A REVIEW ON MACHINE LEARNING TECHNIQUES IN CORONARY ARTERY DISEASE: CHALLENGES AND FUTURE DIRECTION

Abi Izang Igyem[1], Fatima Umar Zambuk [2], Badamasi Imam Yau [3], Kehinde Emmanuel[4]

[1]*Department of Mathematical Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria*
[2]*Department of Mathematical Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria*
[3]*Department of Mathematical Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria*
[4]*Computer Science Department, Federal College of Education, Pankshin, Plateau, Nigeria*

ABSTRACT

According to recent studies, Coronary Artery Disease is said to be one of the leading origins of death worldwide. Early identification of CAD can assist to reduce death rates. Accurate prediction of CAD is necessary and considered to be a difficult attempt to treat a patient effectively before a heart attack occurs. When it comes to prediction using traditional methodologies, the difficulty arises in the intricacy of the data and relationships. This research presents a review of different machine learning algorithms that have been for detecting CAD, the review also presented the challenges of using such predicting methods and hence outlined a better learning algorithm for an efficient detection of CAD as its future direction..

Keyword: - *Coronary artery disease, Machine learning and Heart disease*

1. INTRODUCTION

In the modern world, everyone is so busy with their daily activities and work that they barely have time to take care of themselves. Due to their hurried lives, people frequently face stress, anxiety, despair, and many other conditions. They are becoming ill and developing serious diseases as a result of these major issues. Although several illnesses, including cancer, heart disease, TB, and others, cause deaths each year, heart disease or coronary disease (CVD) is the one that causes the greatest number of them [1]. According to data provided by the World Health Organization (WHO) at the Global Atlas on Coronary Disease Prevention and Control in Switzerland, heart-related diseases account for over 31% of all fatalities worldwide. Coronary artery disease (CAD) is a condition of the heart where the main blood vessels that supply the heart (coronary arteries) have difficulty supplying the heart muscle with enough blood, oxygen, and nutrients. Coronary artery disease is typically brought on by inflammation and cholesterol buildup in the heart arteries [2]. Angina and myocardial infarction, sometimes known as heart attacks, are examples of coronary artery disorders (CAD). Coronary heart diseases (CHD) are conditions where a waxy material called plaque builds up inside the coronary arteries. Heart attacks continue to be a leading cause of death globally, and if they are not treated quickly, they can result in major health issues and even death [3]. The heart is an essential organ of the human body. Blood oxygen and other nutrients are provided to the body by the heart's continuous pumping of blood to every region of the body. If the heart is not functioning properly, other bodily organs may begin to malfunction. Care for the heart and other organs become difficult as a result. Additionally, the risk of heart-related illnesses rises among people as a result of our hectic lifestyles and poor eating habits. After examining the information gathered by the medical community from a variety of patients using various machine learning models, we can define the risk, the symptoms, or the forecast for heart-related disease [4]. Computers are taught to learn knowledge without human involvement in a process known as machine learning (ML), a type of

artificial intelligence. Computers receive data, which they subsequently use to make learning (Regarding medicine, in 2020). The computer is fed data, and using machine learning techniques, the computer learns from the data. The size of the data typically affects how accurate the outcome is. With the development of machine learning (ML) for disease detection and prediction, the health sector has benefited greatly. Using data from the University of California Center for Machine Learning and Intelligent Systems repository, which contains four datasets from four separate hospitals, CAD has been predicted and diagnosed using a variety of machine learning methods. The Neural Network Algorithm produced results with an accuracy of 93%.

The remaining part of this work includes section 2 where we state our motivation and advantages, section 3 is a general overview of machine learning algorithms while section 4 is deep learning, in section 5 we looked at the related work of this research and finally section 6 we discussed the challenges of using supervised machine learning and gave a future direction of research.

