في مجتمع متماسك وهناك العديد من الأمراض التي تهدد حياة الإنسان. ومن هذه الأمراض أمراض القلب والأوعية الدموية. وأمراض القلب والأوعية الدموية هي مجموعة من مشاكل القلب والأوعية الدموية وتشمل أمراض القلب التاجية التي تؤثر على الأوعية الدموية التي تغذي عضلة القلب. بالإضافة إلى أمراض الأوعية الدموية الدماغية، التي تصيب الأوعية الدموية التي تغذي الدماغ؛ وأخيراً أمراض الشرايين الطرفية التي تصيب الأوعية الدموية التي تغذي الذراعين والساقين. وفي هذه الدراسة تم اقتراح خوارزميات التعلم الآلي للتنبؤ الدقيق بأمراض القلب والأوعية الدموية.

تهدف هذه الدراسة إلى الحد بشكل كبير من الفشل المحتمل لخوارزميات التعلم الآلي للتنبؤ بأمراض القلب والأوعية الدموية. وتم ذلك من خلال مقارنة سبعة نماذج للتعلم الآلي Random Forest :و Decision Tree و Support Vector Machine (SVM)و Adaptive Boosting (Adaboost) و Nave Bayes و K-Nearest Neighbours (KNN) والانحدار اللوجستي .(LR) ، وتم الحصول علي البيانات من موقع Kaggles ، خلال عشر سنوات من السجلات التاريخية ل ١٣ متغير. وتم استخدام سبعة نماذج من التعلم الآلي (Decision Treeو Random Forest و Adaboost و SVM و Naive Bayes و KNNو Logistic Regression) كخوارزميات للتنبؤ. وتم استخدام ثلاثة معايير للتقييم والمفاضلة بين الخوارزميات.

الكلمات المفتاحية: التنبؤ بأمراض القلب والأوعية الدموية. خوارزميات التعلم الآلي.