IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005259 |

Predicting the Heart Disease Using Machine Learning Approach

Prashant Ghodeswar, Radha Shirbhate

PG Student, ME Computer Engineering, SPPU Pune, Maharashtra, India

Assistant Professor, BSIOTR Wagholi, Pune, Maharashtra India

ABSTRACT: Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this work, we propose a approach that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of disease. The prediction model is introduced with different combinations of features and several known classification techniques.

KEYWORDS: Segmentation; Support vector machines (SVMs); KNN Algorithm; Decision Tree

I. INTRODUCTION

Heart disease is a broad term that refers to various heart-related diseases. “Heart disease” refers to various conditions, including blood vessel illnesses such as coronary artery disease, abnormal heart rhythms (arrhythmias), and congenital heart abnormalities. Occasionally, the terms “heart disease” and “cardiovascular disease” are used synonymously. Cardiovascular disease (CVD) is a broad term that encompasses diseases characterized by constricted or blocked blood arteries, which may affect a heart attack (myocardial infarction), chest discomfort (angina), or stroke. Other types of the cardiac disease include those that damage the heart muscle, valves, or rhythm [1]. Each year, 17.9 million people worldwide die of CVDs, accounting for about 31% of all fatalities. Nowadays, the healthcare industry generates a great deal of data on patients, illness diagnoses, etc. However, researchers and practitioners do not make effective use of this data. Today, the healthcare sector faces a significant challenge: service quality (QoS). Quality of service entails accurately identifying illness and treating people effectively. Unacceptable outcomes may result from inadequate diagnosis. Numerous aspects raise your coincidental of rising heart disease. Some hazard factors are uncontrollable, such as family history, increasing age, ethnicity, and maleness. However, smoking, diabetes, hypertension, inactivity, and being overweight or obese are all risks that may be managed or avoided [2].

Machine learning enables computer systems to learn from predefined facts, improve performance via experience, and then use what they have learnt to make an educated choice without human involvement. Each time a machine learning algorithm makes a successful choice, the program’s performance is enhanced. The process of knowledge discovery from data (KDD) is shown in the following picture.

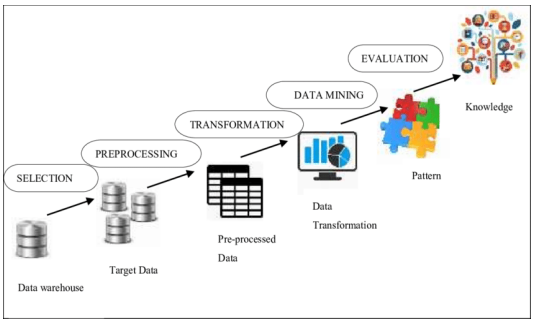

Fig.1. Steps in Knowledge Discovery Process

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005259 |

II. RELATED WORK

In [1] asserts that early diagnosis and ongoing monitoring of cardiac problems may significantly decrease mortality. The exponential rise of information from many causes, such as wearable sensor gadgets used in the Internet of Things health intensive care, spilling systems, and others, has resulted in the continual generation of huge amounts of data. Combining big information analytics with machine learning is a game-changing technique that can revolutionize the healthcare industry, particularly in the early identification of heart disease. This technology has the potential to be more powerful while being more affordable. Apache Spark has been used to offer a real-time cardiac disease estimate system. The system is divided into two major components: stream processing and information storage and display. The first usages Spark MLlib in conjunction with Spark spilling to forecast cardiac disease by applying a classification model to data events. Apache Cassandra is used to storing a massive amount of produced data in seconds.

According to [2] heart disease is one of the major reasons for death worldwide. The authors examined many novel strategies for determining major risk factors for heart disease. They analyzed two sets of heart disease data (Cleveland and Hungarian), split into 33%, 66%, and 100% segments. The values of various individual characteristics are determined in this data to ascertain the disease's significant components. Then, several semi-supervised learning algorithms are employed to evaluate heart disease data. The accuracy, f-measure, and area under these classifiers' receiver operating characteristic curve (ROC) were computed to validate their use and define the optimal semi-supervised learning method. This method investigated both important and irrelevant risk factors for heart disease by systematically eliminating characteristics and monitoring the classification results. The analysis's efficacy and efficiency are shown experimentally on two actual data sets.