2. OUR MOTIVATION AND ADVANTAGES

Our major motivation to conduct this research is the high dead rate recorded daily because of Coronary Artery Disease, also medical personnel's detection and diagnosing of heart diseases are based on the prediction on the patient's medical history, this prediction technique does not give an efficient accuracy. Machine learning techniques provide a high detection accuracy. With this emergent technique, medical doctors will have a reliable and accurate heart detection technique to diagnosis heart disease.

3. A GENERAL OVERVIEW OF MACHINE LEARNING TECHNIQUES

3.1 Naïve Bayes' Classifier

A supervised algorithm is the Naive Bayes classifier. It is a straightforward classification method based on the Bayes theorem. It makes a strong (Naive) assumption of attribute independence. A mathematical idea used to calculate probability is the Bayes theorem. The predictors don't have any connections to one another or a correlation with one another. To increase the likelihood, each attribute independently contributes. It does not make use of Bayesian techniques but can operate with the Naive Bayes model. Naive Bayes classifiers are used in many difficult real-world settings [5]. The classification algorithm Naive Bayes is straightforward, simple to use, and effective when dealing with difficult, non-linear data. As it is based on an assumption and class conditional independence, there is a loss of accuracy.

3.2 Decision Tree

A classification system known as a decision tree can be used with both category and numerical data. Tree-like structures are made using decision trees. A decision tree is a straightforward and popular tool for managing medical datasets. The data in a tree-shaped graph are simple to build and analyze. Three nodes serve as the basis for the decision tree model's analysis.

i. Root node: The primary node, upon which all other nodes are founded.
ii. Interior node: controls different properties.
iii. Leaf node: Display the outcomes of each test.

The results obtained are easier to read and interpret [6]. This algorithm has higher accuracy in comparison to other algorithms as it analyzes the dataset in the tree-like graph. However, the data may be over classified and only one attribute is tested at a time for decision-making. According [6], the results are simpler to read and analyze. As it analyzes the dataset in the tree-like graph, this algorithm is more accurate than other algorithms. However, since only one attribute is checked at a time for decision-making, the data may be over classified.

3.3 K-Nearest Neighbor (K-NN)

K-nearest neighbors is a strategy for supervised categorization. It groups things according to where they are physically located. It is an instance-based method to learning. The distance between two attributes is measured using the Euclidean distance [6]. It employs a group of named points and uses them to indicate a different point. K-NN can be used to fill in the blanks once the data are sorted according to how similar they are. Several prediction approaches are applied to the data set after the missing values have been filled in. By combining these algorithms in different ways, accuracy can be improved. It is easy to use the K-NN algorithm without building a model or making

any assumptions. This approach is flexible and is used for search, regression, and classification. Despite being the simplest method, K-NN's accuracy is impacted by noise and irrelevant features.

3.4 Random Forest Algorithm

The random forest algorithm is a supervised classification algorithmic tool. A forest in this algorithm consists of several trees. Each tree in a random forest emits a class expectation, and the class with the highest votes determines the model's forecast. The random forest classifier is more accurate the more trees there are in it. The following are the three often used techniques: Forest RI (random input choice); Forest RC (random combination); Forest RI and RC combined. It can handle missing variables and is used for both classification and regression tasks, but it excels at classification jobs. Due to the requirement for large data sets and more trees, results are illogical and slow to produce predictions [1].

3.5 Support Vector Machine (SVM)

Support Vector Machines popular known with its acronym SVM, are supervised learning machines based on statistical learning theory that can be used for pattern recognition and regression. Arithmetical learning theory can identify rather precisely the factors that need to be taken into account to learn successfully certain simple types of algorithms, however, real-world applications usually need more complex models and algorithms (such as neural networks), that makes them much harder to analyze theoretically. SVMs can be seen as lying at the intersection of learning theory and practice. They construct models that are complex enough (containing a large class of neural networks for instance) and yet that are simple enough to be analysed mathematically. This is because an SVM can be seen as a linear algorithm in a high-dimensional space [7].

