

Risk Assessment of Cardiovascular Diseases Using kNN and Decision Tree Classifier

Abhijit Dnyaneshwar Jadhav*, Santosh V. Chobe

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune 411018, India

Corresponding Author Email: abhijit.jadhav29@gmail.com



<https://doi.org/10.18280/ria.360118>

ABSTRACT

Received: 8 November 2021

Accepted: 28 December 2021

Keywords:

cardiovascular, machine learning, decision tree, kNN, risk assessment

As per the information provided by WHO, most of the people die from the cardiovascular disease. In 2019, almost 32% of the global deaths were due to cardiovascular diseases, out of which 85% were because of heart attacks and strokes. Hence, it is very important to predict the chances and risk of cardiovascular disease event, to prevent any damage in future. Cardiovascular diseases are the disorders of blood vessels supplying blood to heart, brain and different parts of the body. There are different causes of cardiovascular diseases, which can be quantified with the help of different features and with supporting attributes like age of the person, any diseases like diabetes, blood pressure etc., the risk of the cardiovascular disease can be assessed to prevent the further losses. Machine learning approach is very useful in these circumstances, where quantified data values are available in terms of data set. Machine learning techniques can be used to find the risk of cardiovascular disease. Here, we are proposing to use the two machine learning classifiers such as kNN and decision tree. kNN helps us to find the possibility of cardiovascular disease and decision tree helps us to classify the type of the cardiovascular disease with the risk involved. This approach is very useful, as decision tree is one of the most accurate classifiers, which also helps us to identify the specific cardiovascular disease that can be the future event based on feature values. This proposed methodology is justified with proper research gap stating the important of the proposed architecture and implementation results, which gives effective way for assessing the risk of cardiovascular disease.

1. INTRODUCTION

Cardiovascular diseases are one of the serious reasons of the human deaths worldwide. It damages the heart, kidney, brain etc. body organs. It is mainly caused due to fatty deposits in the body in excess. Though it can be prevented through healthy lifestyle and exercise, it needs awareness about the disease by predicting the value of risk of having the disease. Like, the amounts of fats in the body, there are many other factors, which can be effectively used to measure cardiovascular disease risk prediction [1]. The machine learning, which is the branch of artificial intelligence has a power to develop the powerful prediction value solutions for such problems. Machine learning helps us to develop the models, which can provide high accuracy by using amount of data available [2]. In this research we are considering 30 important features for predicting the cardiovascular disease risk. The features like smoking, excessive drinking, obesity, large waistline, abnormal cholesterol, etc. are few of the important features which have direct and high correlation with the cardiovascular disease. With the help of machine learning techniques, here researchers are analyzing the values of these features and identifying the cardiovascular disease risk [3]. Every medical diagnosis is an important thing and manual process involves the human expertise, where accuracy in diagnosis plays vital role. In manual process, there are the chances of misinterpretation of the data features and it can lead to inaccurate as well as complex conclusions. In cardiovascular disease, such things are very critical and accuracy in diagnosis, prediction of results should properly be interpreted from the

available data records. Hence, there is a need to automate this process and predict the cardiovascular disease risk using automated model. The automation is achieved through machine learning model. Machine learning consists of different types like supervised machine learning, unsupervised machine learning, semi supervised learning. Each type has several techniques for solving the problems in those categories [4]. Classification is the type of Supervised learning, where clustering is the type of unsupervised learning. Classification is the process of classifying the input into the desired class, where the class labels are available. The classification problem can be solved by using different techniques and algorithms such as Support Vector Machine (SVM), Decision Tree, k nearest neighbor (kNN) etc. Clustering is the unsupervised machine learning, where the data inputs are classified into the group of closely similar data items. In clustering, the output group labels are not available and the data items are categorized based on the likeliness of the data items along with consideration of dissimilarity measure [5]. In medical diagnosis of cardiovascular disease, we can predict the results based on the available information, which is specifically defined at the time of problem definition, hence this problem of cardiovascular disease risk assessment and prediction is categorized to supervised learning problem. Hence this problem can be solved by using supervised machine learning techniques. In this research work, we are using the two important classification techniques such as k nearest neighbor (kNN) and decision tree. These are the techniques which can effectively generate the significant results in handling the problem of assessment of risk for cardiovascular disease [6].

