


Supporting Real World Decision Making in Coronary Diseases Using Machine Learning

INQUIRY: The Journal of Health Care
Organization, Provision, and Financing
Volume 58: 1–6
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0046958021997338
journals.sagepub.com/home/inq



Peter Kokol, PhD¹ , Jan Jurman¹, Tajda Bogovič¹, Tadej Završnik²,
Jernej Završnik, MD, PhD^{3,4,5} , and Helena Blažun Vošner, PhD^{3,4,6}

Abstract

Cardiovascular diseases are one of the leading global causes of death. Following the positive experiences with machine learning in medicine we performed a study in which we assessed how machine learning can support decision making regarding coronary artery diseases. While a plethora of studies reported high accuracy rates of machine learning algorithms (MLA) in medical applications, the majority of the studies used the cleansed medical data bases without the presence of the “real world noise.” Contrary, the aim of our study was to perform machine learning on the routinely collected Anonymous Cardiovascular Database (ACD), extracted directly from a hospital information system of the University Medical Centre Maribor). Many studies used tens of different machine learning approaches with substantially varying results regarding accuracy (ACU), hence they were not usable as a base to validate the results of our study. Thus, we decided, that our study will be performed in the 2 phases. During the first phase we trained the different MLAs on a comparable University of California Irvine UCI Heart Disease Dataset. The aim of this phase was first to define the “standard” ACU values and second to reduce the set of all MLAs to the most appropriate candidates to be used on the ACD, during the second phase. Seven MLAs were selected and the standard ACUs for the 2-class diagnosis were 0.85. Surprisingly, the same MLAs achieved the ACUs around 0.96 on the ACD. A general comparison of both databases revealed that different machine learning algorithms performance differ significantly. The accuracy on the ACD reached the highest levels using decision trees and neural networks while Liner regression and AdaBoost performed best in UCI database. This might indicate that decision trees based algorithms and neural networks are better in coping with real world not “noise free” clinical data and could successfully support decision making concerned with coronary diseases machine learning.

Keywords

machine learning, decision making, cardiovascular conditions, diagnosing, knowledge discovery, artificial intelligence, heuristics, cross-validation

How does your research contribute to the field?

Machine learning is used to support decision making in medicine, however, it is not yet proven to be effective using real-world data

How does your research contribute to the field?

Provide assessment and rating of the accuracy and suitability of different machine learning methods when trained on real-world medical data

What are your research’s implications toward theory, practice, or policy?

To improve real-world decision making in cardiology aiming to improve health care delivery

Introduction

Machine learning is a popular and an effective approach for problem-solving when a sufficiently representative set of samples (knowledge base) is available to train its

models in different fields like medicine¹ or engineering.² Knowledge base with enough objects (patient data) and balanced distribution of different patient diagnosis, enables various machine learning algorithms to be successfully trained, meaning that achieved accuracy is high enough to



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

support real world decision making. The transition of hospitals to digitally storing patient data in electronic patients' encounters has created a rich source of data that machine learning algorithms can use.

Cardiovascular diseases are the leading global cause of death. Between 2013 and 2016, according to the NHANES (National Health and Nutrition Examination Survey), 48% of the US population over 20 was diagnosed with cardiovascular disease. The risk of these diseases increases with age.³ However, most cardiovascular diseases can be prevented by reducing risk factors⁴

The most common cardiovascular diseases are coronary artery disease, also known as coronary heart disease, caused by a decrease in blood flow to the heart muscle. This phenomenon is caused by a chronic disease called atherosclerosis, which is caused by the accumulation of fat, cholesterol and other substances on plaque-shaped arterial walls.⁵ Risk factors include family history, high blood pressure, smoking, elevated LDL cholesterol levels, diabetes, obesity, lack of recreation, and excessive alcohol consumption⁶

Different factors must be considered when diagnosing cardiovascular diseases and their classification into subgroups. These include, among other things, laboratory measurements, which are suitable for computer processing and machine learning due to their numerical representation.

