Big Data Analytics Based Data Driven Public Health Care System For Heart Disease Detection

Dr. F Rahman[1], Lalnunthari[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.*

KEYWORDS

Heart Disease, Artificial Intelligence (AI), Machine Learning

ABSTRACT

Everyone should be more conscious of their lifestyle choices in order to live a healthy life. This can be done by following appropriate eating plans, being aware of healthcare issues, and knowing how to maintain good health. The health sector produces enormous amounts of electronic data every second in the present digital era, making it extremely difficult to store and manage the data using traditional software and technology. Additionally, the impending data is generated in varied forms, such as semi-structured, unstructured, and/or structured. The rate at which data are evolving and their abundance necessitate an overwhelming amount of health care data. Thus, it appears that health-related data can be handled using big data. This study demonstrates how quickly things are evolving, resulting in the creation of several approaches and continuous research into fixes for issues that crop up in a range of industries.

Keywords: heart disease, Artificial Intelligence (AI), Machine Learning.

1. Introduction

The heart is a vital component of the human body that maintains the normal functioning of the entire body. A healthy heart is essential to human existence. In the unlikely event that the heart's capacity is inadequate, it affects many bodily parts [1]. Heart and brain are fundamentally related to each other. The human body continues to function until the heart and brain are operating normally. Heart disease, sometimes known as coronary heart disease or cardiovascular disease [6]. If there is an abnormal blood flow, the brain and heart continue to function. When a blood clot occurs in the brain, it is referred to as a mind stroke or brain stroke; a heart attack occurs when a blood clot occurs in the heart. The heart's and brain's functions depend on one another. Thus, the heart serves a vital function in the human body. There are fresh opportunities to investigate a range of clinical and authoritative questions related to non-stop heart attack. According to a study conducted by the World Health Organisation (WHO), heart disease has killed 12 million people on average and kills one person every 34 seconds [2]. Heart disease continues to be the leading cause of death for women. As people age, the smooth internal dividers of the blood arteries that provide blood to the heart may become thin and damaged due to the buildup of fatty substances known as plaque. When a portion of plaque ruptures, platelets and other blood components attach to the injured area and form blood clots. When blood coagulation completely blocks the blood supply and significantly reduces the amount of blood reaching the heart muscle, a heart attack occurs. The patient experiences pain and discomfort in their chest as a result. As a result, some of the heart muscle starts to suffer damage. Heart muscle is undoubtedly damaged if a blockage in the heart persists for a long period without receiving appropriate therapy. If the blood flow isn't restored quickly, the damage to the heart muscle gets worse over time. Sometimes referred to as a myocardial localised

necrosis (MI), severe myocardial dead tissue, coronary artery blockage, or coronary thrombosis, a heart attack occurs [11].

The health of people is a huge worry as the world grows smaller because of factors including pollution, mental stress, abrupt habit changes, poor food quality, and other addictions. These elements have a significant impact on the heart, a vital component of the human body. Heart disease is a common condition that affects a large number of people worldwide and can be fatal. Therefore, big data from the healthcare sector in the field of heart illness is used to identify better solutions for heart-related problems [9]. The research or problem statement is as follows. "In order to produce comprehensive public policy, big data related to global public health will be obtained from open repositories, pre-processed to place in the standard form, and processed using a soft computing approach." Big data offers numerous options for indexing and storing unstructured, previously unutilised healthcare data for later use. [4].