4. RELATED WORK

Multiple approaches to machine learning have been used to accurately predict or identify various forms of cardiac disease. K-means and Artificial Neural Networks were used in a hybrid technique to increase accuracy, identify, and extract the unknown information of heart illness in the prediction of heart disease by [8]. These detected cardiac problems with a 97% accuracy rate. In order to improve the accuracy of coronary prediction in 2021, [9] implemented an AI technique to find relevant traits in a hybrid random forest linear model approach to predict heart disease. The model's accuracy in predicting heart disease was 88.7%. (Aravind, A. et al., 2021) developed predictive models utilizing various machine learning algorithms (Generalized linear model, Decision tree, Random forest, Support vector machine, neural network, and k-nearest neighbor) to aid clinicians in the early detection of coronary artery disease, with neural networks achieving the highest accuracy of 93%. [11] used machine learning to detect heart disease using historical medical records in order to find correlations in the data which greatly increase the correctness of prediction rates. The classifier techniques they employed, Modified Naive Bayes and Random Forests, yielded a 92% accuracy.

The majority of the corresponding literatures that were reviewed used supervised learning to recognize Coronary Disease; for this reason, this research will employ an unsupervised learning technique to address the topic at hand.

Table I: Summary of related work Based on Proposed Method, Problem Solved and Future Direction.

S/N	AUTHOR	TITLE OF THE WORK	OBJECTIVE	METHOD	PROBLEM SOLVED	FUTURE STUDY
1	K.S. Archana et al. / 2022	Automated cardio ailment identification and prevention by hybrid machine learning models	Applying recent machine learning technology to identify heart disease from past medical data to uncover correlations in data that greatly improve the accuracy of prediction rates	Modified Naïve Bayes and random forests classifier algorithm	Produces 92% prediction accuracy compared to other existing methods	Other diseases such as blood pressure, diabetics and the pulse rate can also be considered related to heart failure

2	Savannah L. Bergquist et al / 2017	Classifying Lung cancer severity with ensemble machine learning in health care claims data	To develop a tool for classifying lung cancer severity (early vs. late stage cancer)	Super learner ensemble framework	Classification of lung cancer patients receiving chemotherapy with 93% sensitivity, 92% specificity and overall accuracy of 93%	Exploring multi-level classification and building classification algorithms for other cancer types
3	Koushik Chandra Howlader et al. / 2022	Machine learning models for classification and identification of significant attributes to detect type 2 diabetes	To identify T2D-associated features that can distinguish T2D sub-types for prognosis and treatment purposes	Generalized boosted regression modeling	The best and most frequently accurate outcome predictors of diabetes are glucose levels, body mass index, diabetes pedigree function and age GAMLOESS is the top ranked classifier and FS5 is the most significant feature subset for achieving the best classifications and analyzing the disease	The performance of the model can be inspected using multiple diabetes datasets and explored with high performing machine learning models for various crucial features
4	Mehrbakhsh Nilashi et al. / 2015	Accuracy Improvement for diabetes disease classification: A	To classify diabetes disease by developing an intelligence system using machine learning	Clustering technique	A good classification accuracy was obtained	More attention should be to datasets for disease

