A SURVEY ON MACHINE LEARNING ALGORITHM FOR CORONORY ARTERY DISEASE

R. Mahesh Kumar, S. Veni*

Abstract

Coronary Artery Disease (CAD) is the most common form of cardiovascular disease (CVD) and often results in heart attacks. It is responsible for tens of thousands of fatalities and billions of dollars in annual economic damages. Patients with CAD are diagnosed with invasive and possibly dangerous conventional angiography. Machine learning (ML) techniques are often suggested as fast, inexpensive, and non-invasive means of identifying CAD models. Several datasets, sample sizes, data collection techniques, performance assessments, and machine learning algorithms are used in the published works on ML-based CAD diagnostics. Because of these differences, it is hard to generalize about literary accomplishments. In this article, every significant study on ML-based CAD diagnosis published in the last several years is examined. A comprehensive analysis of the effects of numerous factors, including dataset features (such as geographic location and sample size) and machine learning techniques (such as feature set selection, performance measures, and process) are conducted. Finally, the fundamental difficulties and limits of ML-based CAD diagnosis are investigated in depth.

Keywords: CAD, ML, Heart Disease, Survey, CVD

I INTRODUCTION

Coronary Artery Disease (CAD) resulting from coronary artery stenosis (CAS) is a severe public health problem that substantially burdens wealthy countries with aging and expanding populations. It is the leading cause of death and disability on a global scale. In 2005, 7.6 million people died from cardiovascular disease [1]. In 2006, CAD was responsible for about 400,000 deaths in the United States [2]. The predicted direct cost of CAD in the United States in 2020 was $99 billion [3] [4]. Effective CAD screening, diagnosis, and treatment require a novel CAS detection technique that should be user-friendly, cost-effective, low-risk, and accurate to meet these public health and financial issues [5,6,7]. Data mining (DM) in healthcare is a booming field that is becoming more productive and indispensable [8, 9]. Patients, hospital resources, disease diagnostics, electronic patient information, medical equipment, etc., all comprise the modern healthcare system [10,11]. DM provides many tools and procedures for achieving these aims [12]. Over one million people are killed annually by CAD, a grave condition. CAD is fatal in around fifty percent of patients. More than 335,000 people with Coronary Artery Disease (CAD) die with heart attack.

II BACKGROUND STUDY

Arabasadi, Z. et al. [2] suggested a novel hybrid technique to improve neural network performance. The approach was tried on several HD datasets to see whether it improved performance. The proposed strategy may improve the performance of neural networks in terms of CAD identification. Specifically, CAD may be recognized without angiography, reducing high expenses and substantially adverse effects. Several strong evolutionary and swarm intelligence approaches exist, such as evolution strategy, particle swarm optimization, and genetic algorithm.

Anbarasi et al. (2010) [4] suggested improved prediction of HD by utilizing a genetic approach to identify feature subsets. The features were chosen with the aid of the Genetic

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

algorithm to assist more in identifying cardiac conditions. Using a Genetic search, the number of characteristics is decreased to six. Three classification models are given in the reduced dataset. As the testing mode, the K-fold cross-validation approach is implemented.

Bachmann et al. (2012) [6] emphasized that family history might be a valuable tool for identifying younger men at high risk for risk factor management and researching new possible environmental or genetic causes of cardiovascular disease.

Qin, C.-J. et al. [13] created numerous assessment criteria to quantify the characteristics, integrated with a heuristic search approach for Feature Selection (FS). Furthermore, seven standard classification techniques were utilized to categorize CHD. According to the findings of the comparative studies, FS techniques might minimize the size of the dataset while improving classifier performance. The EA-MFS approach was devised on this premise, which merged these several FS methods into one ensemble algorithm. The program made extensive use of various basic classifiers to represent feature relationships. Finally, in the CAD dataset, the EA-MFS technique demonstrated improved classification performance and resilience, which are key directives for the clinical diagnosis of CAD.

III COMPARATIVE ANALYSIS OF SURVEY

Author	Algorithms Used	Accuracy
Abhay Kishore[1]	RNN	RNN accuracy 92%
Avinash Golande[3]	1) DT 2)KNN, 3) K-mean clustering 4) Adaboost	(86.60% Accuracy)
Mr. Santhana Krishnan. J[12]	NB & DT	Decision Tree (91% Accuracy)
Senthilkumar Mohan [17]	DT and SVM	HRFLM (88.4% Accuracy)

Table 1: comparative analysis of various algorithms with accuracy

IV DISCUSSION

4.1 Data Mining Approaches in Medicine

In medicine, illness diagnosis is a vital and difficult endeavor. The tests are necessary to identify a certain ailment are included in the medical data. The doctors are confronted with imprecision in knowledge and decision-making, which prevents them from arriving at definite conclusions for the recommended resolution. Diverse viewpoints on the topic may be gained by collecting data from various sources, which should then be processed to provide an integrated outcome. For this design, DM methods deliver more accurate and comprehensive information. DM employs intelligent techniques to extract data patterns and provides doctors with an excellent chance to manage vast quantities of data. In specialized medical sectors, prognosis and diagnosis based on DM facilitate the classification and precise prediction of patients' ailments.