Unlike expert systems that diagnose and predict based on human-based definitions of a set of attributes, machine learning algorithms can support decision making by analysing complex patterns in the data.⁷ Generally, machine learning algorithms can be categorized in supervised and unsupervised. While supervised algorithms require that 1 or more attributes in the knowledge base have to contain information about the class of the objects (in our case the diagnosis of the disease), unsupervised machine learning approach enables to search for new knowledge without attributes determined by a human observer, but they can find the most relevant attributes and their relationships with each other. The use of supervised learning, however, mostly addresses the problem of classification and risk determination.⁸ A wide range of applications of machine learning algorithms in the field of medicine can be found in Dinu and Joseph.⁹

Using models created through machine learning, individual patients health status can be classified as healthy or ill on the basis of measured traits. This can benefit the user

in diagnosing the disease or serve as a warning sign for a potentially dangerous condition. Machine learning algorithms that allow the visualization of the internal states of the models can also provide the conditions for individual decisions, allowing the user to check the validity of the results and contribute to new knowledge discovery.¹⁰ Machine learning is widely used in cardiovascular medicine¹¹ for the purpose of data interpretation or the analysis of ECGs and imaging systems.¹² The comparison of decision trees, k-nearest neighbors, and support vector machines (SVM) in image recognition for multi sclerosis detection showed that the k-nearest neighbors performed the best,¹³ however the SVM also showed promising result in detection of dendritic spines.¹⁴ Many studies have used different machine learning models, such as Ensemble, Support Vector Machine, Random Forest, and Clustering for classification cardiovascular illness.¹⁵ According to researcher from the paper,¹⁶ logistic regression has showed a good performance for prediction the risk of common cardiovascular diseases. The study recommends using traditional regression models. Mezzatesta et al¹⁷ investigated the use of non-linear SVM with RBF kernel algorithm, optimized with GridSearch in order to improve the accuracy of the algorithm. The results showed high accuracy of the outbreak of cardiovascular diseases in patients on dialysis.

The aim of the study presented in this paper was to perform machine learning on the routinely collected an Anonymous Cardiovascular Database (ACP database) in the University Hospital of Maribor in the period 1999 to 2019 to both support decision making of physicians in their everyday work and possibly discover new knowledge.

Methodology

We performed the study in 2 phases. During the first phase we trained the different machine learning algorithms on the already tried UCI Heart Disease Dataset,¹⁸ which contains heart disease data to test various machine learning algorithms, aiming to select the "best" candidates to be used on the ACP database during the second phase to induce classifiers. 10-fold cross validation was used to assess the accuracy of classifiers. In the ACP we restricted ourselves to the attributes age, gender, re-hospitalization, AH, SB, SP, KBL, S-CRP, S-Lp(a), S-triglycerides, S-HDL-cholesterol, S-cholesterol. The database contains data on 4477 hospital

¹Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

²University Clinical Medical Centre Maribor, Maribor, Slovenia

³Community Healthcare Center dr. Adolf Drolc Maribor, Maribor, Slovenia

⁴Alma Mater Europaea-ECM, Maribor, Slovenia

⁵Faculty of Natural Sciences and Mathematics University of Maribor, Maribor, Slovenia

⁶Faculty of Health and Social Sciences Slovenj Gradec, Slovenj Gradec, Slovenia

Received 12 January 2021; revised 12 January 2021; revised manuscript accepted 3 February 2021

Corresponding Author:

Peter Kokol, Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška ulica 46, 2000 Maribor, Slovenia.

Email: peter.kokol@um.si

Table 1. The Number of Patients in Each of the I25 Sub Codes.

Sub codes of I25	1	2	5	11
Number of patients	761	299	514	45

visits of 3490 unique patients, to which a total of 98 568 laboratory tests were performed.