Machine learning automated models are mostly very effective and better performance models when developed with a combination of machine learning techniques, instead of standalone techniques. Here, we are taking the benefit of the same and proposing a model with a combination of these two classification techniques. Decision tree helps us to identify the type of the disease and kNN helps us to identify whether there is a risk of cardiovascular disease [7]. This research states about the innovative model for automation of risk assessment of cardiovascular disease, the algorithm, and the results obtained while actually testing the model for different data inputs available. The significance of the results obtained is discussed in the results section. The work is very useful and one of the important innovations in the medical field, as it will help to predict the cardiovascular disease risk and hence alerting the human beings for the same, to avoid the chances of getting into risk of cardiovascular disease. This will definitely help us to save million lives, if implemented in the real-world scenario.

2. RELATED WORK

The thorough literature is done to understand the cardiovascular disease as well as the application of machine learning to cardiovascular disease and risk predictions. Existing systems with results have been studied and the insights from these studies are noted with their advantages and disadvantages. This literature has helped us to find the research gap and frame the solution effectively. The existing work related to this research is stated here.

Krittawong et al. [8] have presented the meta-analysis of cardiovascular disease prediction using machine learning techniques. The results shown are the comparative results analysis of different machine learning techniques. The limitations of the work are, the heterogeneity increases, the data sets used are of small size and that too are divided into parts arbitrarily for training and testing of the models which can affect the results and performance, no proper use of feature selection techniques as using irrelevant features and excluding relevant features leads to bad results, evaluation matrices are also not explicitly prepared. The work should be carried out in systematic and sophisticated way, selection of data, feature selection etc. are important things to generate the correct and trusted results.

Siontis et al. [9] have presented the comparison of different risk predictions in cardiovascular disease models. It is found that, the results obtained by the researchers for their models have several affecting parameters, assumptions which makes their model better as compared to other models. Hence, it is a need to compare all other models in the same operating environment to correctly conclude the better performing models. As per the study, authors are claiming that, the on model can perform better than another model and based on this the selection of the better model should be done. The authors have not presented the study about comparison of few specific models and the approach described in the article is generalize.

Ahmad et al. [10] have presented the work of finding the insights into prognostication, categorization, and assessment of therapeutic heterogeneity. In this work, the machine learning model is developed using clustering technique. The model consists of creating cohort of >40 000 HF patients in Sweden, using the k Means clustering technique. Four novel phenotypes of disease are generated using cluster analysis within this population. Despite being entirely data driven, the

generated phenotypic clusters were recognizable clinically and consisting of strong prognostic value, in more amount as compared to existing one. So, herewith researchers have concluded that, using machine learning approach for heart failure predictions gives better results and helps in diagnosis and predicting the possibilities of the disease.

Weng et al. [11] have stated the use of machine learning in risk prediction of cardiovascular diseases. They have studied different machine learning models in the area. The authors have concluded that, machine learning models are very helpful in predicting the cardiovascular disease risk. The machine learning model helps us to accurately detect the cases of cardiovascular disease and correctly the cases of non-cardiovascular disease.

Willeit et al. [12] used NT-proBNP concentration assessment for prediction of risk assessment in heart failure and cardiovascular disease. The aim of the authors was to study the association between the NT-proBNP concentration and heart disease predictions. The study reveals that, NT-proBNP concentration assessment helps in risk assessments of cardiovascular disease and it is found helpful in predictions of heart disease. From, this it can be realized that, the parameters like NT-proBNP concentration can help us to develop the effective machine learning model for risk predictions of cardiovascular disease.

Ioannidis and Tzoulaki [13] have stated the study of biomarkers in the blood for predictions of cardiovascular disease risk. It is found from the study that, mostly used and preferred by researchers in the world are 10 different biomarkers in the blood. And based on the analytical study, it is found that, these biomarkers have limited or no capability for predictions of cardiovascular disease risk. Hence, it is clear that, we should have a model or system which can help us to predict such risk and machine learning is one of the solutions for the same.

