A REVIEW OF AI MODELS FOR PREDICTION AND DETECTING HEART DISEASES FOR IMPROVED WELLBEING

Akansha Singh,[1] Dr. Nupur Prakash[2]
Research Scholar[1],
University School of ICT, GGS Indraprastha UniversityDelhi, India
Professor[2],
University School of ICT, GGS Indraprastha UniversityDelhi, India

Vivekananda Journal of Research
October, 2021, Vol. 10, Special Issue, Pg No14-25.
ISSN 2319-8702(Print)
ISSN 2456-7574(Online)
Peer Reviewed Refereed Journal
© Vivekananda Institute of Professional Studies
http://www.vips.edu/vjr.php

ABSTRACT

Background- *Coronary artery disease (CAD) is one of the top causes of death worldwide. Over the decades a lot of non- invasive methods were suggested for the prediction and early detection of this disease. However, even after so much rigorous researches, the mortality rate due to CAD is at its peak.*

Aim- *This Scoping review has been carried out with an aim to study and analyze the Machine Learning (ML) models for the diagnosis of CAD and to understand the issues hampering the performance of ML models in order to provide more efficient AI solutions to the clinicians. This study would further benefit the people in the early detection of CAD disease and ensure their wellbeing.*

Methods- *For this study, Google Scholar and PubMed have been considered as the search engines. The term "CAD", "cardiovascular disease", "machine learning", "deep learning", and "data mining" were used to select numerous papers for the study that were published between 2017 and 2021. The major focus of this study was on the different facets for predicting and diagnosing CAD.*

Result- *While reviewing, five aspects have been explored: the methods used for study, different data sources, data handling,feature selection, facets of the CAD study. It has been found that most of the studies have preferred supervised ML algorithms. Only a few studies have considered deep learning, and unsupervised learning. The datasets used in most of thestudies were very small. They were not big enough to be used for training, validation, and testing. Many of the important data sources like Clinical Texts were left out during the study.*

Conclusion- *This study concluded that major issues such as data handling (missing values), poorly selected features, and heterogeneous data, greatly affects the performance of ML models used in the study of CAD. It has shed light on the importance of handling these issues. This study will benefit the researchers in developing ML models while taking all theseissues into consideration. It will also equip the medical professionals and practitioners in providing early diagnosis and timely treatment to the heart patients, leading towards better healthcare and the overall wellbeing of people.*
KEYWORDS : *CAD, machine learning, data collection, feature selection.*

1. INTRODUCTION

Coronary artery disease (CAD) is one of the deadly diseases causing millions of deaths every year. It is one ofthe most important cardiovascular diseases (CVD). CAD is a condition where either one or all the major arteriesgot blocked or narrowed due to the accumulation of plaque inside the major arteries. This narrowing of majorarteries is known as Atherosclerosis. This narrowing can happen because of the accumulation of cholesterol or fatty acids. Calcification is also one of the major causes of thinning of coronary arteries, especially in the casesof diabetic patients [1, 2, 3].

In a report of the World Health Organization (WHO, 2017), it was stated that 17.9 million people die eachyear because of CVD, which is 31% of all the deaths occurring worldwide. More than 75% of deaths caused by CVD occur in low and middle-income countries [4]. It has been estimated that up to 23.3 million people will die because of CVD by 2030. Pagliaro et al. [5] have investigated CAD as the primary risk factor in more than 50%of heart failure in North America and Europe. In [6], the researchers performed a study on CAD in Saudi Arabia. This study showed that the middle-aged population is the most vulnerable to this disease. It also showed that inthe middle-east, the people developing CAD are ten years younger than those in western countries. In 2019, The American Heart Association [7] had estimated that there are 18.2 million Americans of age more than 20years who have CAD. European society of cardiology [8] has also provided their status on CAD. They have found out that middle-income countries in Europe are significant sufferers of CAD. All other countries of the European continent have shown a decline in the number of deaths caused by CVD, except for middle-income countries. Different countries have different causes leading to this disease for example, in south east Asian countries, scarcity of healthcare facilities is the major cause, in USA, obesity and smoking is the major cause ofthis disease etc. All these reports mentioned here clearly indicate the severity of the situation worldwide. CAD is a global problem. The causes of CAD differ from country to country. In some places, major cause is obesityor smoking, high cholesterol and stress at one place and low-income status at some other place. All these factorsneed to be considered while diagnosing this disease. Today, there are a lot of publicly available datasets consisting of various important features causing CAD. With the help of these datasets, important causes of CADcan be determined. Datasets from one region can also help in diagnosing CAD in other areas also. All these datasources including datasets need to be carefully studied and processed. Hence, in order to provide the best CADsolutions to the society, we have considered various aspects of CAD. By providing more efficient AI solutionsnot only the physicians are benefitted but also the heart patients all around the world. Hence, it is important tostudy the gaps in the studies related to CAD so that these gaps can be minimized and better-quality of medicalservices can be provided to the patients. Many preventions have been taken like, reducing the risk factors, creating awareness about this disease, etc. Sometimes the symptoms can be seen in the later stages of CAD condition, but there are times when these symptoms do not even show up until the occurrence of a heart attack.Hence, it is important to detect this disease in the early stages. Although there is an invasive method for the detection of CAD known as Angiography, it is very expensive and can even cause death. To overcome this issuea lot of non-invasive methods were suggested which use machine learning (ML). Over the decades, extensive research has been done regarding this issue [9-12].

