

Machine Learning Techniques for improving study of Heart Disease Prediction System

Abdelmegeid Amin Ali¹, Ashraf Heikal², Eman M. Anwar^{2,3}, Shaimaa M Hussien^{2,3}

¹Faculty of Computers and Information, Department of computer science, Minia University, Minia, Egypt

²Al-Obour High Institute for Management and Informatics, Obour, Egypt, Department of information systems

³Faculty of Computers and Information, Department of Information System, Minia University, Minia, Egypt

Corresponding Authors Email: abdelmegeid@yahoo.com

Received 18 May 2021; revised 3 July 2021; accepted 23 August 2021

Abstract – The heart is an important organ in human beings. Because even a slight error might result in weariness or death, diagnosing and forecasting cardiac disorders requires increased precision, perfection, and accuracy. There are innumerable heart-related deaths, and the number is increasing significantly every day. To address the issue, researchers use a variety of data mining and machine learning approaches to evaluate massive amounts of complex medical data, assisting healthcare providers in the prediction of heart disease. Using various data mining approaches, the suggested research predicts the likelihood of heart disease and categorizes the risk level of patients. When compared to other machine learning algorithms, the trial results show that the bagging technique with decision tree algorithm has the highest accuracy of 88.56%.

Keywords – Heart disease Classification, Machine Learning, Ensemble method

I. INTRODUCTION

One of the most essential organs in humans is the heart. It is a muscular organ that pumps blood into the body and is an essential component of the cardiovascular system [1]. The cardiovascular system is made up of all blood vessels, such as arteries, veins, and capillaries, which form a complex network of blood vessels all over the body [2]. Any obstruction or abnormality in normal blood circulation or flow from the heart could result in a slew of pains caused by coronary heart disease. These are commonly known as cardiovascular diseases (CVDs) and are among the world's deadliest ailments. CVDs include diseases of the coronary heart, cerebral vascular diseases, and blood vessel diseases [3]. CVDs are the greatest cause of mortality and disability worldwide, according to the World Health Organization (WHO) Report international Atlas on upset interference and management [4]. Despite the fact that CVDs can be prevented through lifestyle changes and other related measures, they are nevertheless on the rise on a daily basis, according to several WHO publications.

However, multiple WHO investigations have shown a global increase in CVDs, which is extremely concerning. Cardiopathy affects both men and women in the same way. According to the World Health Organization, 17.9 million people died in 2016 as a result of heart disease, accounting for 31% of all fatalities worldwide. Stroke and coronary failure account for 85% of these deaths (WHO2016) [26]. Heart diseases occur when neither the heart nor the blood arteries work normally. Another issue with upset is induration, which is commonly defined as arterial hardness. In this instance, the arteries become thicker and more rigid. Arteriosclerosis is the narrowing of blood vessels, resulting in decreased blood flow through the buildups.

Heart attacks happen when blood clots, or blockages, form in the arteries, disrupting blood flow. These illnesses are also the result of a severe cellular physiological state disturbance produced by broad genetic and molecular abnormalities in cells. As a result, people who are afflicted with or at high risk of developing certain diseases would benefit from early detection and tailored medical treatment. Smoking is one of the most dangerous things that has an impact on heart health. High cholesterol, high blood pressure, physical inactivity, a poor diet, obesity, and poorly treated polygenic disease are only a few of the variables that contribute to the development of heart disease [27]. As a result, particular aspects related to mode must be addressed in order to examine the hazard of vascular disease. As a result, essential tests such as cholesterol, electrocardiograms, chest discomfort, blood pressure, highest heart rate, and excessive sugar levels should be performed by patients to swiftly expose and predict suitable guiding scenarios. Some estimates and circumstances make it even more difficult to compare medical practitioners' work to existing patient check outcomes [28]. Cardiopathy is usually diagnosed by collecting a medical history, using a stethoscope, ultrasound, and performing a diagnostic technique (ECG).

The doctor's guess, expertise, and experience are used to announce the conclusion and compare it to the prior information stored in the database to determine if a patient with a specific illness is normal or aberrant [29]. One of the disadvantages of the

- [16] Dr.S.Seema Shedole, Kumari Deepika (2016). Predictive analytics to prevent and control chronic disease, <https://www.researchgate.net/publication/316530782>, January 2016.
- [17] K.Gomathi, Dr.D.Shanmuga Priyaa (2016). Multi Disease Prediction using Data Mining Techniques, *International Journal of System and Software Engineering*, December 2016, pp.12-14.
- [18] Ashwini Shetty A, Chandra Naik (2016). Different Data Mining Approaches for Predicting Heart Disease, *International Journal of Innovative in Science Engineering and Technology*, Vol.5, May 2016, pp.277- 281.
- [19] Ayon Dey, Jyoti Singh and Neeta Singh (2016). Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using Principal Component Analysis), *International Journal of Computer Applications* (0975 – 8887) Volume 140 – No.2, April 2016.
- [20] Eva Tuba et al. “Classification and Feature Selection Method for Medi- cal Datasets by Brain Storm Optimization Algorithm and Support Vector Machine”. In: *Procedia Computer Science* 162 (2019), pp. 307–315.
- [21] Alexander Wood et al. “Private naive bayes classification of personal biomedical data: Application in cancer data analysis”. In: *Computers in biology and medicine* 105 (2019), pp. 144–150.
- [22] Sandhya Harikumar. “Blended Models for Nearest Neighbour Algorithms for High Dimensional Smart Medical Data”. In: *Smart Medical Data Sens- ing and IoT Systems Design in Healthcare*. IGI Global, 2020, pp. 48–75.
- [23] Diyang Xue, Adam Frisch, and Daqing He. “Differential Diagnosis of Heart Disease in Emergency Departments Using Decision Tree and Medi- cal Knowledge”. In: *Heterogeneous Data Management, Polystores, and An- alytics for Healthcare*. Springer, 2019, pp. 225–236.
- [24] Changsheng Zhu, Christian Uwa Idemudia, and Wenfang Feng. “Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques”. In: *Informatics in Medicine Unlocked* 17 (2019), p. 100179.
- [25] David, H. B. F. Impact of ensemble learning algorithms towards accurate heart disease prediction.
- [26] Ali, A. A., Hassan, H. S., & Anwar, E. M. (2020, July). Improve the accuracy of heart disease predictions using machine learning and feature selection techniques. In *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences* (pp. 214-228). Springer, Singapore.
- [27] Gao, Xiao-Yan, et al. "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method." *Complexity* 2021 (2021).
- [28] Ali, A. A., Hassan, H. S., Anwar, E. M., & Khanna, A. (2021). Hybrid technique for heart diseases diagnosis based on convolution neural network and long short-term memory. In *Applications of Big Data in Healthcare* (pp. 261-280). Academic Press.
- [29] Ali, A. A., Hassan, H. S., & Anwar, E. M. (2020, July). Heart diseases diagnosis based on a novel convolution neural network and gate recurrent unit technique. In *2020 12th International Conference on Electrical Engineering (ICEENG)* (pp. 145-150). IEEE.