Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Comparison of Machine Learning Models to Predict Heart Attack: A Review

Vivek Rathee[1], Renu Vadhera[2], Parveen Kumari[3]
M. Tech Scholar[1] – DPGITM Engineering College, Department of CSE & IT, Gurugram, Haryana, India
Assistant Professor[2,3]– DPGITM Engineering College, Department of CSE & IT, Gurugram, Haryana, India
vivekrathee@gmail.com[1], renu.csed@dpgitm.ac.in[2], parveen.csed@dpgitm.ac.in[3]

Article Info
Page Number: 1265-1273
Publication Issue:
Vol. 72 No. 1 (2023)

Article History
Article Received: 15 February 2023
Revised: 24 March 2023
Accepted: 18 April 2023

Abstract
According to the latest report published by the World Health Organization (WHO), heart disease is becoming increasingly prevalent globally due to various reasons associated with modern lifestyle. The fatality rate due to cardiac arrest is over 18 million per year worldwide. Unfortunately, due to the large population and insufficient healthcare infrastructure, it is often not feasible to identify heart disease in its early stages and initiate treatment. However, with the emergence of AI, DL, and Soft Computing, it has become possible to investigate these health issues in the initial stage. Therefore, the primary objective of this research is to develop a solution to predict heart disease accurately and in advance. Machine learning has become a critical component in healthcare, and many researchers have published papers investigating appropriate algorithms to predict heart disease. After reviewing various research papers, it was observed that different algorithms yield different accuracies on the same or different datasets. In the research work, various machine learning approaches and analysis will be examined for better accuracy on a validated dataset.
Keywords: Random Forest, Decision Tree, SVM, Machine Learning, Heart Disease, KNN, Logistic Regression

1. Introduction

Cardiovascular diseases are a leading cause of death worldwide, according to a report by the World Health Organization (WHO). More people die from cardiovascular diseases than from any other illness. In 2016 alone, almost 18 million people died from these diseases, accounting for almost 30% of all global deaths. Out of these 30%, 84% of the deaths were due to stroke and heart attack. The high mortality rate and expensive surgical procedures associated with heart disease make it a serious threat to many families, particularly those in developing countries [6-7]. It is essential to explore the relationship between various human features and the likelihood of developing heart disease [8-9]. Therefore, a robust system is necessary to predict which individuals are at higher risk of heart disease, so that preventive measures can be taken well in advance. Machine learning, which is based on statistics, aims to use optimization techniques to deliver accurate results in various domains, such as industry, medicine, social sciences, business, and more [10-12]. The development of modern electronic health technologies has enabled the generation of vast amounts of data in our daily lives [9-

10]. With every online transaction, 2.5 quintillion bytes of data are generated per day, which can be stored online for future analysis. Figure 1 illustrates the prevalence of cardiovascular diseases in advanced countries.

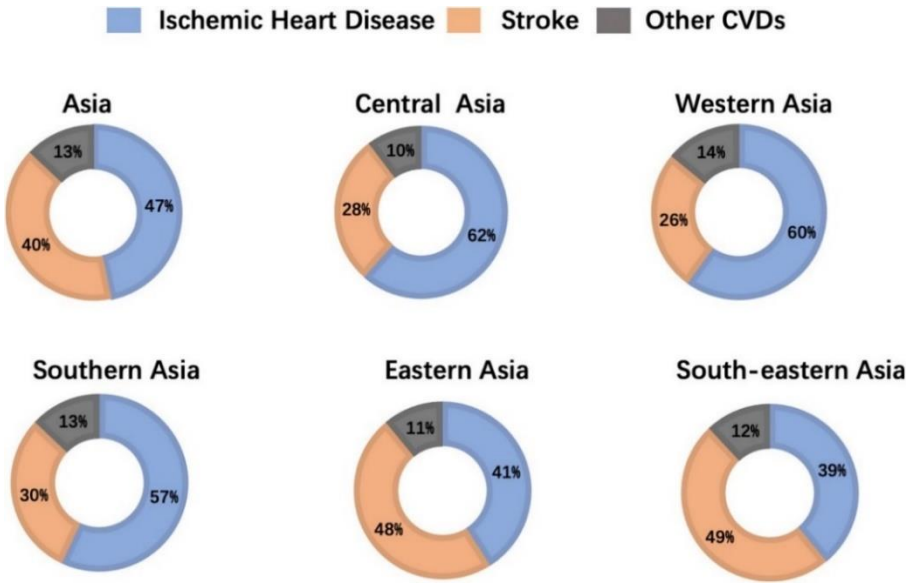

Fig.1: Proportions of IHD and Stroke Deaths in Total CVD Deaths in AISA [JACC].