Unlike the Cleveland database, which is limited to only heart disease data, the ACP database contains cardiovascular disease data. According to the International Statistical Classification of Diseases and Health Related Problems the patients were diagnosed with the diseases code I25 (Chronic Ischemic Hard Disease) and its sub codes (Table 1). After the cleanup, and discarding of sub types containing a smaller number of patients the database contained 1619 patients.

The performance of different machine learning algorithms used in this study was assessed with different metrics.¹⁹ Due to the fact that Accuracy is the most standard metric used in the health sciences machine learning community and that it is easy to understood by health professionals we used only it in the rest of the paper.

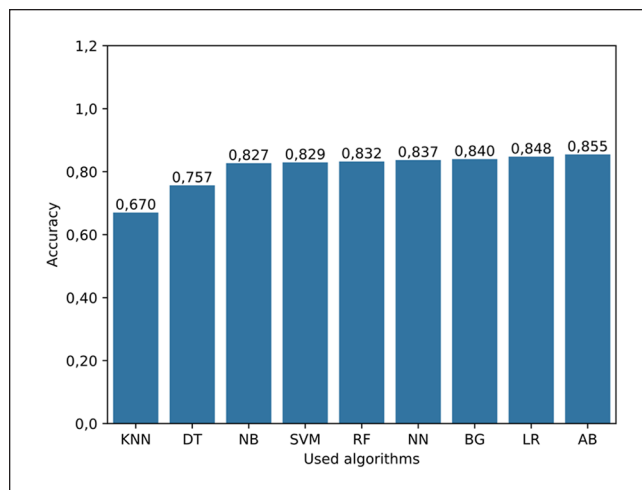
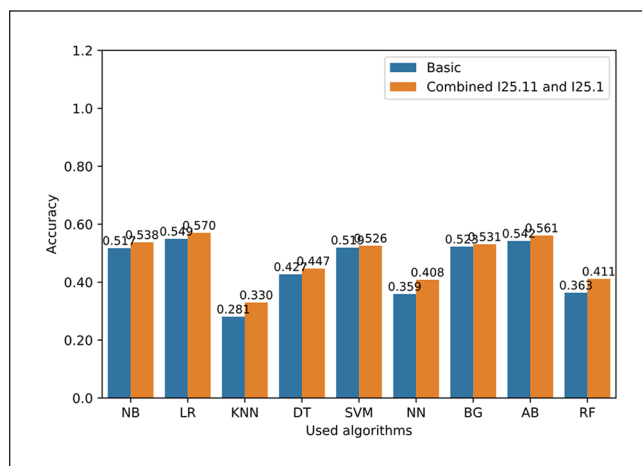
Results

The best classifiers created with individual machine learning algorithms together with their accuracies are presented on Figure 1. Algorithms presented are k-nearest neighbor's (KNN), decision tree (DT), naive Bayes (NB), support vector method (SVM), random forest (RF), neural network (NN), bagging using support vector method (BG), logistic regression (LR), and AdaBoost using logistic regression (AB).

Due to the simplicity of the binary classification problem, most algorithms have achieved a high degree of accuracy. The best results were achieved by a classifier that used AdaBoost and logistic regression as the core classifier (0.855) and the worst by k-nearest neighbor's (0.670). The difference in the accuracy of the use of 10 or 50 base classifiers was negligible, being 0.002. Due to similar results, the classifier with 10 core classifiers was selected as the best, whose learning process is faster due to its smaller size.

The confusion matrices of all classifiers reflected a higher proportion of false negatives than false positives. The algorithms were thus more accurate in the classification of healthy individuals and slightly worse in the classification of patients with heart disease.