Machine learning is one of the branches of Artificial Intelligence. Machine learning is a type of learning, wherein, models are created using statistical or mathematical models that can reason and work like the human brain without explicit programming. Over the few decades, machine learning has been used not only in the Cardiology field but also in various other sectors of healthcare like automated image screening, cancer detection, precision medicine, genomic analysis, drug discoveries, design and prognosis, etc. It has shown its success byproviding efficient and cost-effective models to the clinicians and medical

practitioners in the early detection of various diseases including CAD. Early detection of CAD enables timely medication of heart patients, thereby, reducing the risk of heart attacks, leading towards a good quality of life. In this review, numerous papers on coronary artery disease and other cardiology related topics have been studied to examine the type of algorithms used in these papers and the various health parameters used for diagnosing CAD such that major benefits can be given to the society.

The paper is organized as follows. Section 2 discusses the previous state of the art models designed for the detection of CAD, Section 3 describes the methodology used in order to extract all the relevant papers. It also describes the objectives and the review questions on the basis of which gaps in previous studies were identified. Section 4 deeply explains all the mentioned review questions and identify the gaps in the previous studies on CAD. In the last section, i.e., Section 5, conclusion and areas of study in the future are discussed in detail.

2. LITERATURE SURVEY

Yahaya et al [13] have performed a comprehensive review on the prediction of various heart diseases using data mining and machine learning techniques. This study shows that different researchers have performed different machine learning techniques to get the best accuracies. However, this is just marginal success. As the medical data is increasing, there is a need for more complex models which can handle such large volume of data. Kilic A [14] has reviewed various artificial intelligence and machine learning techniques in the cardiovascular field. The study analysed that there are mainly two different ML methods that are being used in the healthcare sector. They are supervised ML and unsupervised ML. The study also reviewed various performance measures, and ML applications in various fields of healthcare like automated Image interpretation, predictive diagnosis, quality data, etc. These areas need to be searched more as the medical data is growing day by day. The AI solutions need to be more cost-effective and efficient.

Jiang et al [15] have surveyed to determine the current status of AI in healthcare. So that the future of AI can be determined in the field of healthcare. The study shows that in healthcare there are two kinds of data- structured data and unstructured data which are handled by different techniques. For example, structured data is handled by various supervised machine learning algorithms like naïve Bayes, neural network, etc. whereas unstructured data is handled by NLP (natural language processing). It also shows the journey of how the data is collected from different sources, filtered through various NLP and ML techniques, and finally provides a clinical decision support system. The comparison of various ML techniques used in healthcare over the years is also discussed in this paper. The study concluded with the description of various AI tools and the problems it faces in the real-life deployment of such tools.

Rong et al [16] have conducted a study to inspect the current trends of AI in the healthcare domain. Since the technology and data keeps on increasing, it is not easy to keep track of all the technical innovations or their implications in the medical field. Hence, this study has tried to keep track of the accomplishments of AI in various fields of healthcare like disease diagnostics, biomedical information processing, and research. This can help researchers to pave new paths in the different domains of healthcare. This study also conducted two case studies. One is “the prediction of epileptic seizure occurrences”, and the other is “the filling of a dysfunctional urinary bladder”. Dalal [17] performed a study to analyze the implications of ML in healthcare and how ML is influencing and helping the healthcare industry. The study showed that although there are a lot of ML applications in healthcare, all of them are still in their initial stages. With the increasing complexity of data and various issues regarding healthcare, there is a need for improving the models for handling such tasks. Yadav et al [18] have performed a

study on a UCI machine learning repository dataset to determine the best algorithms for predicting heart disease. The various algorithms studied in this paper are naive Bayes, logistic regression, KNN. Hybrid algorithms were also generated and evaluated. After performing experiments on a python platform, it was determined that neural network and fuzzy KNN has outperformed the other algorithms. The study has performed regularization on the models to achieve better accuracy.