To extract knowledge from large and complex datasets, models for data extraction and classification are built using machine learning and statistical approaches [13-14]. Clinical decision making is crucial to all patient care activities, including diagnosis, treatment, and disease management, and is a critical skill for medical professionals.

Ensuring that medical treatments are affordable is a crucial factor that affects the selection of medical strategies to apply to patients. In many cases, doctors have a limited amount of time to make decisions about the most effective treatment options for their patients [15]. This is especially true in today's healthcare environment, where patient visits are often shorter than they used to be. Figure 2 provides an overview of death rates caused by various illnesses in developed countries.

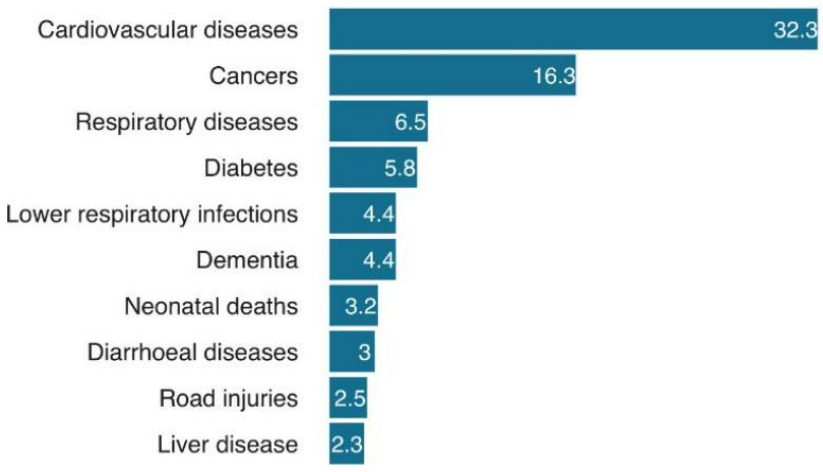

Fig.2: Death status caused by various illness in developed countries [IHME].

2. Types of Cardiac Disease

Types of Cardiac Disease: Heart disease is a major health concern in modern times and can manifest in various stages. These types of heart disease can have different attributes that can disrupt the proper functioning of the heart [20]. Based on clinical evidence, heart disease is mainly classified into various categories. According to medical experts, there are different types of heart diseases as shown in Figure 3.

- Cardiomyopathy
- Heart Failure
- Angina Pectoris
- Myocardial infarction
- Atrial Fibrillation
- Heart arrhythmia

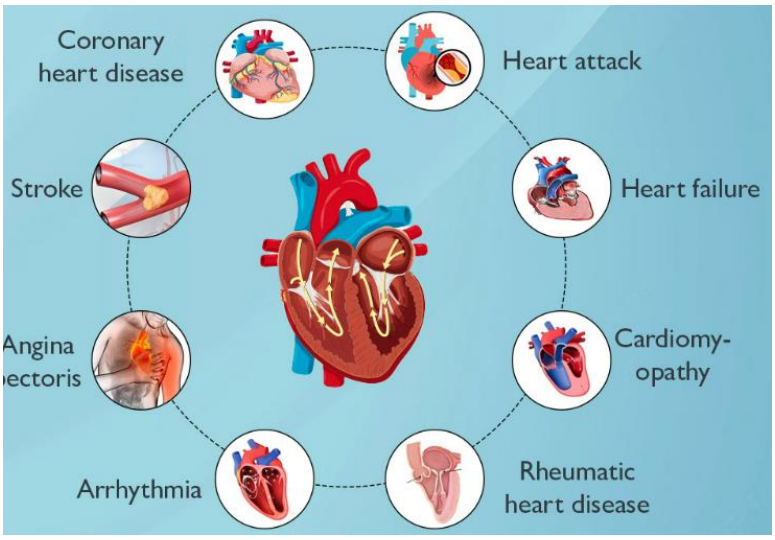

Fig.3: Types of Cardiac Disease [University Diagnostic Medical Imaging].

Coronary Artery Disease: Proper blood circulation is vital for the healthy functioning of the human body. However, when the blood flow is reduced, it can cause discomfort in the coronary artery, leading to coronary artery disease [21]. This condition can cause damage to the veins and affect the normal functioning of the heart, leading to restless and irregular changes in the systolic and diastolic parameters of the heart.