Gaziano et al. [14] have presented the laboratory and non-laboratory methods for predictions of cardiovascular disease risk. The laboratory-based predictions required the highly equipped laboratories, where the large equipment of high cost is required. It is not possible in developing countries to develop such laboratories easily with the huge investments. Hence, the authors have suggested that, by using the large size datasets, the non-laboratory methods can help us for predicting the cardiovascular disease risk. Also, in laboratory, the parameters we use for the predictions are limited to few only, but when we use datasets, a large and relevant parameter values help us to achieve the accurate and useful insights for the predictions.

Krittawong et al. [15], the authors have presented about using artificial intelligence models for cardiovascular disease and clinical care management. The authors have presented the work, how artificial intelligence techniques like deep learning helps to detect the novel genotypes and phenotypes in the cardiovascular disease. Artificial intelligence ensures the precision in cardiovascular medicines also, as well the predictions in cardiovascular risk for the patients. It helps in managing the clinical care of the disease by predicting the things in risk assessment and management of cardiovascular disease.

Wilson et al. [16] have presented the predictions of cardiovascular disease risk using the risk factors. The authors have presented the work with the logistic method. It is compared and continued from the previous work using the regression methods. The authors have considered the risk

evaluation factors like blood pressure, age, sex, cholesterol level, diabetes etc. which does help in predicting the cardiovascular disease risk with high accuracy.

Jabbar et al. [17] have proposed a novel approach using the genetic algorithm. The cardiovascular disease prediction using the association rule mining is the first step of the model. After finding the association rules, the optimized genetic algorithm is applied to predict the risk of cardiovascular disease. The authors have concluded that, the effective results can be obtained if, the genetic algorithm is applied to optimize the generated association rules with association rule mining. The fitness function is the challenging issue, when using the genetic algorithms, which affects the accuracy of the results.

Dangare and Apte [18] have implemented and compared the three different classification techniques for cardiovascular disease prediction. The techniques used are Decision Tree, Naïve Bayes and neural networks. The authors have claimed that, neural networks provide the higher accuracy as compared to decision trees and naïve bayes. To get the desired accuracy importantly two attributes are used for the said work, attributes are obesity and smoking. The disadvantage is that, the results are generated using very small number of samples, which affects to greater extent in predictions and also, the attributes used are small in number, where it is very important to extract all the features which have high correlation with the prediction variable.

Deekshatulu and Chandra [19] have presented the work of heart disease predictions using the classification approach of k nearest neighbor and the genetic algorithm. The authors have concluded that, the convincing accuracy is observed during execution of the model and it can help doctors for better diagnosis of the patients using predictions in heart diseases. The use kNN and genetic algorithm is useful for properly classifying the predictions and hence, generating the correct results for both heart disease cases and non-heart disease cases.

Amma [20] has used the neuro-fuzzy and genetic algorithm for developing the model for predicting the cardiovascular diseases. The work is based on the theoretical foundation and assumptions. The authors are claiming that, if genetic algorithms are used with neuro fuzzy system, error rate can be reduced and also doctor expertise can help to generate the better results.

From this detailed survey study, the conclusion is that, no accurate and proper working model is available to predict the cardiovascular diseases and risk associated with it. Each method used can generate the better results considering the features used, in that too machine learning is the effective approach for these predictions [21]. Hence, by using the effective combination of two classification machine learning techniques like decision tree and k nearest neighbor, derived as per research gap observed in the literature, the proposed methodology is implemented.

3. DEFINITIONS AND CONCEPTS

In the proposed architecture the combination of different machine learning techniques is used.