The same type of classifiers was induced in the second phase of our study for the 4 diagnoses from the Table 1. The accuracies are shown in the Figure 2 (blue columns). Because the diagnoses I25.1 and I25.11 are similar and due to the limited number of characteristics considered by the classifiers, those 2 were difficult to distinguish. When physician diagnosed patients, they also used symptoms not presented in the ACP database (such as the presence of

**Figure 1.** The accuracies of the best classifiers used on the UCI database in increasing order.**Figure 2.** The accuracies of machine learning algorithms for the ACP database.

angina pectoris), which might affect the accuracy of induced classifiers. When we combined the diagnoses I25.1 and I25.11 we achieved better accuracies shown as orange columns in Figure 2.

We tested different parameter settings for individual classifiers and selected the best ones. In that manner, we used a weighted neighborhood of 20 neighbors for the k-closest neighbor classifier. We did not limit the depth of the decision tree and used entropy for the splitting criterion. A linear hyperplane was used as the kernel for the classifier using the support vector method. The topology of the neural network with the highest accuracy was 3 hidden layers with 200 neurons each. We used 50 base classifiers for bagging and AdaBoost. In the boosting method, we used the support vector method for the base classifier and logistic regression for AdaBoost. For the random forest, we used 500 trees with no depth limit.

In the binary classification, the highest accuracies were achieved by logistic regression (0.570) and its implementation using the AdaBoost ensemble method (0.561), however boosting slightly reduced the accuracy. A high improvement in accuracy when combining I25.1 and I25.11 diagnosis was accomplished in the k-nearest neighbor method, neural network, and random forest, of which the highest improvement was achieved by the neural network whose accuracy increased for 0.049.

In our last experiment we used the ACP database to analyse the accuracy of binary classification (presence or absence of cardiovascular disease). Because the database didn't contain individuals without the cardiovascular condition (PWCD) and we did not have access to data on the simultaneous measured values of the required attributes, we generated samples for PWCD individuals. We balanced the created patient database by generating 1619 PWCD patients, of whom each individual had the same probability of being male and female. We limited the age of patients between 20 and 60 years. The value of the repeat hospitalization attribute was set at 0 for PWCD patients. The number of patients with positive AH, SB, or KBL was followed by the proportion of the global population with the listed medical conditions. 429 PWCD patients had arterial hypertension AH and 142 diabetes mellitus. The heart failure attribute of SP was set to 0 in the generated set, since we treated patients with heart failure as ill. There were 162 PWCD patients with chronic kidney disease.

Although the intervals of the possible values of the attributes specified by the reference values are recommended for the healthy individual, they may deviate due to factors such as lifestyle habits. Thus, for example, a healthy person has higher cholesterol than 5.2, but will not necessarily have cardiovascular disease. We also observed discrepancies between the reference intervals in the literature. Because of this, we have extended the reference intervals by 25% when generating patient attribute values. We did not extend the interval limits below zero. The classification was performed with different parameters for individual classifiers and the best ones were used in our experiment. We employed 30 neighbors and weighted distances for the k-closest neighbors, although the weightless approach had similar accuracy. The decision tree achieved the best results without limiting depth. A linear hyperplane was used as the kernel for the support vector method. The neural network achieved the best results with 3 hidden layers of 20 neurons and without limitation of learning iterations. For the ensemble method, 50 base classifiers were used, where the base classifier was the support vector method for bagging, and logistic regression for AdaBoost. The highest accuracy of patient classification among the presented algorithms was achieved by a random forest with 100 trees and no depth limitation. The accuracy of the binary classification algorithms in healthy and sick patients is shown in Figure 3. The best accuracy was achieved by random forest (0.997)