Acute myocardial infarction: Acute myocardial infarction, also known as a heart attack, is caused by the presence of fatty material that affects the proper circulation of blood in the body and heart [22]. As a result, the arteries can become blocked and damaged, leading to tissue damage and a severe condition in which oxygen-rich blood supply is unable to reach various organs, resulting in abnormal functions.

3. Literature Review

H. Jindal et al: The primary focus of this research paper is to predict the likelihood of patients developing heart disease based on various medical traits. The researchers have developed a heart disease prediction system that utilizes the medical history of patients to identify which individuals are at a higher risk of being diagnosed with a heart ailment. They employed different machine learning (ML) algorithms such as logistic regression and KNN to classify and predict patients with heart diseases. The researchers have used a collaborative approach to improve the accuracy of the prediction model. The proposed model was able to predict the occurrence of heart diseases in specific individuals with high accuracy using KNN and Logistic Regression. The results of this study showed that the accuracy of the model was better than previously used classifiers such as naive Bayes [1].

P. Motarwar et al: The medical field is faced with numerous challenging tasks, such as predicting patterns to prevent and manage diseases. To accomplish this task, supervised machine learning algorithms such as LMT, SVM, and Naive Bayes are utilized. In order to build a model, a dataset is necessary, which is then divided into training and testing datasets. Next, the dataset is preprocessed to extract important features while disregarding non-important features to enhance accuracy. Among the implemented algorithms, the random forest algorithm has been found to provide the best accuracy [2].

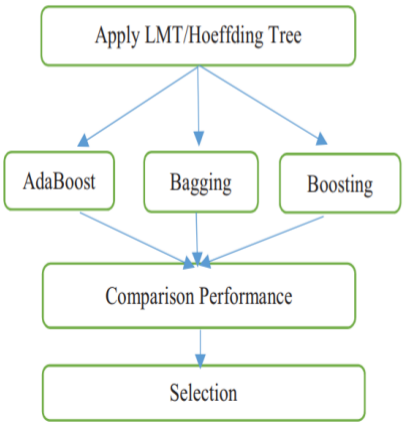

Fig.4: Steps for heart diseases prediction [2]

V. Sharma et al: The author of this paper has utilized a widely recognized UCI Heart disease dataset that comprises 14 distinct factors associated with heart diseases for their research work. In developing a model, ML algorithms such as Random Forest, Support Vector Machine (SVM), Naive Bayes, and Decision Tree were employed. The author also attempted to identify correlations between various parameters present in the dataset using standard ML techniques, and based on these parameters, the likelihood of cardiac diseases can be determined efficiently. According to the results obtained, the Random Forest algorithm is the most accurate and provides rapid prediction results, which can be beneficial to medical practitioners in their clinics as decision-support systems [3].

S. Mohan et al: This research article proposes a technique with the main aim of identifying significant features from a dataset and applying ML algorithms to achieve better accuracy in predicting cardiovascular diseases. According to various reports, heart disease is one of the

leading causes of mortality worldwide. In the medical healthcare system, predicting cardiovascular diseases is a daunting task in the field of medical data analysis.

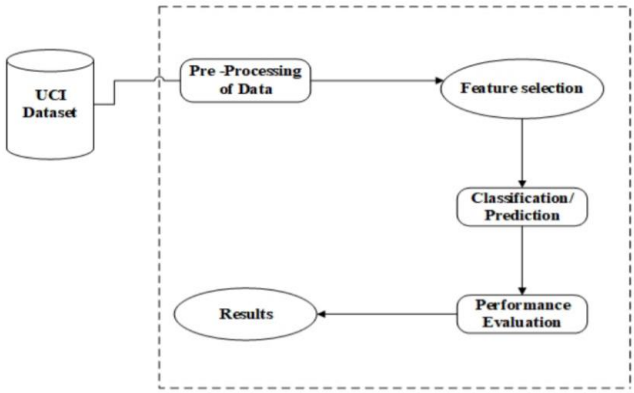

Fig.5: Experiment workflow with UCI dataset [5].

The large amount of data generated by the healthcare industry on a daily basis makes manual analysis impractical, but AI is playing a crucial role in various domains, particularly in healthcare. With the assistance of ML, data scientists can accomplish this difficult task with ease. The proposed system utilizes an integrated technique, and the Random Forest algorithm yields better accuracy in predicting heart attacks [5].