Minou et al [19] have performed a study to predict the 10-year risk of future coronary heart disease in patients. For this study, they have taken a heart dataset from the UCI dataset library. This dataset consisted of4270 patients and 14 attributes. This is a binary classification problem. SMOTE filter was used to tackle downthe class imbalance problem. Training and testing sets were created using cross-validation. To evaluate the performance of various classifiers, various performance metrics were used like F1-score, AUC, accuracy, recall, etc. after the experiments, it was concluded that random classifiers and decision trees have given the best performance. Martin-Isla et al [20] have performed an analysis on how cardiac imaging is also an importantpart of diagnosing cardiovascular diseases. The study provides a great insight into the importance and implications of ML in various fields of healthcare. With the constantly growing data, new opportunities are also evolving, leading to the building of new models. These models can further be used in a different applicationof healthcare like automated imaging, atherosclerosis, coronary artery disease, etc.

Chen et al [21] have performed a study to evaluate the performance of the deep learning model in detecting obstructive stenosis in coronary arteries. It was evaluated using an analysis of computed tomography angiography in 124 patients. Receiver operating characteristics curves are used to compare the performance ofthe deep learning model and the reader model. The AUCs were calculated for three different scenarios. In the patient-based analysis, the resulted AUC value was 0.78, in segment-based analysis, its value was 0.84 and in vessel-based analysis, its value was 0.87. Finally, it was concluded that deep learning is an effective and less time-consuming method for determining obstructive stenosis in coronary arteries as compared to the reader model. Chowdary GJ [22] has conducted a study to predict whether a person has heart disease or not. The study suggested an ensemble approach which is a combination of Logistic regression, random forest, K-nearest neighbors, naïve Bayes, and artificial neural network with ReLU function. To evaluate the performance of this model, different metrics were used. The study resulted in an accuracy of 91%, precision of 91.6%, the sensitivity of 86%, and specificity of 91%. After comparing the results with different individual models, it wasdetermined that the ensemble modeling provides much better results.

Yan et al [23] have performed an analysis regarding the issues of AI in healthcare. The study shows that, although AI has been around for many years, it's application in healthcare is still not clearly understood. Thisstudy has tried to show the importance of using of AI in medicine and its application in various healthcare sectors like disease diagnosis, prognosis, automated imaging process, etc. Many tools have been developed to provide the best healthcare services, but no research has clearly defined the impact of AI technology on clinicians, medicos and their jobs. The paper explores a lot of unsolved problems that need to be tackled.

3. METHODOLOGY

In this study, a scoping review was conducted on various papers related to CAD with the following objectives:

- To determine the implementations of various ML methods on CAD diagnosis and treatment.
- To find various gaps in the studies conducted on CAD by answering some review questions.

This study elaborates the issues and problems related to various facets of CAD that were not covered in theprevious works. This study has been carried out to help medical practitioners and researchers in developing better ML models for predicting CAD while keeping these issues in mind. The work presented in this paper isfocused in determining the various issues related to the studies conducted on CAD that are becoming the hindrances in the good performance of ML models. There are some questions which are considered in order toidentify such issues. As the goal of healthcare sector is to provide best services to its patients, it has become imperative for researchers to provide efficient solutions to the clinicians.

In our scoping review, the objectives of the review were focused on the pertinent issues specifically relatedto the prediction and diagnosis of CAD, and were guided by the following research questions:

a) What are the kinds of models used in such studies?
b) What are the different sources of data?
c) Was the data handled correctly?
d) Whether any feature selection was done or not?
e) What facets of CAD were considered in the study?

A literature survey of electronic databases was conducted for published articles on CAD from 2017 to 2021. The database used were PubMed and Google Scholar. The search term used in both databases were “CAD”, “machine learning in CAD”, “AI in CAD”, and “deep learning”. Search strategy incorporated two concepts: i) CAD (i.e the condition), and ii) methods used for predicting CAD (i.e. ML, DL).