		case on a public medical dataset	techniques		using the combined method of clustering, PCA and NN	classification by using the incremental machine learning approaches
5	Chaymae Benfares et al. / 2020	A clinical support system for classification and prediction of depression using machine learning methods	To analyze patient data based on their represented symptoms, In order to help clinicians and mental health practitioners classify and refine the type of depression disorder characterized in patient's intelligently, in order to make a relevant decision	Supervised classification(random-forest algorithm)	A clinical system for the classification and prediction of disorders of the depression characterized in patients based on relevant data using ML algorithms	Other algorithms can be improved on for a more satisfactory results
6	Loic Yengo et al. / 2016	Impact of statistical models on the prediction of type 2 diabetes using non-targeted metabolomics profiling	To classify specific metabolites in sub-clinical phases preceding the onset of type 2 diabetes to enable efficient preventive and personalized interventions	Logistic regression for the probability of incidence and Cox regression for the age at diagnosis of diabetes	Few biomarkers with an efficient combination as risk scores can improve the identification of incident type 2 diabetes cases, especially in those poorly recognized by classical risk factors	The use of statistical methodologies such as mediation analyses and mendelian randomization could provide avenues for further improvement
7	Nongyao Nai-arun et al. / 2015	Comparison of classifiers for the risk of diabetes prediction	To predict the risk of diabetes for everyone without the need of blood test or going to a hospital	Bagging and boosting techniques combined with four classifying algorithms	The creation model of a diabetes risk prediction model and application	
8	Senthilkumar Mohan et al.	Effective heart disease prediction	Finding significant features by applying	Hybrid random forest, and novel method by	All the features	New feature-selection methods

	/ 2019	using hybrid machine learning techniques	machine learning techniques resulting in improving the accuracy in the prediction of coronary disease	using linear model (HRFLM)	selected and ML techniques used, prove effective in accurately predicting heart disease of patients	can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction
9	Jian Ping Li et al. / 2020	Heart disease identification method using machine learning classification in E-healthcare	Proposing an efficient and accurate system to diagnosis heart disease	Classification algorithms	Achievement of good accuracy as compared to previously proposed methods	Using the proposed system on the treatment and recovery of critical diseases such as heart, breast and diabetes
10	G. Parthiban and S.K. Srivatsa / 2012	Applying machine learning methods in diagnosing heart disease for diabetic patients	To predict the heart disease for diabetic patients using diabetic diagnosis attributes	Support vector machines	Shown that mining helps to retrieve useful correlation even from attributes which are not direct indicators of the class to be predicted	Using the SVM model for the classification of a diabetic dataset
11	J. Beschi Raja et al. / 2019	Diabetic prediction using gradient boosted classifier	Creating a model to predict diabetics	Gradient boosted classifier	An optimized prediction accuracy by reducing loss function	Figuring out the features impacting by hybrid of feature selection methods with classifiers
12	Jaymin Patel et al. / 2016	Heart disease prediction using machine learning and data mining technique	To extract hidden patterns by applying data mining techniques which are noteworthy to heart diseases and to predict the presence of heart disease in patients	Decisions tree classification using J48 algorithm	The creation of a predictive model for heart disease patients with an accuracy of 56.76% and the total time build	Making use of multivariate decision tree approach on smaller and larger amount of data

					model is 0.04 seconds	
13	Luiz Henrique A. Salazar et al / 2022	Application of Machine learning techniques to predict Patient's No-show in the health sector	To explore the main causes that contribute to a patient's no-show and develop a prediction model able to identify whether the patient will attend their scheduled appointment or not	Random Forest Classifier	Attributes that can influence a patient's attendance in a medical consultation	Large dataset that includes additional information on patients and medical appointments
14	Ahmad Hammoudeh et al. / 2018	Predicting hospital readmission among diabetics using deep learning	To predict hospital readmission among diabetic patients	Convolutional neural networks and data engineering	A combination of convolutional neural networks and data engineering were found to outperform other machine learning algorithms when employed and evaluated against real life data	The approach can be used to predict patients with other diseases readmission
15	Angeline Yasodhara et al. / 2020	Identifying modifiable predictors of long-term survival in liver transplant recipients with diabetes mellitus using machine learning	To examine factors affecting survival in patients with Diabetes Mellitus post Liver transplantation	Machine learning	Identification of mortality risk factors, specifically delineating factors in the highest risk diabetic recipients	Recommendations to improve the long-term survival of Liver transplantation recipients with diabetes mellitus
16	Shwet Ketu and Pramod Kumar Mishra / 2021	Empirical analysis of machine learning algorithms on imbalance electrocardiogram	To critically analyze and summarize the state-of-the-art research articles on heart disease detection over electrocardiogram(ECG)	Machine learning algorithms (i.e., Support Vector Machine, K-Nearest Neighbors, Random	Finding out the impact of class balancing on the performance	An extension of this critical analysis with the algorithmic and data perspective