A. Nikam et al: This research proposes an ML technique that employs various characteristics to predict cardiac diseases, with BMI being the primary method used for estimation. BMI is a crucial measure for predicting cardiovascular diseases. The article mainly concentrates on the impact of BMI on the prediction of cardiovascular diseases. The model has been designed with several different features as well as regression and classification methods. The author concludes that BMI is a significant factor in the prediction of cardiovascular diseases [4].

4. Machine Learning

The emergence of modern technologies in electronic healthcare has enabled both humans and machines to generate vast amounts of data on a daily basis. With each online transaction, data is generated at a rate of 2.5 quintillion bytes per day, which can be stored online for future analysis [1-4]. To extract knowledge from these large and complex datasets, models are built using machine learning and statistical approaches for data extraction and classification.

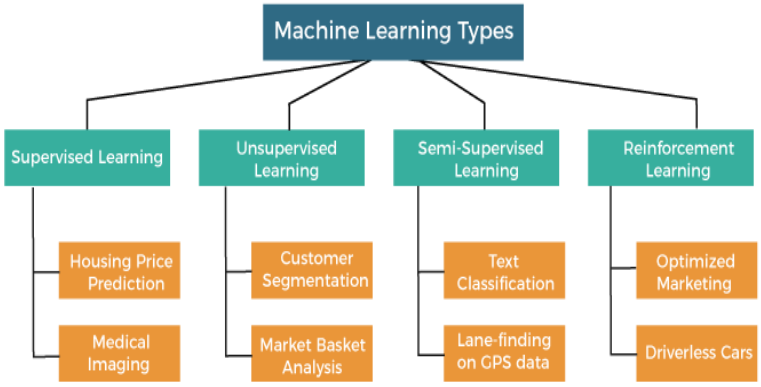

Fig.6: Machine Learning Classification.

Preserving medical statistics applications is amazingly diverse, and at a advanced level, one can inspect a patient's delicate data such as blood pressure, sugar level, temperature and glucose for recognizing and preventing illness in the early stages. At a lower level, a single feature such as blood pressure level can be analyzed to provide proper medication [6]. Various methods and techniques are utilized in all aspects to obtain meaningful knowledge from raw data prevailing in the actual ecosystem. In this article, a combinational approach of traditional statistical methods with model-based machine learning techniques will be used to efficiently classify heart disease from datasets with reduced attributes and highly improved accuracy measures.

Supervised Techniques

These types of techniques construct the training model using known evidence with minimal uncertainty. Figure 7 displays the classification of supervised learning. The prediction models are developed using supervised learning with classification and regression techniques [8]. Some of the common classification algorithms include SVM, Naive Bayes classification techniques, NN, LR (Logistic Regression), KNN, decision trees, and discriminant analysis.

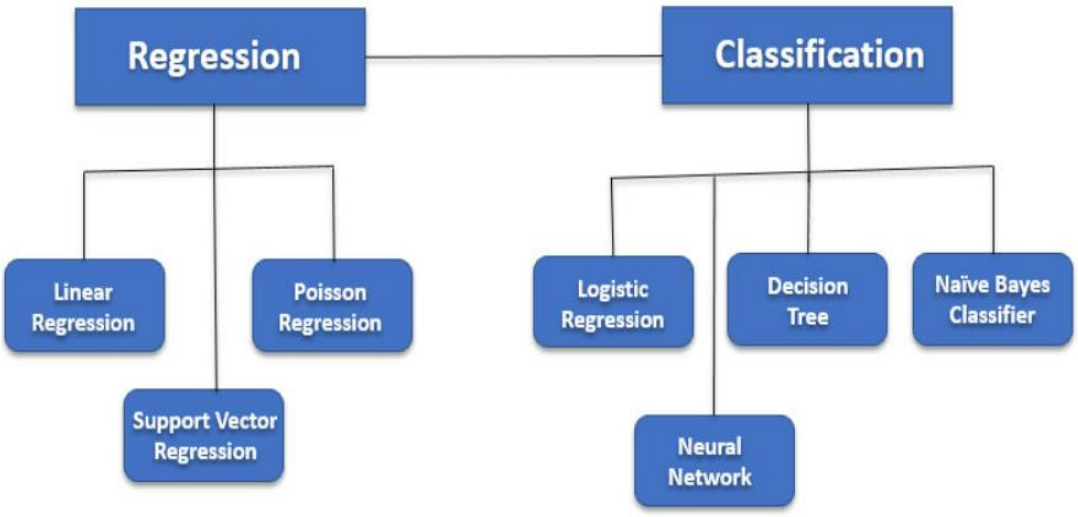

Fig.7: Supervised Learning Techniques.