Inclusion criteria - main focus was to extract papers based on the methods used for predicting and diagnosing CAD using non-invasive methods. Some of the other criteria are mentioned below:

Outcome based: all the papers which were based on different facets on CAD were considered i.e whether the person has CAD or not or which artery is stenotic or what is the percentage of stenosis in each artery.

Dataset used: all the studies which included some kinds of datasets were considered in order to determine the scale of input fed to the models. Some of the studies were based on symptoms, some were base don ECG and Lab reports etc.

Language used- all the papers written in English were considered.

After the extraction of the relevant papers, several gaps were noticed in the previous studies on CAD. These gaps are discussed in the next section in detail.

4. EVALAUTION AND DISCUSSION

4.1 Models used in CAD studies

It can be seen in various studies that ML models have been implemented in the cardiology field. Different types of machine learning algorithms are Supervised learning, unsupervised learning and reinforcement learning, but, the most popular machine learning algorithm used in CAD diagnosis is Supervised machine learning. Since the datasets considered for CAD diagnosis have labels which make it a classification task. Different kinds of supervised learning models were used in the studies like Support vector machines (SVM) [24, 25],

Naïve Bayes, decision trees [26], Random forest [27, 28], Artificial neural networks (ANN) [29], etc. The idea behind these models is almost the same except for the function that is used to map the inputs to the correct output labels. Training data is fed to these models as input (x). The model learns the parameters to connect the input to the correct label class using a function f(x), which varies in different models. When given testing data, the model uses this function f(x) to determine the correct output (y).

There are only a few studies, where, unsupervised learning was considered [30, 31]. There is another type of learning known as Deep Learning. It is a subfield of machine learning which mimics human cognitive behavior to solve complex tasks. It is inspired by the neurons inside the human brain. It has become popular after the year 2006 because earlier it was not able to converge at an optimum solution. Now, the existing deep neural networks have better activation functions, better regularization methods to solve large and complex tasks.

In recent years, deep learning has shown great success in the cardiology field especially in the imaging segment [32-35]. As the medical data doubles every few months on average, it would be better to use deep learning models as they can deal with it very efficiently.

4.2 Different sources of data

Medical data is spread widely. There are many sources available for medical data. All these sources of data aredivided into two parts, the structured data and unstructured data. Structured data can be inferred as the data ina patient matrix form, having proper rows and columns like Electronic Medical Records (EMR), billing information, laboratory tests, etc. Unstructured data do not have a specific form. They need to be handled differently to use them. Examples of unstructured data are imaging data like Computer Aided Tomography (CAT) scans, clinical texts, X-rays, etc.

There are also datasets available from various open-source platforms like the UCI dataset repository. Some studies use their local datasets, but most of the studies have used datasets from open-source platforms. Differenttypes of datasets used in the literature are Cleveland dataset [35-40], Hungarian dataset [41, 42], Z-Alizadeh Sani dataset [43-47], etc. All of these datasets contain relevant information regarding the CAD. The problem with these datasets is that they are very small in size. To obtain good quality results, the training data should be large enough which can help the model to learn better. In medical applications, datasets hold a great importance. Dataset belonging to different countries contain different parameters regarding CAD. Their analysis does not only help locally but also globally. Hence, for providing better protection from CAD to the population of any country, it is necessary that these datasets should be processed properly as they will form thebasis of the creation of non-invasive models.

Sometimes, to increase the number of samples, a lot of extra information is added to the existing retrospective data. But we need to be careful while doing that because adding data is a whole lot different from adding information. ML models can find unnecessary correlations in the data which can further provide dangerous medical results.

Another problem with these datasets is that they do not contain complete information. Some of the unstructured data were left out like clinical texts. To have a good insight into a particular disease, it is importantto include all the medical information available and then refine it.

4.3 Data handling

The backbone of any AI solution is the data. The success of any ML model highly depends upon the quality ofthe data being used. In many of the previous researches, the data has been directly used

without being pre- processed. Before using all this data as an input to different models, it needs to be cleaned and pre-processed. No matter how great the model is, it won't produce the best results until you have good quality data.

As the data comes from various sources, it might contain many and different kinds of flaws. For example-

Missing values

The data may contain missing values, e.g. for a certain disease diagnosis, some patients have gone for a laboratory test and some may not. Those empty entries are known as missing values. There are various methods available for treating missing values like putting "$" or writing "Na" in missing entries.