Here, we are introducing two important machine learning techniques as follows:

1. k Nearest Neighbor (kNN)
2. Decision Tree

3.1 k Nearest neighbor (kNN)

K nearest neighbor is the machine learning techniques used or classification. Sometimes, it is also used as kNN clustering, because of its working structure [22]. kNN is one of the faster classification techniques. It is popular because of negligible training time requirement. In this technique the algorithm trains itself with every input value passed for the classification, hence no separate training phase is required resulting in faster execution of the algorithm [23]. During classification of the input data, the closeness of the data item to be classified is measured as a distance with other classified data items of different classes and the data item will be classified to the class, where the distance is minimum with more number of data points in a particular class. Hence, kNN works based on similarity measure of the data points and very useful in classifying the unknown data points in most of the cases.

3.2 Decision tree

Decision Tree is the supervised machine learning technique used for multiclassification of the input data. Decision tree are constructed with decision nodes as internal nodes and the class labels as the leaf nodes of the tree. Each internal node is defined for the decision for classification path, based on the values of different features used for classification as input variables [24]. There are number of algorithms used for construction of the decision tree such as ID3, C4.5, VFDT, etc. During training phase, with the available input and output data values, the decision tree is constructed. During testing phase, based on the input values of the features, the path of the output class is decided and finally the input is classified [25]. Decision trees are popular for their multiclassification property, as well as speed of the execution with accuracy in results.

4. SOLUTION METHODOLOGY

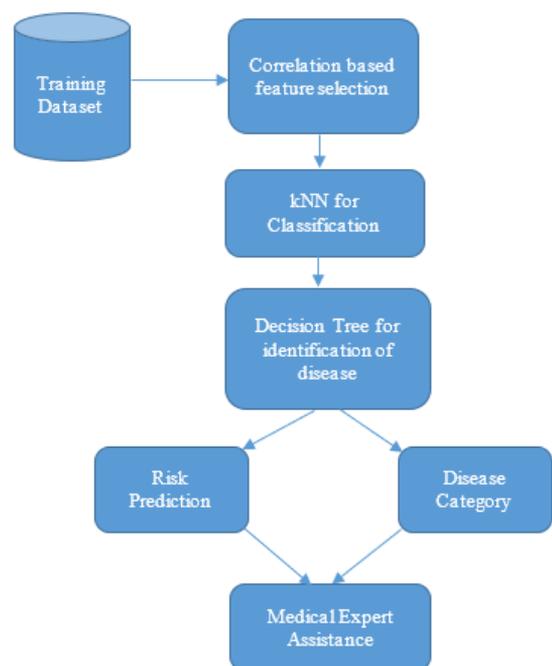


Figure 1. System architecture

As shown in Figure 1, the machine learning techniques: kNN and decision tree are used for identification of the cardiovascular disease and risk assessment of the same. The correlation-based feature selection approach is used for removing redundant and irrelevant features from the dataset. After feature selection kNN is used for classification of the disease, which in turn will be passed to decision tree, which will identify the category of the disease as well as risk involved to the health of the patient. This architecture is helpful for the unknown or new input data values also. The model will provide good accuracy as compared to existing approaches, as it is a hybrid model acting as an expert system for cardiovascular disease identification. This model helps us to first categorize the case as cardiovascular disease or non-cardiovascular disease by using the classification technique kNN. After which, the input is passed to the decision tree to classify the type of cardiovascular disease based on the decision attributes which are applied at the internal nodes of the decision tree. Here, different decision attributes are used like gender, age, blood pressure, diabetes, smoking etc. to classify the input data to the appropriate class label of the cardiovascular disease. This architecture model helps to automate the process of medicinal prescriptions also based on the risk associated with the disease. As based on the risk, the medical assistance will be decided and provided to the patients if needed, without any manual intervention in decision making of predictability of the cardiovascular decision process.

5. DATA SET AND DATA PREPROCESSING

The dataset used here is heart disease dataset for the implementation purpose. Data set consists of 489 records. It does not have the redundant entries of the records. Number of attributes present in the dataset are 30 attributes. These attributes specify the characteristics of the human beings like gender, age, smoking, diet, cholesterol level, biomarkers identified through blood reports etc. which are helpful in predicting the risk of cardiovascular diseases. To develop the model and test its accuracy, in machine learning, model should be trained with sample data records and then by passing the input it will be tested [26]. So, for training and testing of the model, the data set is divided into two parts, training dataset which consists of 295 records and testing dataset which consists of 194 records from the actual heart disease dataset as shown in graph of Figure 2. To achieve the desired accuracy, it is important to use the better quality dataset for training of the model. Hence, if any noisy data available in the dataset, then it should be removed and relevant data values should be available in the dataset.