Bandyopadhyay et al. [7] (2015) used Bayesian classifier models to determine cardiovascular event risks. Individual patient characteristics were used as a crucial management strategy.

Milovic et al.[9] identified the process of evaluating raw data and extracting its significance with the aid of a computer (2012). Currently, healthcare companies create and gather a vast amount of data. An exponential increase in the volume of data necessitates a flexible, automated method for data extraction. Due to the abundance of data in the healthcare industry, DM applications are unavoidable.

In Healthcare department, DM starts with an assumption in medical research, and outcomes are altered appropriately. Medical DM is of tremendous relevance in health care since it satisfies the full wants of the healthcare organization by studying their requirements in depth. Massive advancements in information and communication technology will allow DM to realize its full potential for uncovering knowledge

buried in the enormous healthcare database.

Rafe et al. [14] (2013) claimed that DM in medicine provides new opportunities for examining the hidden laws inside clinical data sets. These guidelines are applicable for illness identification in clinical diagnosis. The user-centric approach of DM technologies uncovers a previously unknown pattern. Diagnosis in the medical data set is increasingly important, but it is still a challenging process that must be performed precisely and competently.

4.2 Contribution of Data Mining to Disease Diagnosis

DM in healthcare is mostly utilized for forecasting different illnesses and supporting physicians with their clinical decision-making. The discussion of different healthcare sector practices is presented.

Acharya et al. [5] (2012) established a CAD approach that helps to distinguish between normal and fatty disease-affected livers. In assessing the suggested approach using a database of 58 abnormal and 42 normal liver ultrasonography patients, the authors applied the decision tree classifier technique and achieved a classification accuracy of 93.3%.

Mamiya et al. [10] (2015) have created a logistic regression model for tuberculosis (TB) patients using clinical and demographic data. As a primary public health activity, they researched the contacts of a newly diagnosed TB patient to prevent TB transmission.

Meng et al. [11] (2013) assessed the efficiency of ANNs and DT models for predicting diabetes using common risk markers. They examined the accuracy, sensitivity, and specificity of all three models. The classification accuracy of logistic regression, ANN model, and decision tree (C5.0) was 76.13 percent, 73.23%, and 77.87 percent, respectively. The decision tree model is one of the three models (C5.0) that had the highest accuracy in classifying.

4.3 Coronary Heart Disease Risk Factors

Dhanashree et al. [8] (2013) introduced the classifier strategy for identifying cardiovascular disease using naive Bayes. The accuracy of the data acquired from the Cleveland clinic foundation with 14 criteria was 89 percent. This model is constructed and refined to discover the appropriate categorization process which will be utilized for clinical diagnosis.

Soni et al. [15] (2011) have used three distinct supervised machine learning methods, Nave Bayes, Neural network, and Decision tree, to evaluate a dataset of 3000 cases with 14 different variables to predict HD.

Srinivas et al. [16] (2010) presented the investigation and prediction of CHD events and heart attacks in coal mining locations using DM approaches. They've created a behavioral risk factor monitoring system with dependent variables such as chest discomfort, stroke, and heart attack. Cardiac care research identifies eleven predictors of morbidity.

Topol et al. [18] (2011) Identify a condition halfway between stable chronic angina and cardiac infarction. It is a clinical diagnosis determined by the absence of myocardial infarction (MI) by electrocardiography (ECG) or cardiac enzyme testing.

S. No.	Merits	Demerits
1	Among the multitude of Selecting the important variables for CHD prediction yields effective performance	Diverse data types in a dataset cannot provide an accurate forecast.
2	The review indicates that huge datasets provide superior performance.	Minimal dataset mining is incapable of producing reliable findings.
3	By mixing and contrasting additional algorithms, interpretability, dependability, and generalizability may be enhanced.	The complicated medical dataset may not suit a single DM approach.