		based arrhythmia dataset for heart disease detection	datasets using machine learning	Forest, Extra Tree, Bagging, Decision Tree, Linear Regression, and Adaptive Boosting)	of machine learning algorithms which will be very helpful in developing a machine learning-based direction toward robotic or smart machine based solutions for social well-being	
17	Wenqi Li et al. / 2022	Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks	To predict the grade of coronary heart disease through heart color Doppler echocardiography report, blood biochemical indicators and ten basic body information items about the patients	Combined reinforcement multitask progressive time-series	A prediction of the degree of coronary vascular occlusion	The model can be applied to predict other diseases as well as other multi-task learning areas
18	Arthur M. Lee et al. / 2022	Using machine learning to identify metabolomics signatures of pediatric chronic kidney disease etiology	To identify metabolomics signatures in pediatric CKD based on diagnosis	Logistic regression, support vector machine, random forest and extreme gradient boosting	The identification of metabolomics signatures associated with pediatric CKD cause through untargeted metabolomics quantification	The method should be used in the identification of other disease
19	U. Sivaji et al. / 2021	A hybrid random forest linear model approach to predict the	To identify significant characteristics by applying AI techniques to improve the coronary	Random forest linear model algorithm	Detecting the accuracy of heart disease by using	The process can also be used in detecting accuracy of other

		heart disease	expectation accuracy		RFLM with a highest accuracy of 88.7%	disease
20	Yun-Chung Liu et al. / 2022	Evaluation of the need for intensive care in children with pneumonia: machine learning approach	to develop machine learning (ML) algorithms to predict ICU care needs for pediatric Pneumonia patients within 24 hours of admission, evaluate their performance, and identify clinical indices for making decisions for pediatric pneumonia patients.	Random forest and extreme gradient boosting models	An algorithm that could accurately classify the risk of early ICU transfer within 24 hours of admission for children with pneumonia	Imaging data such as chest x-ray can be included in the indices and also blood gas values and procalcitonin measures can be included in the algorithm training
21	Aravind Akella et al. / 2021	Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution	To build predictive models using machine learning algorithms to assist clinicians in timely detection of coronary artery disease	Machine Learning Algorithms (Generalized linear model, Decision tree, Random forest, Support vector machine, neural network, k-nearest neighbor)	Predicting coronary artery disease with an accuracies greater than 80%	Multiple and larger datasets can be used for further testing /validation of the algorithms used.
22	Amita Malav et al./ 2019	Prediction of heart disease using K-means and Artificial Neural Network as Hybrid approach to improve accuracy	To determine and extract the unknown knowledge of heart disease using hybrid combination of K-means clustering algorithm and artificial neural network	Hybrid approach (K-Means and Artificial Neural Network Algorithms)	97% accuracy rate of heart disease detection	Larger dataset can be used for testing the proposed approach
23	Intisar Ahmed/2022	A study of heart disease diagnosis using machine learning and data mining	To explore how machine learning algorithms can be used in predicting heart disease by building an optimized model	Random forest classification algorithm	Accuracy of 89.4%	Larger dataset with different hyper parameters
24	Varun Sapra et al./2023	Integrated approach using deep neaural network and CBR	To diagnose coronary artery disease among patients based on their clinical data using a	Deep neural network	96.2% accuracy	Can be extended to other non-communicable diseases such as

	for detecting severity of coronary artery disease	deep neural network	cancer, diabetes, hepatitis, etc

Table I presents the summary of related work based on methods applied in detecting Coronary Artery Disease, the targeted problem solved and the future direction of the paper. Based on the table, a lot of research had been conducted using supervised learning categorically single machine learning algorithm for classification, only few literatures use hybrid machine learning algorithms.