Here, in the cardiovascular disease cases, the parameters and their values has a specific set of values or specific domain of values and there are very less chances that, the value of any of the parameter falls outside of the domain. In this dataset, the records present are having values from all possible values in the domain. So, when the data is specifically known with the corresponding output label, the problem can be categorized to a classification problem category. Hence, the cardiovascular risk assessment problem is identified as a classification problem and is solved using the classification machine learning model with the heart disease dataset [27]. This testing dataset consists of maximum of the records samples which are passed during the training of the model also, hence it helps to test the classification model effectively.

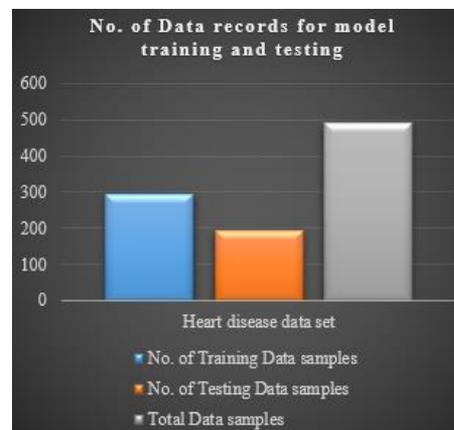


Figure 2. Graph showing training and testing records

As earlier described, the dataset consists of the data with 30 attributes. But, to enhance the accuracy of the classification model, it is important to test the correlation among the input data parameters and the output class label. Here, the correlation based feature selection is used for finding the correlation among the attributes and the class label [28]. Those attributes with high correlation are used and attributes with low correlation value are removed during the training and testing of the model. By using this approach, the accuracy of the model is increased. So, data preprocessing plays a vital role in developing the accurate machine learning models for solutions. From the 30 attributes in the heart disease dataset, by using correlation based feature selection, 21 attributes are selected for training and testing of the implemented model. The list is as given in Table 1. In Correlation based feature selection, the correlation matrix is generated, which consists of the factor value for all possible pairs of independent features and with the given cutoff value only dissimilar features are used for model training and testing. So, as per requirement of machine learning solutions, the correlation matrix helps to determine the correlation between input data parameters. This feature selection helps to use only relevant features and improve the accuracy of the machine learning model to the significant extent.

Table 1. Feature details

Sr. No.	Feature/Attribute
1	Smoking
2	Alcoholism
3	F/U cardiology clinic
4	SBP
5	age
6	Pulse pressure
7	Heart rate, bpm
8	PCI
9	Hypertension
10	Atrial fibrillation
11	Diabetes mellitus
12	Myocardial infarction
13	Stroke/TIA
14	Hemoglobin
15	Creatinine clearance
16	Potassium
17	BNP
18	LDL
19	HbA1C
20	Cholesterol
21	b-Blockers

All these 21 attributes have high correlation with the output class label used for prediction of cardiovascular disease risk.

6. RESULTS AND DISCUSSION

In this machine learning model, the two important classification techniques are used for prediction of cardiovascular disease risk. The k nearest neighbor and the decision tree. K nearest neighbor is used to handle the classification of the data record in either of the category, such as case of cardiovascular risk (Yes) and case of cardiovascular risk free (No). By using, the decision tree, the cardiovascular disease is identified, if kNN categorizes the input data to ‘Yes’ class label. There are different cardio vascular diseases like heart disease, arrhythmias, Aorta disease, congenital heart disease, Coronary artery disease etc. So, when there is a risk of cardiovascular disease, kNN classifies the input to ‘Yes’ class label, enabling the prediction of risk of getting affected with the disease in future and after that, the decision tree by using its internal decision nodes, identifies the cardiovascular disease type. With the help of decision tree classifier, four different types of cardio vascular diseases are considered for model training and testing. The cardiovascular diseases such as Coronary heart disease, Strokes and TIAs, Peripheral arterial disease and Aortic disease are the sample disease records considered in the research work. The graph in Figure 3 shows the number of samples for each cardiovascular disease in training and testing. As shown in the graph, with kNN the accuracy rate is 99.29% for identification of the risk and non-risk cases, which is a better accuracy rate. After passing the input ahead to decision tree algorithm, the correct identification of the cardiovascular disease type is generated by the decision tree and the accuracy obtained is 96.52%, which is one of the better results generated by machine learning model. The results are as shown in Table 2.