Table 2. Merits and Demerits of the previous research studies on CHD

S.No.	Methods	Advantages	Limitations
1.	Decision Tree	It reduces the uncertainty of difficult judgments. Assigns precise values to the results of many characteristics. In addition to handling both numeric and categorical data, the decision tree can process both types of information.	It is an unstable classifier, meaning its performance relies on the kind of dataset.
2.	Support Vector Machine (SVM)	Superior precision compared to other classifiers. Avoids Overly accommodating	The primary challenge is selecting the proper kernel function. It overcomes the issue of several classes by dividing them into two classes, such as one -against-one and one-against.
3.	Neural Network	Determine complicated correlations between dependent and independent variables with relative ease. Capable of managing noisy data.	Over-fitting. Large neural networks make the processing of ANN networks harder to comprehend and take a great deal of processing time.
4.	. . . .	It facilitates the calculation process. Improve the speed and precision of massive datasets.	In situations when variables are interdependent, it does not provide correct findings.
5.	Hybridization	Compute the difficult problem with the fewest possible characteristics	The efficiency of the hybridization is dependent on the data-driven algorithm selection.

Table 3 Advantages and Limitations of DM method selected for CHD risk prediction

V CONCLUSION

Several forms of ML algorithms for HD [19] prediction have been covered. Various ML algorithms are constructed and aimed to determine the optimum strategy by examining its properties. In distinct conditions, each algorithm produced a different result. Further analysis indicates that relatively little accuracy is acquired for the prediction model of heart illness, suggesting that more sophisticated models are essential to forecast early heart sickness. This papere presents a method for accurately predicting cardiac disease at a cheap cost and with little complexity.

REFERENCES

[1] Abhay Kishore, Ajay Kumar, Karan Singh, Maninder Punia, Yogita Hambir," Heart Attack Prediction Using Deep Learning," International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04, Apr-2018.

[2] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Computer Methods and Programs in Biomedicine, 141, 19–26. doi:10.1016/j.cmpb.2017.01.004.

[3] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[4] Anbarasi, M, Anupriya, E & Iyengar, N 2010, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology, vol. 2, no. 10, pp. 5370-5376.

[5] Acharya, UR, Sree, SV, Ribeiro, R, Krishnamurthi, G, Marinho, RT, Sanches, J & Suri, JS 2012, "Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm", Medical Physics, vol. 39, no. 7, pp. 4255-4264.

[6] Bachmann, MJ, Willis, BL, Ayers, CR, Khera, A & Berry, JD 2012, "Association Between Family History and Coronary Heart Disease Death Across Long-Term Follow-Up in Men", Circulation, vol. 125, no. 25, pp. 3092-3098.

[7] Bandyopadhyay, S, Wolfson, J, Vock, DM, Benitez, GV, Adomavicius, G, Elidrisi, M, Johnson, PE & O" Connor, PJ 2015, "Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data", Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1033-1069.

[8] Dhanashree, SM, Mayur, PB & Shruti, DD 2013, "Heart Disease Prediction System using Naive Bayes", International Journal of Enhanced Research in Science Technology & Engineering, vol. 2, no. 3, pp. 1-5.

[9] Milovic, B & Milovic, M 2012, "Prediction and Decision Making in Health Care using Data Mining", International Journal of Public Health.

[10] Mamiya, H, Schwartzman, K, Verma, A, Jauvin, C, Behr, M & Buckeridge, D 2015, "Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes", Journal of Biomedical Informatics, vol. 53, pp. 237-242.

[11] Meng, XH, Huang, YX, Rao, DP, Zhang, Q & Liu, Q 2013, "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors", Kaohsiung Journal of Medical Sciences, vol. 29, no. 2, pp. 93-99.

[12] Mr.Santhana Krishnan.J, Dr.Geetha.S," Prediction of Heart Disease Using Machine Learning Algorithms",2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.

[13] Qin, C.-J., Guan, Q., & Wang, X.-P. (2017). Application of Ensemble Algorithm Integrating Multiple Criteria Feature Selection In Coronary Heart Disease Detection. Biomedical Engineering: Applications, Basis and Communications, 29(06), 1750043. doi:10.4015/s1016237217500430

[14] Rafe, V & Farhoud, RH 2013, "A Survey on Data Mining Approaches in Medicine", International Research Journal of Applied and Basic Sciences, vol. 4, no. 1, pp. 196-202.

[15] Soni, J, Ansari, U, Sharma, D & Soni, S 2011, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, vol. 17, no. 8, pp. 43-48.

[16] Srinivas, K, Rao, GR & Govardhan, A 2010, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", Proceedings of the fifth ICCSE conference, China, pp. 1344-1349.

[17] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava —Effective Heart Disease Prediction Using Hybrid Machine Learning

Techniques ‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, —On the Compatibility of Adaptive Controllers, ‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings).

[18] Topol, EJ, Califf, RM, Prystowsky, EN, Thomas, JD & Thompson, PD 2011, Textbook of Cardiovascular Medicine, Lippincott Williams and Wilkins, USA.

[19] Yang, J., Tian, S., Zhao, J., & Zhang, W. (2020). Exploring the mechanism of TCM formulae in the treatment of different types of coronary heart disease by network pharmacology and machining learning. Pharmacological Research, 105034. doi:10.1016/j.phrs.2020.105034.