Unsupervised Techniques

It is not always necessary to have labeled datasets to predict critical information. Unsupervised machine learning can successfully predict outcomes for unlabelled datasets. Clustering is the most commonly used method of unsupervised learning techniques. Hidden Markov models, Fuzzy clustering, hierarchical clustering, K-means clustering algorithm, and many others are the primary clustering algorithms [10]. Researchers have developed numerous algorithms that fall under unsupervised and supervised categories, but their application varies significantly from one scenario to another. Therefore, selecting an appropriate machine learning algorithm yields better results in prediction and classification processes. However, choosing the right algorithm can be a challenging task. Figure 8 displays the classification of unsupervised learning techniques.

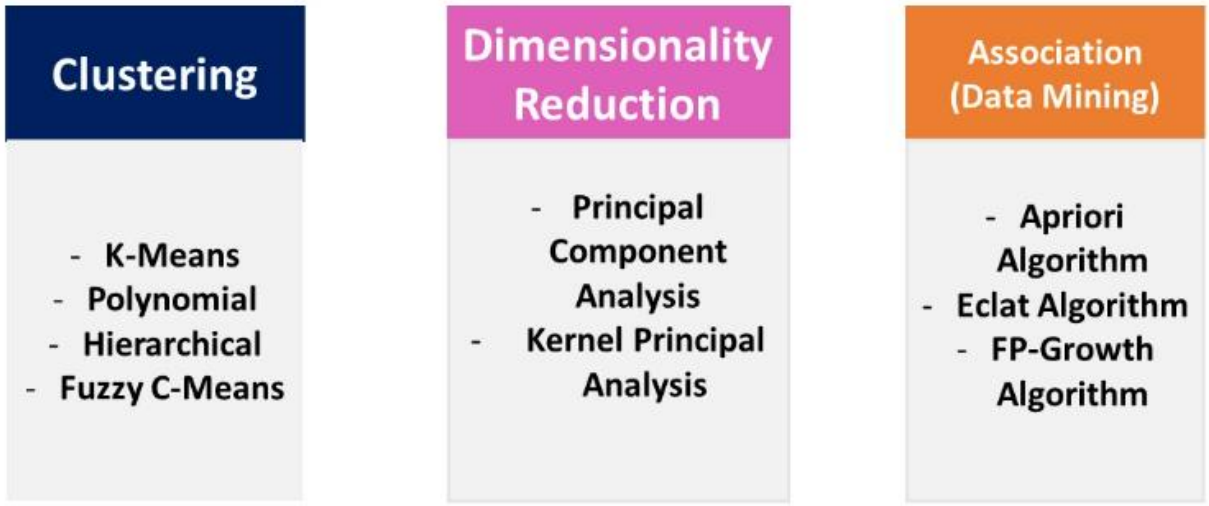

Fig.8: Un-supervised Learning Techniques.

5. Conclusion

The main objective of this review paper is to analyze various machine learning algorithms that aid in the effective prediction of heart disease. After analyzing numerous research papers, it was discovered that different algorithms predict varying accuracy on the same or different datasets. In the modern lifestyle, healthcare issues are a major concern, and diagnosing diseases in the early stages can reduce the mortality rate. Emerging fields such as Soft Computing, ML, and DL play a crucial role in healthcare by predicting diseases effectively. To achieve better accuracy in heart disease prediction, more attributes can be added to the dataset, along with the integration of machine learning algorithms.

References

[1]. H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in ICCRDA 2020, IOP Conf. Series: Materials Science and Engineering, vol. 1022, IOP Publishing, 2021, p. 012072. doi: 10.1088/1757-899X/1022/1/012072.

[2]. P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning," in 2020 International Conference on

Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE, 2020. doi: 10.1109/ic-ETITE47903.2020.242.

[3]. V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, Dec. 18-19, 2020. doi: 10.1109/ICACCCN51052.2020.9362842.

[4]. A. Nikam, S. Bhandari, A. Mhaske, and S. Mantri, "Cardiovascular Disease Prediction Using Machine Learning Models," in 2020 IEEE Pune Section International Conference (PuneCon), IEEE, Dec. 16-18, 2020. doi: 10.1109/PuneCon50868.2020.9362367.