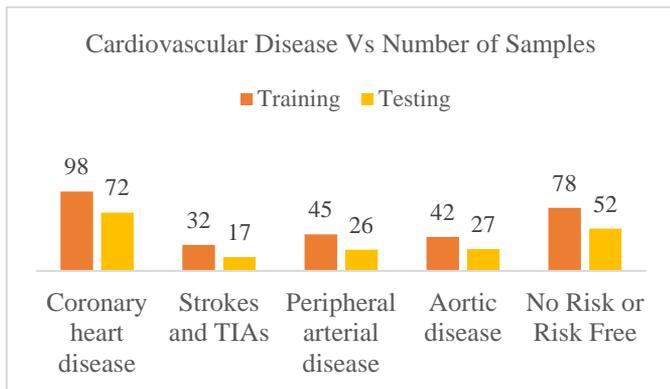


Figure 3. Accuracy comparison with existing models

Table 2. Accuracy results of model with techniques used in the model

Sr. No.	Technique in the Model	Desired Output by the technique	Accuracy Result (%)
1	Our Model: kNN	Class Label as ‘Yes’ or ‘No’	99.29
2	Our Model: Decision Tree	Classifying type of cardiovascular disease	96.52
3	Our Model: Overall Accuracy	Classifying Label as (Risk Vs No Risk) i.e. ‘Yes’ or ‘No’	99.29

The implementation model in [18], have shown the results with 13 and 15 attributes, where the number data samples used by the researchers was only 121, hence accuracy results are almost 99.62%. But, it is important to cover all the cases and samples of the cardiovascular diseases along with all the relevant attributes for developing the model. Here, the result is generated with 489 data records along with 21 attributes selected by using feature selection. The results obtained here have statistical significance, because the cases used for testing are consisting of almost every possible combination of attribute values for identifying the possibility of the cardiovascular disease. The graph of Figure 4 shows the performance enhancements using the implemented model.

This model is very useful in the real scenario and if added some human expertise, can generate the excellent results by helping for early detection of the cardiovascular disease, which will definitely help for saving the lives. In developed countries, the rate of the death due to cardiovascular diseases is huge and hence, such automated machine learning models are useful for identifying the risk of cardiovascular disease in advance and alerting the citizens to prevent any happenings to their health.

Here, the accuracy results are discussed based on one of the important assumptions, that the cardiovascular disease identification problem is a classification problem and testing accuracy samples are the samples from the training samples only, hence the accuracy obtained is 99.29% for the implemented model.

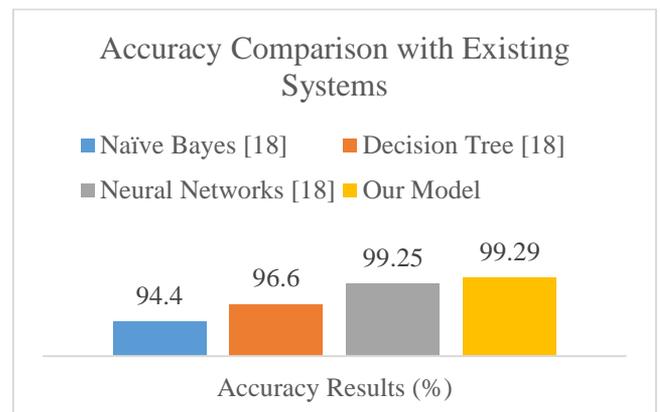


Figure 4. Accuracy comparison with existing models

If, such case is not happening, and the unknown input is passed, then also, the model is using kNN technique, which helps to classify the input based on the distance between the classified data points and the data input to be classified and based on closeness, it classifies it correctly. Hence, the kNN is used in the model which helps in case of unknown data input values also.