Sparse data

Sparsity occurs in a dataset when for a particular column (feature) most of the row entries (patient entries) are missing. This can greatly affect the performance of the machine learning models. If there are very few entries for that particular feature, the column can be dropped or we can use any other method like "imputation".

Imbalance data

This is one of the most recurring problems in the datasets. An imbalance class problem occurs when one of the classes contains a majority of the instances and the other class contains only a few instances. For example, in the Z-Alizadeh Sani dataset, there are a total of 303 instances out of which 227 belong to the majority class (CAD people) and 76 belong to the minority class (Healthy people). This problem will result in biased results. To tackle down this problem, supervised or unsupervised filters can be used, e.g., "Class Balance" [46],

"*resample*" and "*spreadsubsample*" [48] filter.

Clinical text processing

Some of the datasets might leave out clinical text data, but they are as important as any EMRs or any laboratory tests. One reason might be that they are unstructured data. Hence, Natural Language Processing (NLP) is used to process the information in clinical texts. They are not written in plain grammar English and also contain various negation terms, e.g. a doctor may write that the patient does not have a certain symptom. While processing clinical notes, their negative terms should be removed as they can affect the results.

Sometimes it happens that good accuracy is achieved despite having bad quality data. This condition is knownas the "accuracy paradox". This has happened in many of the previous studies, where, some of the data-relatedproblems remained unsolved like the class imbalance problem despite achieving high accuracy. In such situations, the solution would not be considered as a good medical solution as the result would be biased and such results are riskier in medical applications.

Feature engineering

After having good quality data, some of the important features need to be selected. Most of the previous studieshave performed feature selection using the "*InfoGain*" method and the "*weigh-by-SVM*"

method. However, some of the studies have preferred the complete feature set. Feature selection is one of the most important steps in the prediction of a certain disease. The feature set might contain some redundant or correlated features. Suchkind of features need to be removed from the set. There should be neither too many nor too few features in thefeature set.

There are various feature selection methods available, such as, "*GainRatioAttributeEval*", "*weight-by- SVM*" and many more. These methods are based on ranking. They evaluate the importance of a feature concerning the class feature. Those features having the same value are considered to be redundant and need tobe removed from the feature set. Some other feature selection methods are wrapper class and swarm searches methods.

Facets of CAD

In coronary artery disease studies, one facet that is considered in most of the studies is to determine whether aperson has CAD or not. The datasets used for such studies have two classes: class1 (those people who have CAD), and class 2 (healthy people).

Another refined way of determining CAD in people is to know the status of the three major coronary arteries. These major arteries are Left Anterior Descending (LAD), Left Circumflex (LCA), and Right Coronary Artery (RCA). It is known that even if one of the arteries is stenotic, then the person would have CAD, otherwise not. A few studies based on this facet were also done in the literature. However, if the percentage of stenosis or blockage in the arteries is known, then it would be much easier to tell whether a person has CAD or not. If the percentage of stenosis is more than 50% then the person has CAD otherwise not.This is also another facet of diagnosing CAD. The researches on this facet have spiked up in the past years [21,25].

CONCLUSION

In this scoping review, various studies on Coronary Artery Disease have been studied and analyzed todetermine the various machine learning models implemented in predicting CAD. In order to strengthen thequality of the healthcare services provided to the society, it is necessary to identify these gaps (the issues related to CAD) so that they can be removed or minimized while designing a new ML model. Various issues relatedto such studies were identified and discussed in detail. This study has given insight into various gaps andimpediments related to applying various ML models on prediction, diagnostics and treatment related to CAD.The previous studies were based on one facet of CAD which determines whether a person has CAD or not, and such studies were implemented using a Supervised machine learning algorithm. As the medical data is growingday by day, it is important to use better models such as deep learning in order to get better insight of the disease.It is also found that most of the studies have used open-source platform datasets which were very small in size. Clinical text and other types of unstructured data were left out in some of the datasets. When the data usedfor the study is not properly processed, it results in poor accuracy. The aim of a study should not be just to provide a high accuracy model, but an efficient model which has some clinical utility.

This review work may benefit many researchers and practitioners in designing new models for predicting CAD while keeping the aforesaid issues in mind. This, in turn, will enable the medicos and para medicos in providing timely treatment to heart patients using appropriate AI based predictive techniques and early diagnostics. Eventually, it will help people live a cardio vascular disease-free life through lifestyle modifications and proper medication, leading towards their overall wellbeing and a healthier society.

