Research Article

Don Bosco Institute of Technology Delhi Journal of Research
Year 2024, Volume-1, Issue-2 (July - Dec)

# Study on Data Mining Techniques in Healthcare Sector: A Review

Charanpreet Kaur[1] , Rosy Madaan[2]

[1]Research Scholar Department of Computer Science & Engineering FET, MRIIRS

[2]Associate Professor Department of Computer Science & Engineering FET, MRIIRS

**Keywords:** *Data Mining, Healthcare, Heart Disease, Classification, Clustering*

**doi:** 10.48165/ dbitdjr.2024.1.02.05

## ABSTRACT

Today, the Healthcare sector is generating bulks of data be it from the medical history of the patients to their personal details, their clinical data, or the genetic data. Electronic Health Records (EHR), the medical data, is very complex and varied and hence cannot be processed using the traditional manual tools. Hence, Data Mining Analysis is used extensively in the Healthcare Industry to uncover the hidden patterns and relationships to study the similarity between patients, identify their symptoms and diagnose the disease at an early stage so that proper treatment could be given to the patients well in time. Today, Heart Diseases are very common and can lead to the risk of life. Due to the lack of extensive medical facilities and resources in the healthcare sector, it is very important to diagnose the risk factors leading to cardiovascular disease. This paper reviews and compares the work done by the different researchers and highlights the various risk factors that can cause heart problems and apply the different data mining algorithms that can be used to diagnose the early symptoms so that the disease can be cured.

## Introduction

In the back 1990s, data mining was being introduced and its roots are being traced along with statistical tests, regression analysis, database theory, and machine learning. The analysis of data gained importance when computers changed into decision support systems. With the growth of data warehouses and data mining concepts, Data Analytics has evolved with time. It includes the discovery and interpretation of hidden patterns and relationships. Earlier statistics were used for the result interpretation of data and discovering new patterns and associations. Ancient Egyptians even used the statistics concepts for building pyramids. The Government of different countries has been using statistics to estimate the population growth and plan the various activities for their city. Statisticians and data analysts began to use the data mining term by the late '80s.

Business Analysts were keen on getting all the information about their data as the businesses and the markets expanded and grew more complex. The decision support systems (DSS) provided the basic information to run the day-to-day business activities but what the top management executives

*Corresponding author.
*E-mail address:* charanpreet.dbit@gmail.com

required was the statistical information to take strategic decisions [19]. Data warehouses were then introduced to store the huge volumes of data and data mining is the concept used to extract the unknown and hidden information from these databases. "*Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses*" [1]. Data mining tools are used to predict future trends and analyse the data very efficiently. Data mining is an interdisciplinary field and is considered a misnomer. Data mining is the technology that analyses past data with the present to predict the future. There are tons of data i.e. generated from the various healthcare organizations. It includes patients history, clinical data, treatment details, hospital records, etc. which needs to be analysed to improve the healthcare.

## Data Mining in Healthcare

In the healthcare sector, patient data has been increasing exponentially today that has led to the rise of Electronic health records (EHR). With data mining, healthcare organizations have optimized treatments efficiency, increased patient's life; diagnose the disease at an early stage, customer relationship management, etc. The amount of data generated in the medical field contains valuable information which can be used for analytical processing. There are many challenges faced while dealing with such a large amount of data [8]. Data mining and predictive analysis have opened new access for the diagnosis and treatment of diseases at an early stage in the Healthcare industry. Medical data mining helps in evaluating and predicting early symptoms to treat any disease or condition. This reduces the money spend and saves the lives of the patients. Effective use of data mining algorithms in the healthcare industry can further reduce hospital admissions and claim to aid better healthcare facilities [4]. Data mining concepts can be used to detect fraud and abuse in case of unusual or abnormal patterns or fraudulent medical policies or insurance. Several data mining techniques like classification, clustering, regression, K- Nearest Neighbour (KNN), Decision Tree, Naive Bayes, and Artificial Neural Network (ANN), etc. are the common techniques that can be used in the healthcare domain. In this paper, different risk factors that can lead to heart disease were analysed using various data mining techniques and then their accuracy rates were compared.

If the problem of heart disease is not identified in time, it can lead to the death of the patient. In most of the cases, the problem is identified at the last stage of the disease or after death. Also, our health resources are very limited be in terms of hospitals or treatment. The collection of the medical data and the survey is also the task as we need to be sure regarding the target population as it is the matter of life and death. Hence, early diagnosis of the disease can save the life of the patient and can reduce the cost and hospitalization.

## Literature Review

The authors Alex P, Mamatha and Shaji, Shaicy [2] studied the different data mining algorithms used to diagnose the heart disease at an early stage. The data was collected from Jubilee Mission Medical College and Research Institute Thrissur" and also from the discharge summary of the respective patients. A total of 2200 samples was collected with 20 attributes. The different attributes taken were:- Age, Sex, Blood pressure(DI), Sugar, Urea, Creatin, Bilirubin, SGOT (Serum glutamic oxaloacetic transaminase), SGPT (Pyruvic), PR (Per Rectum), Sodium, Potassium, HB(Haemoglobin), Family history, Stress level, Sedentary lifestyle, Height, Weight, Smoker(yes/no). They predicted that the ANN algorithm gave the best accuracy rate and it can be further increased using more no. of attributes. The accuracy estimated for the Support Vector Machine (SVM) was 85.88%, for Random forest was 85.88%. It was observed that for K- Nearest Neighbour (KNN), the accuracy rate was 83.21% and for Artificial Neural Network (ANN), it was 92.21%.