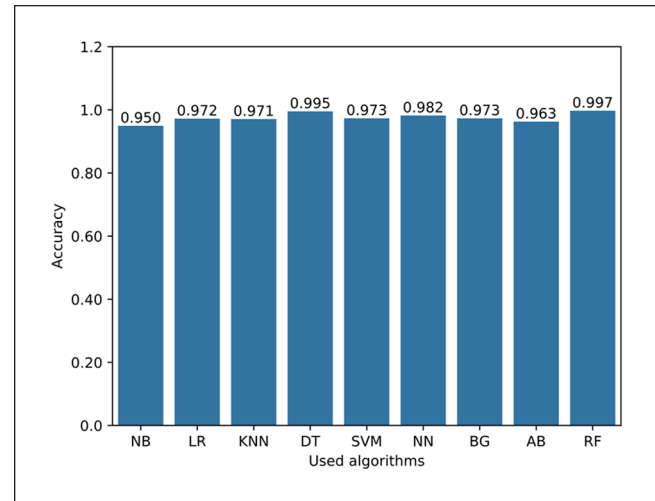


Figure 3. The accuracies of machine learning algorithms for the ACP database (binary classifications).

Table 2. Most Accurate Machine Learning Algorithms.

UCI	ACP-4 diagnoses	ACP-3 diagnoses	ACP-2 diagnoses
AB	LR	LR	RF
LR	AB	AB	DT
BG	BG	NB	NN

followed by decision tree (0.997), and the worst by naïve Bayes (0.950).

The ranking of 3 most accurate machine learning algorithms is shown in the Table 2. In the ACP database the reduction from 4 to 3 diagnoses didn't significantly changed the order. However the reduction from 3 to 2 diagnoses did not only change the order, but also sets of the most accurate algorithms. The most accurate algorithms for the UCI and ACP databases are also totally different.

Feature significance analysis using decision trees¹ revealed that 3 features are the most important in distinguishing between PWCD and ill patients (Figure 4). These are in descending order: Lp(a), CRP, and HDL-cholesterol.

Discussion

The final results show that machine learning can accurately support diagnosing in patients with/without cardiovascular diseases and even help by more precise diagnosing into 4 different cardiovascular diagnoses. It can also support knowledge discovery that is, revealing the various kinds of association of Lp(a) (Lipoprotein A) to CVD events.²⁰ In that manner machine learning can help physicians in cooperation with data scientist to improve the research on CVD and also translate knowledge in clinical practice and thus help to improve health care.²¹ We chose among the most commonly used algorithms for machine learning methods. Our analysis