[5]. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, Special Section on Smart Caching, Communications, Computing And Cybersecurity For Information-Centric Internet Of Things. doi: 10.1109/ACCESS.2019.2923707.

[6]. D. K. G, S. K. D, A. K, and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," in Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India.

[7]. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," in Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), IEEE Conference, IEEE Xplore ISBN:978-1-5386-0965-1.

[8]. S. Zhang and Y.-L. S. A., "Deep learning-based recommender system: a survey and new perspectives," Journal of ACM Computing Surveys, vol. 1, no. 1, pp. 1–35, 2017.

[9]. A. Khatami, A. Khosravi, and C. L., "Medical image analysis using wavelet transform and deep belief networks," Journal of Expert Systems with Applications, vol. 3, no. 4, pp. 190–198, 2017.

[10]. A. Shetty and C. Naik, "Different data mining approaches for predicting heart disease," International journal of innovative research in science, engineering and technology, vol. 3, no. 2, pp. 277–281, 2016.

[11] S. Aydin, "Comparison and evaluation data mining techniques in the diagnosis of heart disease," Indian Journal of Science and Technology, vol. 6, no. 1, pp. 420-423, 2016.

[12] N. Bayasi and Tekeste, "Low-power ECG-based processor for predicting ventricular arrhythmia," IEEE Transactions on Very Large-Scale Integration (VLSI) Systems, vol. 24, no. 5, pp. 1962-1974, May 2016.

[13] B. Berikol and Yildiz, "Diagnosis of acute coronary syndrome with a support vector machine," Journal of Medical System, vol. 40, no. 4, pp. 11-18, 2016.

[14] Z. Wang, X. Liu, and J. G., "Identification of metabolic biomarkers in patients with type-2 diabetic coronary heart diseases based on metabolomic approach," Journal of Cardiovascular Diseases & Diagnosis, vol. 6, no. 30, pp. 435-439, 2016.

[15] M. Singh and Martins, "Building a cardiovascular disease predictive model using structural equation model and fuzzy cognitive map," Journal of Fuzzy Systems, vol. 2, no. 6, pp. 1377-1382, 2016.

[16] S. Prabhavathi, "Analysis and prediction of various heart diseases using DNFS techniques," International Journal of Innovations in Scientific and Engineering Research, vol. 2, no. 7, pp. 678-684, 2016.

[17] R. Sali and M. Shavandi, "A clinical decision support system based on support vector machine and binary particle swarm optimisation for Cardiovascular disease diagnosis," International Journal of Data mining and Bio-informatics, vol. 15, no. 1, pp. 312-327, 2016.

[18] P. K. Ghadge, "Intelligent heart attack prediction system using big data," International Journal of Recent Research in Mathematics, Computer Science and Information Technology, vol. 2, no. 2, pp. 73-77, 2016.

[19] G. Purusothaman and K. Krishnakumari, "A survey of data mining techniques on risk prediction: heart disease," Indian Journal of Science and Technology, vol. 8, no. 5, pp. 643-651, 2015.

[20] A. Richter, J. Listing, M. Schneider, T. Klopsch, A. Kapelle, J. Kaufmann, A. Zink, and A. Strangfeld, "Impact of treatment with biologic dmards on the risk of sepsis or mortality after serious infection in patients with rheumatoid arthritis," Annals of the Rheumatic Diseases, pp. 147-153, 2015.

[21] S. Sairabi and D. Mujawar, "Prediction of Heart Disease using Modified K-means and by using Naive Bayes," International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, 2015.

[22] M. Vafaie and M. Ataei, "Heart diseases prediction based on ECG signals classification using a genetic-fuzzy system," Journal of biomedical signal processing and control, vol. 14, no. 5, pp. 291–296, 2014.

[23] Z. Wang, "Study on qi-deficiency syndrome identification modes of Coronary heart disease based on metabolomic biomarkers," Journal of evidence-based complementary and alternative medicine, vol. 24, no. 16, pp. 192–198, 2014.

[24] X. Yang, M. Li, Y. Zhang, and J. Ning, "Cost-sensitive naive bayes classification of uncertain data," Journal of Scientific World, vol. 9, no. 8, pp. 1897–1904, 2014.

[25] D. Chandna, "Diagnosis of heart disease using data mining algorithm," International journal of computer science and information technologies, vol. 5, no. 2, pp. 1678–1680, 2014.