2. Literature Review

Several machine learning methods, including C4.5, Random Forest, Naïve Bayes, Decision Tables, and Logistic Regression, were put into use. The accuracy of prediction using the suggested model is higher than that of other techniques, as figure 4.15 illustrates. Compared to all other algorithms, Algorithm C 4.5 yields incredibly low accuracy [5][12]. Computer and mobile applications have been built for the Windows and Android platforms, respectively, based on the above suggested rule set. The literature research indicates that structured data is utilised in the prediction process, and machine learning methods are employed to verify and enhance the precision of the outcomes [3]. Numerous studies on qualities were conducted, and the necessary attributes to improve the accuracy of the results were discovered. The suggested study focusses on every dataset attribute, extracting prominent attributes using statistical techniques and correlation algorithms. Machine learning techniques were applied to every feature characteristic in a variety of environments, and the effectiveness of each approach was examined [7]. Increasing prediction accuracy is the primary goal of feature extraction. The rule set, prediction system, and application have all been constructed with the aid of the Apriori algorithm's properties. Benefits of health-related Big Data offers us several options, including: preventing disease, identifying risk factors, improving the probability component of an illness, paying for treatment, making decisions, etc. Big data has the potential to prevent disease, which could be beneficial for people with various ailments. Big data research may reveal trends that could aid in the early identification of danger factors. Big data can also be used to develop treatment models that significantly save costs. Big data analysis can also lead to better decision-making [8].

3. Methodology

It is possible to combine data in the form of tables, remarks, records, charts, figures, and messages in order to assess and derive better future planning. As was previously said, big data is available in numerous locations and is dispersed globally in both organised and unstructured formats. Therefore, it is a difficult task to recognise the many sources and gather the necessary data. Clinical data, lab results, clinical references in text form (such as medication information, health goods, measurements, etc.) are among the data used by the health care industry. New methods of data storage, such as electronic health records, telehealth, smart devices, etc., were employed. In a similar vein, the digital world offers data from millions of people in an unstructured format via social media [13].

Social media platforms are nothing more than an arrangement of many entities (individual people) brought together for electronic communication that is connected to mutual acquaintances, families, shared interests, workplaces, and other things. Users can sign up and use any platform by building a profile with their name, city, qualifications, mobile number, location, place of employment, interests, and other details. Social networking platforms come in a variety of packages with various practices and amenities. Here are a few social media sites that are in use right now. These platforms are used by both professionals and non-professionals and offer services in a variety of fields, including social, health, and scientific.

Big data is a source of information that can aid in making important decisions. To support these decisions, the information must be handled, organised, analysed, and examined in a variety of formats, including unstructured, semi-structured, and structured, in order to yield patterns that are practical and goal-oriented. The phrase "data analytics" refers mostly to the study of data to detect associations or relations between data and relevant findings using machine learning algorithms, enabling timely decision-making. Analysing large amounts of data to uncover hidden patterns, hidden relationships, and other useful information that may be used to make better decisions is known as data analytics. Data analytics can be carried out using a variety of methodologies and readily available tools, producing definitive results.