III. PROPOSED ALGORITHM

Machine learning is used to develop a prediction system for heart disease. The primary advantage of machine learning is that it has a high degree of discrimination. The suggested model's flowchart is shown in Figure.

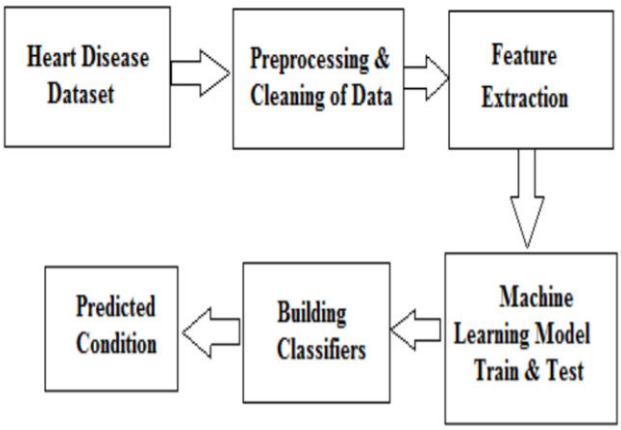

Fig.2. Framework for Heart Disease Classification

The dataset contains various parameters of the patients like age, sex, blood pressure, type of chest pin etc. As shown in the figure the dataset is first preprocessed and cleaned to remove any missing values and uniform the parameters [12]. Then it is passed on to feature extraction stage in which the features are extracted for all the individuals. The classifiers are then utilized to classify the various features based on ground truth taken from other sources for classification. Machine learning models can be developed using numerous techniques like SVM, KNN, Artificial Neural Network etc. These machine learning models are trained based on the features to classify the dataset as belonging to either healthy or affected patients

IV. SIMULATION RESULTS

Machine learning has a significant contribution in dealing with the problems of extracting features from input dataset to solve predictive tasks such as prediction of a disease, decision support, analyzing risk anomalies. Thus, predictive models are applied in the diagnosis and forecasting of the diseases. EMR is communal database for clinical disease and risk prediction, giving exclusive diagnostic opportunities & challenges. A continuously growing large amount of EMR has to be processed and analyzed using data mining and text mining to improve accuracy of disease prediction and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005259 |

future health risks. Further, algorithms ranging from machine learning to deep learning can be applied to achieve better results in healthcare systems.

From From the review it is observed that, most of the researchers consider only structured data for the prediction system, but a lot of valuable medical information is accumulated by unstructured texts. Thus, it becomes important to process and analyze the unstructured information due to lack of common structure using efficient algorithm. Machine learning models yield better performance in many tasks and require less manual feature engineering. In the review, several methods of Machine learning and Deep learning are discussed but low accuracy problem still exists, and there is a scope to improve the prediction performance of disease to achieve higher results in terms of accuracy.

REFERENCES

1. M. Ghassemi, T. Naumann, P. Schulam, L. A. Beam, I. Chen, and R. Ranganath, "A Review of Challenges and Opportunities in Machine Learning for Health", University of Toronto and Vector Institute, Toronto, Canada, pp. 191–200, 2020.
2. J. Lopes, T. Guimarães, M. F. Santos, J. Lopes, T. Guimarães, and M. F. Santos, "Predictive and Prescriptive Analytics in Healthcare: A Survey", Procedia Computer Science, Published by Elsevier, vol. 170, pp. 1029–1034, 2020, doi: 10.1016/j.procs.2020.03.078.
3. K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, "Prescriptive analytics: Literature review and research challenges", International Journal of Information Management, vol. 50, pp. 57–70, February, 2020, doi: 10.1016/j.ijinfomgt.2019.04.003.
4. K. Hasan, A. Alam, D. Das, and E. H. Senior, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers", IEEE Access, Special Section on Deep Learning Algorithms for Internet of Medical Things, vol. X, pp. 1–19, 2020, doi: 10.1109/ACCESS.2020.2989857.
5. F. Ali, S. Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, "A Smart Healthcare Monitoring System for Heart Disease Prediction Based On Ensemble Deep Learning and Feature Fusion", Information Fusion, Published by Elsevier, 2020, doi: 10.1016/j.inffus.2020.06.008.
6. A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 Int. Conf. Wirel. Technol. Embed. Intell. Syst. WITS 2019, pp. 1–5, 2019, doi: 10.1109/WITS.2019.8723839.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details