In future works, large volume of data or large datasets should be used in order to obtain much better results. With the medical data keeps growing, it would be better to use deep neural networks as it would give better performance than traditional machine learning models.

REFERENCES

1. Mori, H., et al, 2018. Coronary artery calcification and its progression: what does it really mean?. *JACC: Cardiovascular Imaging*, *11*(1), pp.127-142.

2. Yahagi, K., et al, 2017. Pathology of human coronary and carotid artery atherosclerosis and vascular calcification in diabetes mellitus. *Arteriosclerosis, thrombosis, and vascular biology*, *37*(2), pp.191-204.

3. Ahmed, S. and Sobh, R., 2019. Predictive Value of Osteoprotegerin for Detecting Coronary Artery Calcification in Type 2 Diabetes Mellitus Patients in Correlation with Extent of Calcification Detected by Multidetector Computed Tomography. *Endocrine, Metabolic & Immune Disorders-Drug Targets (Formerly Current Drug Targets-Immune, Endocrine & Metabolic Disorders)*, *19*(6), pp.845-851.

4. Cardiovascular Disease (CVDs), [accessed on 15/2/2021] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

5. Pagliaro, B.R., et al, 2020. Myocardial ischemia and coronary disease in heart failure. *Heart failure reviews*, *25*(1), pp.53-65.

6. Albugami, S., et al, 2020. The Perception of Coronary Artery Disease and Cardiac Catheterization in Saudi Arabia:"What the Public Know". *Cureus*, *12*(1).

7. Virani, S.S., et al, 2020. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation*, *141*(9), pp.e139-e596.

8. Timmis, A., et al, 2020. European Society of Cardiology: cardiovascular disease statistics 2019. *European heartjournal*, *41*(1), pp.12-85.

9. Marimuthu, M., et al, 2018. A review on heart disease prediction using machine learning and data analyticsapproach. *International Journal of Computer Applications*, *181*(18), pp.20-25.

10. Martins, B., et al, 2021. Data Mining for Cardiovascular Disease Prediction. *Journal of Medical Systems*, *45*(1),pp.1-8.

11.. Buchan, K., et al, 2017. Automatic prediction of coronary artery disease from clinical narratives. *Journal ofbiomedical informatics*, *72*, pp.23-32.

12. Ayon, S.I., et al, 2020. Coronary artery heart disease prediction: a comparative study of computationalintelligence techniques. *IETE Journal of Research*, pp.1-20.

13. Yahaya, L., et al, 2020. A comprehensive review on heart disease prediction using data mining and machinelearning techniques. *American Journal of Artificial Intelligence*, *4*(1), pp.20-29.

14. Kilic, A., 2020. Artificial intelligence and machine learning in cardiovascular health care. *The Annals of thoracicsurgery*, *109*(5), pp.1323-1329.

15. Jiang, F., et al, 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascularneurology*, *2*(4).

16. Rong, G., et al, 2020. Artificial intelligence in healthcare: review and prediction case studies. *Engineering*, *6*(3),pp.291-301.

17. Dalal, K.R., 2020, July. Analysing the Implementation of Machine Learning in Healthcare. In *2020 InternationalConference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 133-137). IEEE.

18. Yadav, S.S., et al, 2020, March. Application of machine learning for the detection of heart disease. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 165-172). IEEE.

19. Minou, J., et al, 2020. Classification Techniques for Cardio-Vascular Diseases Using Supervised MachineLearning. *Medical Archives*, *74*(1), p.39.

20. Martin-Isla, C., et al, 2020. Image-based cardiac diagnosis with machine learning: a review. *Frontiers incardiovascular medicine*, *7*, p.1.

21. Chen, M., et al, 2020. Diagnostic performance of deep learning-based vascular extraction and stenosis detection technique for coronary artery disease. *The British journal of radiology*, *93*(1113), p.20191028.

22. Chowdary, G.J., 2020. EFFECTIVE PREDICTION OF CARDIOVASCULAR DISEASE USING CLUSTEROF MACHINE LEARNING ALGORITHMS. *Journal of Critical Reviews*, *7*(19), pp.1865-1875.

23. Yan, Y., et al, 2019. The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?. *Journal of geriatric cardiology: JGC*, *16*(8), p.585.