7. CONCLUSION AND FUTURE SCOPE

So, herewith the authors have studied the detailed background work in prediction of cardiovascular disease risk. After studying in detail, it is realized that machine learning application to cardiovascular disease risk prediction gives amazing results and helps a lot to improve the disease prediction system, eliminating the erroneous manual disease diagnosis system. In the implemented model use of kNN and decision tree gives better results using the labeled data sets.

The accuracy for risk identification with known data values is 99.29% and proves the significance of the machine learning classification model for cardiovascular disease risk prediction. This system helps the society to get the prior predictions early in the stage to prevent such dangerous diseases and hence save the human lives. While studying the present prediction systems, it is observed that, no proper prediction model is present which can generate faster and accurate results. Hence by using the faster and efficient classification machine learning techniques kNN and decision tree a better model is developed for achieving better cardiovascular disease risk predictions.

ACKNOWLEDGMENT

We are thankful to our respected Management, Principal, Research and Development Department members and all respected faculty members of Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri for their continuous support and motivation for research work and research writing.

REFERENCES

- [1] Alaa, A.M., Bolton, T., Di Angelantonio, E., Rudd, J.H., Van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PloS one*, 14(5): e0213653. <https://doi.org/10.1371/journal.pone.0213653>
- [2] Thomas, M.R., Lip, G.Y. (2017). Novel risk markers and risk assessments for cardiovascular disease. *Circulation Research*, 120(1): 133-149. <https://doi.org/10.1161/CIRCRESAHA.116.309955>
- [3] Kremers, H.M., Crowson, C.S., Therneau, T.M., Roger, V.L., Gabriel, S.E. (2008). High ten-year risk of cardiovascular disease in newly diagnosed rheumatoid arthritis patients: A population-based cohort study. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 58(8): 2268-2274. <https://doi.org/10.1002/art.23650>
- [4] D'Agostino Sr, R B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation*, 117(6): 743-753. <https://doi.org/10.1161/CIRCULATIONAHA.107.699579>
- [5] Bhopal, R.S. (2008). Cardiovascular risk prediction and ethnicity: QRISK2 strengthens creaky foundations. *BMJ*, 336: 1475. <https://doi.org/10.1136/bmj.39609.449676.25>
- [6] Yancy, C.W., Jessup, M., Bozkurt, B., Butler, J., Casey, D.E., Drazner, M.H., Wilkoff, B.L. (2013). 2013 ACCF/AHA guideline for the management of heart failure: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*, 62(16): e147-e239. <https://doi.org/10.1161/01.cir.0000437741.48606.98>
- [7] Emerging Risk Factors Collaboration. (2012). C-reactive protein, fibrinogen, and cardiovascular disease prediction. *New England Journal of Medicine*, 367(14): 1310-1320. <https://doi.org/10.1056/NEJMoa1107477>
- [8] Krittanawong, C., Virk, H.U.H., Bangalore, S., Wang, Z., Johnson, K.W., Pinotti, R., Tang, W.H. (2020). Machine learning prediction in cardiovascular diseases: A meta-analysis. *Scientific Reports*, 10(1): 1-11. <https://doi.org/10.1038/s41598-020-72685-1>
- [9] Siontis, G.C., Tzoulaki, I., Siontis, K.C., Ioannidis, J.P. (2012). Comparisons of established risk prediction models for cardiovascular disease: Systematic review. *BMJ*, 344: e3318. <https://doi.org/10.1136/bmj.e3318>
- [10] Ahmad, T., Lund, L.H., Rao, P., Ghosh, R., Warier, P., Vaccaro, B., Desai, N.R. (2018). Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *Journal of the American Heart Association*, 7(8): e008081. <https://doi.org/10.1161/JAHA.117.008081>
- [11] Weng, S.F., Reys, J., Kai, J., Garibaldi, J.M., Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS One*, 12(4): e0174944. <https://doi.org/10.1371/journal.pone.0174944>
- [12] Willeit, P., Kaptoge, S., Welsh, P., Butterworth, A.S., Chowdhury, R., Spackman, S.A., Di Angelantonio, E. (2016). Natriuretic peptides and integrated risk assessment for cardiovascular disease: an individual-participant-data meta-analysis. *The lancet Diabetes & endocrinology*, 4(10): 840-849. [https://doi.org/10.1016/S2213-8587\(16\)30196-6](https://doi.org/10.1016/S2213-8587(16)30196-6)
- [13] Ioannidis, J.P., Tzoulaki, I. (2012). Minimal and null predictive effects for the most popular blood biomarkers of cardiovascular disease. *Circulation Research*, 110(5): 658-662. <https://doi.org/10.1161/RES.0b013e31824da8ad>
- [14] Gaziano, T.A., Young, C.R., Fitzmaurice, G., Atwood, S., Gaziano, J.M. (2008). Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: The NHANES I Follow-up Study cohort. *The Lancet*, 371(9616): 923-931. [https://doi.org/10.1016/S0140-6736\(08\)60418-3](https://doi.org/10.1016/S0140-6736(08)60418-3)
- [15] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, 69(21): 2657-2664. <https://doi.org/10.1016/j.jacc.2017.03.571>
- [16] Wilson, P.W., D'Agostino, R.B., Levy D, Belanger, A.M., Silbershatz H, Kannel W.B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18): 1837-1847. <https://doi.org/10.1161/01.cir.97.18.1837>
- [17] Jabbar, M.A., Deekshatulu, B.L., Chandra, P. (2012). An evolutionary algorithm for heart disease prediction. In *International Conference on Information Processing*, pp. 378-389. https://doi.org/10.1007/978-3-642-31686-9_44
- [18] Dangare, C.S., Apte, S.S. (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10): 44-48. <https://doi.org/10.5120/7228-0076>
- [19] Deekshatulu, B.L., Chandra, P. (2013). Classification of heart disease using k-nearest neighbor and genetic algorithm. *Procedia Technology*, 10: 85-94. <https://doi.org/10.1016/j.protcy.2013.12.340>