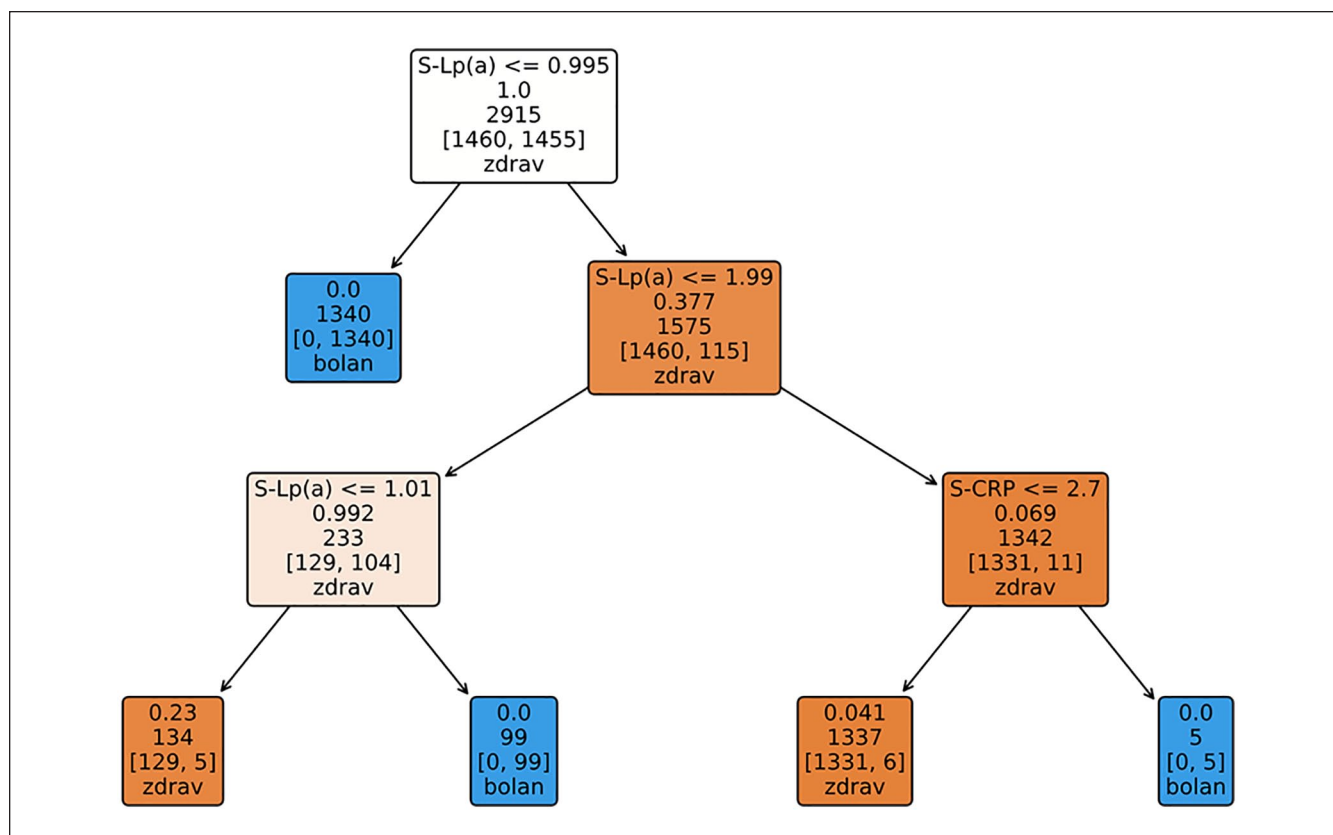


Figure 4. The decision tree induced on the APC database for binary classification.

indicates a significant difference in the datasets, since we used the same set of classifiers in both cases. *In the case of the ACP database most of the algorithms achieved a higher accuracy comparing to UCI dataset.* The UCI dataset was smaller, containing only 297 patients and 13 input attributes were used to classify the disease. The data might have also be affected by other unknown health condition of the patients. Unlike the UCI database, which contained only heart disease data, the APC database also had cardiovascular disease data. As the dataset did not contain healthy individuals we generate a synthetic set of patients, which might be reflected in the accuracy. It is noticed that Random forest, NN and DT proved to be the most accurate algorithms as in many other medical applications REF. In addition to Lp (a), CRP, and HDL-cholesterol were the most important attributes for the presence of a cardiovascular disease as is also shown by Rhee et al.²² CRP is as one of the key indicator of inflammation and tissue damage also known to be highly associated with cardiovascular conditions.²³

Our study had some limitations. The datasets used in our study varied in the presentation of the data. The UCI dataset was significantly smaller and only patients with cardio diseases were included. Therefore, there were a small number of samples and a limited set of attributes. Contrary, there were no k of healthy patients in in the APC database, consequently the generation of healthy individuals might have

influenced the classification results. However, our study has several advantages. We analysed data from a large number of patients with reasonably good data quality which also included the “noise” of the real world clinical environment. The purpose of medical classifiers is to support the physician in real world decision making. We used state-of-the-art ML algorithms which are able to develop the best predictive models.