5. CHALLENGES AND FUTURE RESEARCH DIRECTION

Even though there are advances in healthcare and medicine research, the rate of coronary heart disease has been progressively increasing over time, and researchers have been working to identify the causes that are associated with future risk of the disease. As a result, the medical field requires an automated intelligent system to accurately predict heart disease to help in the decision-making process. In recent years a lot of research had been conducted using machine learning in the detection of heart disease, supervised machine learning algorithms have the advantage of having a clear goal: predicting the label of interest. There is rising worry about the insufficiency of early detection and preventive strategies for coronary artery disease (CAD) in high-risk patients despite advances in medical knowledge [11] and other literatures had proposed a hybrid (random forest and naïve Bayes) model, but the two algorithms are known to have the following disadvantages:

Naïve Bayes Algorithm
- There must be no interaction between classes.
- Interrelationships between qualities have a negative impact on labeling performance.
- It makes the assumption that numerical attributes follow a normal distribution.

Random Forest Algorithm
- More difficult to compute and expensive.
- It is necessary to specify how many base classifiers there are overall.
- When estimating variable importance, it highlights characteristics or variables with a wide range of potential values.

A common problem is overfitting.

In this paper our review shows that even with the advancement of machine learning in heart diseases there are numerous limitations to these algorithms as mentioned above. Hence there is a need to try out other sophisticated machine learning algorithms to foster a fast solution to the trending problems.

6. CONCLUSION

Coronary Artery Disease is said to be one of the leading origins of death worldwide. So research in this research has become a necessity. In this paper, we conducted a comprehensive review on different methodologies that were used in combating the coronary artery disease. Machine learning algorithms are said to be the most sophisticated method but there is also need for a hybrid machine learning algorithm for the coronary artery disease detection.

Acknowledgement

We wish to appreciate the expert support of our supervisors Dr. F. U. Zambuk and Dr. B. I. Yau for their guides towards the success of this research.

REFERENCES

[1] Ramalingam, V. Dandapath, V, A, & Karthik Raja M. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology. 2018;7(2.8):684–687. https ://doi.org/10.14419 /ijet.v7i2.8.10557

[2] Mayo Clinic (2022). Coronary artery disease https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613

[3] Swathy, M. & Saruladha, K. (2021). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques, ICT Express (2021), doi: https://doi.org/10.1016/j.icte.2021.08.021

[4] Mohan, S. Thirumalai, C. & Srivastava, G.(2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. Access IEEE. 2019;7:81542–54.

[5] Fatima, M. & Pasha, M. (2017). Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. 2017;9:1–16. https ://doi.org/10.4236/jilsa .2017.91001. Foresee Medical (2020) Benefits of Machine Learning in healthcare https://www.foreseemed.com/blog/machine-learning-in-healthcare

[6] Weng, S.F. Reps, J. Kai, J. Garibaldi, J.M. & Qureshi, N. (2017). Can machine learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944

[7] Awad, M. & Khanna, R. (2015). Support Vector Machines for Classification. https://www.researchgate.net/publication/300723807

[8] Aditya, M. Prince, K. Himanshu, Arya, & Pankaj, K. (2015). EarlyHeart Disease Prediction Using Data Mining Techniques", CCSEIT, DMDB, ICBB, MoWiN, AIAP – 2014, pp. 53–59, DOI: 10.5121/csit.2014.4807

[9] Animesh, H. Subrata, K. Mandal, A. Gupta, Arkomita Mukherjee & Asmita Mukherjee (2017). Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159

[11] K. S. Archana, B. Sivakumar, Ramya Kuppusamy , Yuvaraja Teekaraman , & Arun Radhakrishnan (2021). Automated Cardioailment Identification and Prevention by Hybrid Machine Learning Models. *Computational and Mathematical Methods in Medicine* Volume 2022, Article ID 9797844, 8 pages https://doi.org/10.1155/2022/9797844