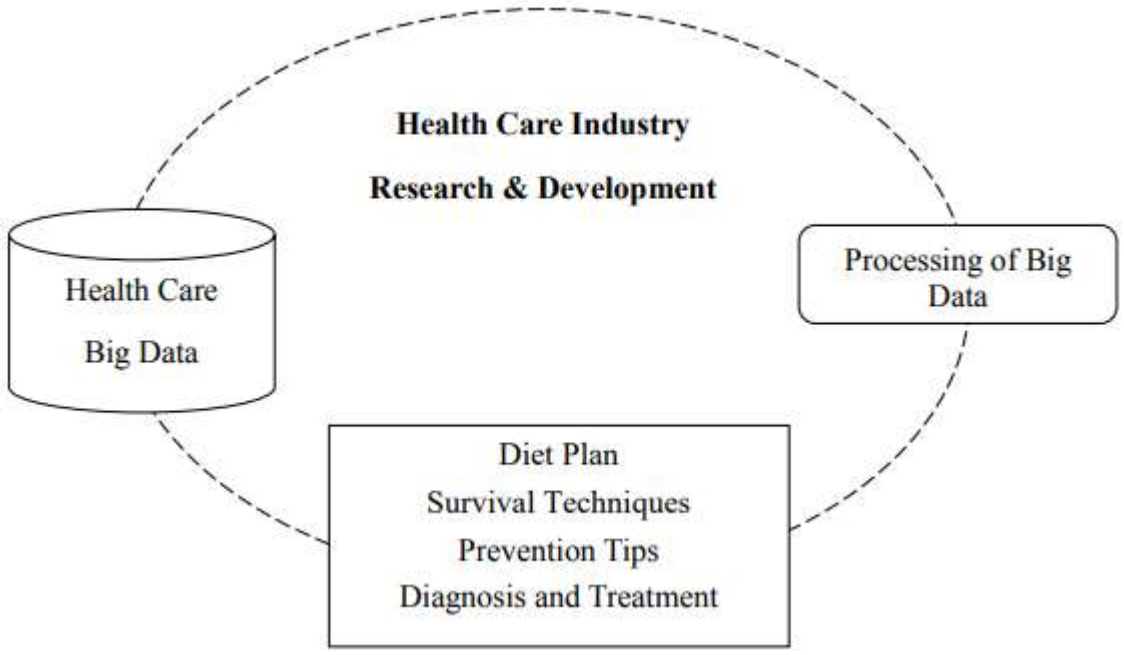

Figure 1: overall proposed framework:

4. Results and discussion

Machine learning techniques are used for the forecast. The accuracy of the result may be improved by the feature selection of the attributes. Pre-processed data from the UCI dataset, which contained missing values and duplicates eliminated, were used in the experiment. The machine learning techniques are applied in several contexts, including Python and Weka. Linear regression and Random Forest are implemented in Python. Weka uses methods for classification such as C 4.5, Decision Tables, and Naive Bayes [10].

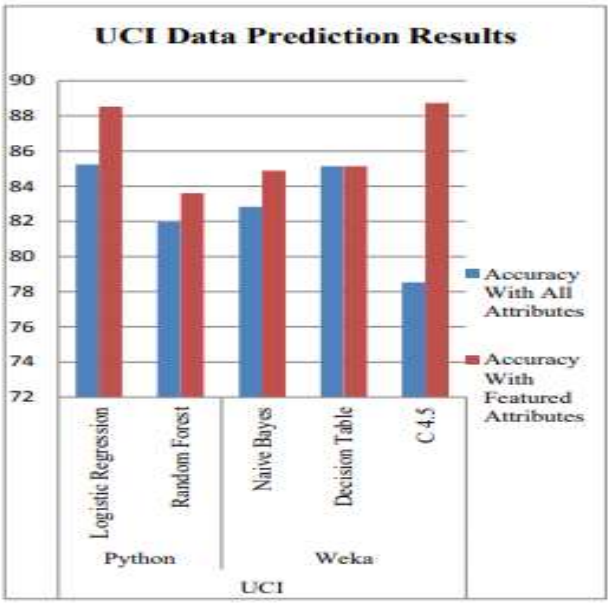

Figure 2: UCI dataset

Figure 3's results demonstrate how highlighted features increase prediction accuracy more quickly. To increase prediction accuracy, the Apriori algorithm's rules are created and updated in collaboration with subject matter experts. A system has been created for Windows and mobile applications using a planned rule set.