Conclusion

In summary, cardiovascular disease still poses a high risk of death, so detecting risk factors is very important. We showed that diagnosing could be improved by using machine learning algorithms. Therefore, we implemented, analysed and compared some of the most common used ones. For that purpose, we used an already tried UCI database and a real world database of patients with cardiovascular disease. The accuracy of machine learning differs significantly between the 2 databases, indicating that decision trees based algorithms and neural networks might be better in coping with real world not “noise free” clinical data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Peter Kokol  <https://orcid.org/0000-0003-4073-6488>

Jernej Završnik  <https://orcid.org/0000-0002-5023-0483>

References

- Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst*. 2002;26(5):445-463. doi:10.1023/a:1016409317640
- Zhou J, Qiu Y, Zhu S, et al. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. *Eng Appl Artif Intell*. 2021;97:104015. doi:10.1016/j.engappai.2020.104015
- Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*. 2019;139(10):e56-e528. doi:10.1161/CIR.0000000000000659
- WHO | Global atlas on cardiovascular disease prevention and control. WHO. Accessed March 19, 2020. http://www.who.int/cardiovascular_diseases/publications/atlas_cvd/en/
- Mayo Clinic. Diseases and conditions. Accessed March 19, 2020. <https://www.mayoclinic.org/diseases-conditions>
- Hajar R. Risk factors for coronary artery disease: historical perspectives. *Heart Views*. 2017;18(3):109-114. doi:10.4103/HEARTVIEWS.HEARTVIEWS_106_17
- Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
- Deo Rahul C. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593
- Dinu AJ, Joseph F. A study on Deep Machine Learning Algorithms for diagnosis of diseases | Semantic Scholar. Accessed March 19, 2020. <https://www.semanticscholar.org/paper/A-study-on-Deep-Machine-Learning-Algorithms-for-of-DinuA.-Joseph/513ea1b5f2bbe25522f8c07636d7447fb-d9ea5c>
- Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018;104(14):1156-1164. doi:10.1136/heartjnl-2017-311198
- Kokol P, Mernik M, Završnik J, Kancler K, Malčič I. Decision trees based on automatic learning and their use in cardiology. *J Med Syst*. 1994;18(4):201-206. doi:10.1007/BF00996704
- Tajik AJ. Machine learning for echocardiographic imaging. *J Am Coll Cardiol*. 2016;68(21):2296-2298. doi:10.1016/j.jacc.2016.09.915
- Zhang Y, Lu S, Zhou X, et al. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. *Simulation*. 2016;92(9):861-871. doi:10.1177/0037549716666962
- Wang S, Li Y, Shao Y, Cattani C, Zhang Y, Du S. Detection of dendritic spines using wavelet packet entropy and fuzzy support vector machine. *CNS Neurol Disord Drug Targets*. 2017;16(2):116-121. doi:10.2174/187152731566616111123638
- Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Heart Assoc*. 2019;8(17):e012788. doi:10.1161/JAHA.119.012788
- Nusinovici S, Tham YC, Chak Yan MY, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol*. 2020;122:56-69. doi:10.1016/j.jclinepi.2020.03.002
- Mezzatesta S, Torino C, Meo PD, Fiumara G, Vilasi A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Comput Methods Programs Biomed*. 2019;177:9-15. doi:10.1016/j.cmpb.2019.05.005
- UCI Machine Learning Repository. Heart Disease Data Set. Accessed March 21, 2020. <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Zhou J, Li X, Mitri HS. Classification of rockburst in underground projects: comparison of ten supervised learning methods. *J Comput Civ Eng*. 2016;30(5):04016003. doi:10.1061/(ASCE)CP.1943-5487.0000553
- Marais AD. Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease. *Pathology*. 2019;51(2):165-176. doi:10.1016/j.pathol.2018.11.002
- Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8(2):020303. doi:10.7189/jogh.08.020303
- Rhee E-J, Byrne CD, Sung K-C. The HDL cholesterol/apolipoprotein A-I ratio: an indicator of cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes*. 2017;24(2):148-153. doi:10.1097/MED.0000000000000315
- Halaris A. Inflammation-associated co-morbidity between depression and cardiovascular disease. In: Dantzer R, Capuron L, eds. *Inflammation-Associated Depression: Evidence, Mechanisms and Implications*. Current Topics in Behavioral Neurosciences. Springer International Publishing; 2017:45-70.