- [20] Amma, N.B. (2012). Cardiovascular disease prediction system using genetic algorithm and neural network. In 2012 International Conference on Computing, Communication and Applications, pp. 1-5. <https://doi.org/10.1109/ICCCA.2012.6179185>
- [21] Mohan, S., Thirumalai, C., Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7: 81542-81554. <https://doi.org/10.1109/ACCESS.2019.2923707>
- [22] Jadhav, A.D., Pellakuri, V. (2021). Highly accurate and efficient two phase-intrusion detection system (TP-IDS) using distributed processing of HADOOP and machine learning techniques. *Journal of Big Data*, 8(1): 1-22. <https://doi.org/10.1186/s40537-021-00521-y>
- [23] Jadhav, A.D. (2021). Two Phase-Intrusion Detection System (TP-IDS) model using Machine Learning Techniques. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(9): 417-425. <https://doi.org/10.17762/turcomat.v12i9.3096>
- [24] Oo, A.N., Naing, T. (2019). Decision tree models for medical diagnosis. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, 3(3): 1697-1699. <https://doi.org/10.31142/ijtsrd23510>
- [25] Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I. (2002). Decision trees: An overview and their use in medicine. *Journal of Medical Systems*, 26(5): 445-463. <https://doi.org/10.1023/A:1016409317640>
- [26] Nalavade, K.C. (2020). Using machine learning and statistical models for intrusion detection. *International Journal of Computer Applications*, 175(31): 14-21.
- [27] Alzahrani, A.O., Alenazi, M.J. (2021). Designing a network intrusion detection system based on machine learning for software defined networks. *Future Internet*, 13(5): 111. <https://doi.org/10.3390/fi13050111>
- [28] Wosiak, A., Zakrzewska, D. (2018). Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis. *Complexity*, 2018: 2520706. <https://doi.org/10.1155/2018/2520706>