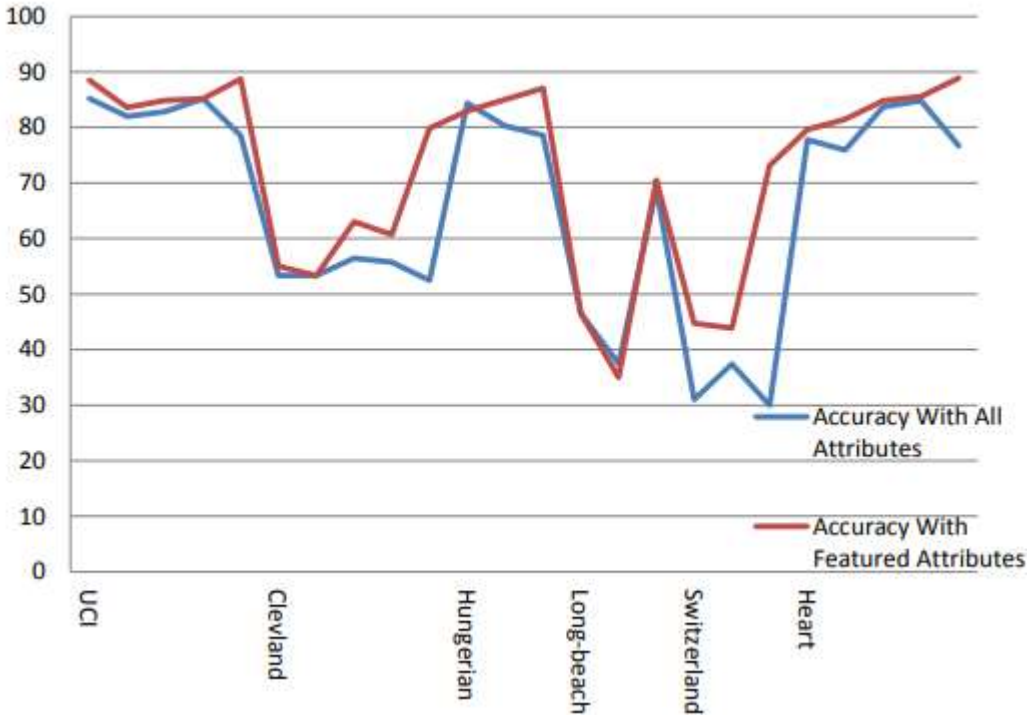

Figure 3: Comparative analysis of Prediction results of UCI with all other datasets

Figure 3 above displays the prediction results for the UCI dataset. Several machine learning methods, including C4.5, Random Forest, Naïve Bayes, Decision Tables, and Logistic Regression, were put into use. The accuracy of prediction using the suggested model is higher than that of other techniques, as figure 4.15 illustrates. In comparison to all other algorithms, Algorithm C 4.5 yields extremely low accuracy. Computer and mobile applications have been built for the Windows and Android platforms, respectively, based on the above suggested rule set.

Table 1: detection ratio for DCNN

Diseases	Models	Accuracy (%)	Specificity (%)	Sensitivity (%)	F1-Score	Precision (%)	Recall (%)
Heart disease	*Logistic regression*	80	82	81	0.75	80	79
	Random forest	80	94.16	91.15	0.89	92	80.85
	Proposed	89	92.05	93	0.90	92.15	95.05

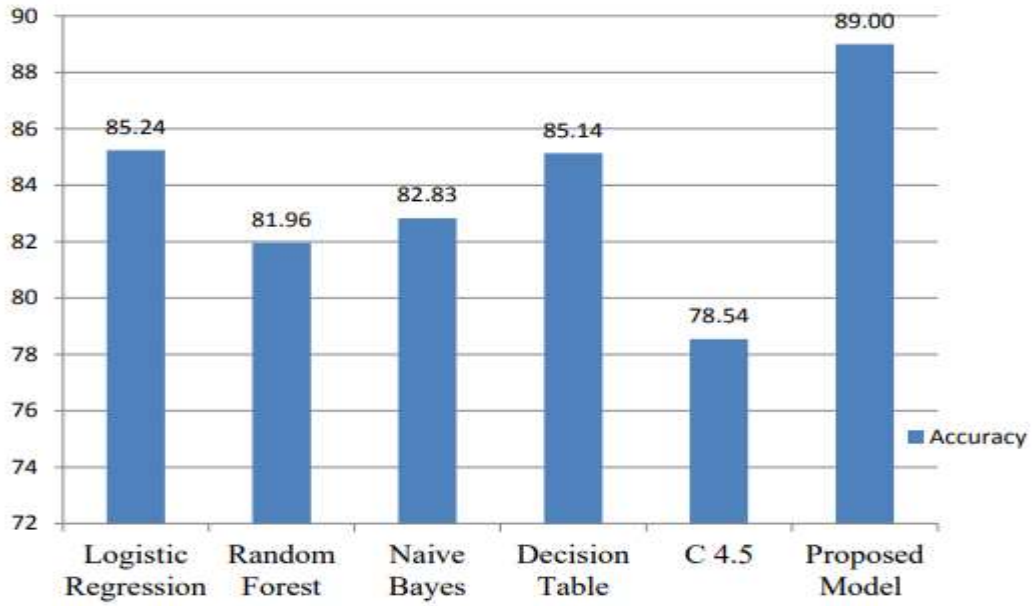

Figure 4: Prediction accuracy

Using Python and the Weka tool, various feature selection techniques were performed to the dataset in order to extract features for every characteristic. The outcomes of the application of filter method feature selection methods, such as Univariate selection, Feature Importance, and Correlation matrix with heat map, are stored. A selection of features were employed for the prediction using the weighted average approach. When the suggested rule set is used, the accuracy of the diagnosis of heart disease can be improved to 89%. Therefore, this collection of rules can be applied to enhance the precision of diagnosis and prediction.