24. Terzi, M.B. and Arikan, O., 2019, October. Coronary Artery Disease Detection by using Support Vector Machines and Gaussian Mixture Model. In *2019 Medical Technologies Congress (TIPTEKNO)* (pp. 1-4). IEEE.

25. Alizadehsani, R., et al, 2013. A data mining approach for diagnosis of coronary artery disease. *Computer methodsand programs in biomedicine*, *111*(1), pp.52-61.

26. Abdar, M., et al, 2019, February. Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 26-30). IEEE.

27. Mohan, S., et al, 2019. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, *7*, pp.81542-81554.

28. Yang, L., et al, 2020. Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific reports*, *10*(1), pp.1-8.

29. Arabasadi, Z., et al, 2017. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer methods and programs in biomedicine*, *141*, pp.19-26.

30. Rokeakh, R., et al, 2020, December. Unsupervised learning reveals two phenotypes of left ventricle contraction in patients with heart failure. In *AIP Conference Proceedings* (Vol. 2313, No. 1, p. 070026). AIP Publishing LLC.

31. Ross, E., et al, 2018. Phenotype Discovery in Cardiovascular Patients Using Unsupervised Learning. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *38*(Suppl_1), pp.A245-A245.

32. Poongodi, T., et al, 2021. Deep Learning Techniques for Electronic Health Record (EHR) Analysis. In *Bio- inspired Neurocomputing* (pp. 73-103). Springer, Singapore.

33. Dai, L., et al, 2021. Deep learning-based classification of lower extremity arterial stenosis in computedtomography angiography. *European journal of radiology*, *136*, p.109528.

34. Nishi, T., et al, 2021. Deep learning-based intravascular ultrasound segmentation for the assessment of coronary artery disease. *Int J Cardiol.* 2021 Jun 15;333:55-59. doi: 10.1016/j.ijcard.2021.03.020. Epub 2021 Mar 16. PMID: 33741429.

35. Lin S., et al, 2020. Feasibility of using deep learning to detect coronary artery disease based on facial photo. *EurHeart J.* 2020 Dec 7;41(46):4400-4411. doi: 10.1093/eurheartj/ehaa640. PMID: 32818267.

36. Zhou, C., et al, 2021. Automated Deep Learning Analysis of Angiography Video Sequences for Coronary ArteryDisease. *arXiv preprint arXiv:2101.12505*.

37. Verma, L., et al, 2016. A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. *Journal of medical systems*, *40*(7), pp.1-7.

38. Feshki, M.G. and Shijani, O.S., 2016, April. Improving the heart disease diagnosis by evolutionary algorithm ofPSO and Feed Forward Neural Network. In *2016 Artificial Intelligence and Robotics (IRANOPEN)* (pp. 48-53).IEEE.

39. Latha, C.B.C. and Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensembleclassification techniques. *Informatics in Medicine Unlocked*, *16*, p.100203.

40. Ayon, S.I., et al, 2020. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. *IETE Journal of Research*, pp.1-20.

41. Caliskan, A. and Yuksel, M.E., 2017. Classification of coronary artery disease data sets by using a deep neural network. *The EuroBiotech Journal*, *1*(4), pp.271-277.

42. Setiawan, N.A., et al, 2020. Diagnosis of coronary artery disease using artificial intelligence based decision support system. *arXiv preprint arXiv:2007.02854*.

43. Alizadehsani, R., et al, 2012. Diagnosis of coronary artery disease using data mining based on lab data and echofeatures. *Journal of Medical and Bioengineering*, *1*(1).

44. Alizadehsani, R., et al, 2019. Machine learning-based coronary artery disease diagnosis: A comprehensive review. *Computers in biology and medicine*, *111*, p.103346.

45. Alizadehsani, R., et al, 2018. Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. *Computer methods and programs in biomedicine*, *162*, pp.119-127.

46. Abdar, M., et al, 2019, February. Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 26-30). IEEE.

47. Alizadehsani, R., et al, 2012. Diagnosis of coronary artery disease using data mining techniques based on symptoms and ecg features. *European Journal of Scientific Research*, *82*(4), pp.542-553.

48. Singh, A. and Payal, A., 2021. CAD Diagnosis by Predicting Stenosis in Arteries Using Data Mining Process.In *International Conference on Innovative Computing and Communications* (pp. 689-699). Springer, Singapore.