5. Conclusion and future scope

The conclusion presents research supporting the theory that "Big data cannot just help but essentially delivers actual big decisions after processing available data thoroughly." A well-thought-out big data processing plan combined with thorough analysis of the results yields intentional outcomes such as proposals and policies related to public health, predictions, etc. searching for and properly organising big data about healthcare on the internet, social media, pollution statistics, etc. retrieving the necessary information about chronic cardiac disease. Big data is an enormous collection of connected data that, when properly processed and using the right approaches, can yield useful findings. Large amounts of organised and unstructured big data are available in the healthcare industry. Heart disease is a serious health concern that requires accurate data collecting from multiple source.

Reference

[1] Sorte, P., Golande, A., Yermalkar, A., Suryawanshi, V., Wanjare, U., Satpute, S.: Smart hospital for heart disease prediction

using IoT. Int. J. Innov. Technol. Explor. Eng. (IJITEE) 8(9), 321–326 (2019)

[2] S.Nandhini, Monojit Debnath, Anurag Sharma, Pushkar, Heart Disease Prediction using Machine Learning: International Journal of Recent Engineering Research and Development.

[3] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. Indian Journal of Information Sources and Services, 14(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[4] Ganesan, M., Sivakumar, N.: IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In: IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–5. IEEE, Pondicherry, March 2019

[5] Lade, Swati, Sushil Burle, Satish B Kosalge, and Madhuri N Bansode. 2022. ANTIMICROBIAL AND ANTIOXIDANT ACTIVITY OF HIBISCUS SABDARIFFA. LINN (ROSELLE). *International Journal of Pharmacy Research & Technology*, 12 (1), 22-27. doi:10.31838/ijprt/12.01.04

[6] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna, "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Engineering & Technology, 7 (2.32) (2018) 363-366.

[7] Sumiati., Penny, H., Agung, T., Afrasim, Y., & Achmad, S. (2024). Classification of Heart Disorders Using an Artificial Neural Network Approach. Journal of Internet Services and Information Security, 14(2), 189-201.

[8] Sunil S. Khatal and Dr. Yogesh Kumar Sharma, "Analyzing the role of Heart Disease prediction system using IoT and Machine Learning", International Journal of Advanced Science and Technology, Vol. 29, No. 9s, (2020), pp. 2340-2346.

[9] Archna Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 International Conference on Electrical and Electronics Engineering (ICE3- 2020).

[10] Sathyanarayanan, S., & Srikanta, M.K. (2024). Heart Sound Analysis Using SAINet Incorporating CNN and Transfer Learning for Detecting Heart Diseases. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 152-169. https://doi.org/10.58346/JOWUA.2024.I2.011

[11] Fizhar Ahmed, An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients: International Journal of Computer Applications (0975 – 8887) Volume 164 – No 6, April 2017

[12] Thai, D.T., Minh, Q.T., Phung, P.H.: Toward an IoT-based expert system for heart disease diagnosis. In: The 28th Modern Artificial Intelligence and Cognitive Science Conference (MAICS 2017), pp. 157–164. IPFW, Fort Wayne (2017)

[13] M. Ganesan and Dr. N. Shivakumar, "IoT based heart disease prediction and Diagnosis model for healthcare using Machine Learning models", Proceeding of International Conference on System Computation Automation and Networking 2019.

[14] Chitambaram T, Logesh Kannan N, Gowsalya M, "Heart Disease Detection Using Machine Learning", Advances in Computational Science and Technology ISSN 0973- 6107, Volume 10.