J. Thomas and R. T. Princy [6] highlighted the various data mining algorithms that can be used to predict the early diagnosis of the heart problems. The different attributes taken into consideration were age, gender, blood pressure, cholesterol, pulse rate. Different classification techniques like Naïve Bayes, K- Nearest Neighbour, Decision Tree Algorithm, and Neural Network were used to predict the heart-related issues. When more no. of attributes was used, the accuracy level was very high. Taking the above attributes, the authors predicted the accuracy rate of 40.3% and it was further observed that that accuracy can be increased up to 80.6% when smoking and previous medical history, etc. more attributes were added. The proposed model used K- Nearest Neighbour (KNN) and Iterative Dichotomiser 3 (ID3) algorithms. First, the attributes were classified using KNN. The class obtained was analyzed by comparing with standard values. Then ID3 was used to predict the risk level of the heart problem. After calculating the risk levels, the results were generated and were compared with the pre-medical records of the patients. These records were used by the doctors to predict the further medication. The data was classified on the basis of age attributes using KNN. The risk level was then predicted using ID3 and each class had standard values. If the predicted value was more than the standard value then the patient was at high risk.

Priyanka N.; Pushpa Ravi Kumar [7] used two techniques of Naïve Bayes & Decision tree to predict the heart issues in the patients. The model is build using the .NET framework as the front end and the MSQL database as the backend server. The connection was established between the database server and the frontend. The first step was the data pre processing step

in which missing values were checked and inconsistent data was removed. Then the classification techniques were used to predict the class. These algorithms were implemented using the .Net framework. Different test cases were performed and the results were compared. It showed that out of the two data mining techniques, the Decision Tree showed more accuracy.
Srabanti Maji and Srishti Arora [9] proposed a hybridization technique to predict the heart issues using Decision Tree and Artificial Neural Network (ANN). The accuracy, sensitivity, and specificity of the individual classifier and hybrid technique were analyzed and compared. To predict and analyze heart diseases RIPPER classifier, decision tree, artificial neural networks (ANNs), and support vector machine (SVM) were implemented using WEKA Tool. The hybrid model was built using the decision tree technique combined with ANN and was concluded that this model gave the best performance in terms of accuracy rate and efficiency. The Dataset was collected from the UCI repository. There were a total of 13 attributes consisting of 270 instances in the dataset.
Chauhan, R et al. [13] built the model to diagnose early symptoms of heart issues in patients. The classifier models utilized in this paper were random tree, decision rule and random forest. Further, these techniques were analysed and the classifier model determined the maximum accuracy for the diagnosis of the disease. The tool used for implementing algorithms was rapid miner. The data from the online repository of the University of Lyon was used for the research. A total of 209 records were taken to analyse the data. The different attributes taken into consideration were Age, Chest pain, Blood sugar, Blood pressure in resting phases, Electrocardiographic results. The decision tree algorithm showed an accuracy rate of 75.10%, while the random tree gave an accuracy rate of 69.90%, and the random forest technique proved to be the best out of the three with an accuracy of 75.60%.
Hnin Ei Ei Cho, Nan Yu Hlaing [15] used data mining algorithms to diagnose the heart issues early and with proper medical guidance and medical treatment, the life of the patient can be saved. The results were then compared and evaluated and the best algorithm was chosen for the diagnosis process. Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM) techniques were implemented in Python Language. The highest accuracy of 88.52% was achieved by using the Random Forest with 5-fold cross-validation. Also, the Neural Network with 72.99% accuracy performed better compared to all the other classification algorithms for a larger dataset. The dataset was collected from the Cleveland dataset provided by DHDSP (303 samples) consisted of 14 attributes in total. Out of these, 165 instances (54.46%) were having cardiovascular disease. Another dataset was collected from the National Cardiovascular Disease Surveillance System and Cardio train dataset from Kaggle (700,000 samples) and it contained 11 attributes. Around 34,979 instances i.e. (49.97%) were diagnosed with the disease.
Hazra et al. [16] discussed the various data mining techniques that were implemented to diagnose the early symptoms of heart attack. The authors throw light on the work done in this field and explain the research done in the prediction of heart issues using the various data mining tools. It also explains and defines the scope of various data mining tools that can be used to implement these algorithms in real life. The authors further compare all the techniques and predict which technique is the most efficient and yields better performance for the detection of the heart problems.
B. S. S. Rathnayakc and G. U. Ganegoda [17] analyzed the different set of risk parameters that can be the cause of heart disease. These were Gender, Blood Pressure, Cholesterol Level, Diabetes, Diet, Tobacco intake, stress, Age, etc. There are so many attributes that can cause heart issues, to reduce the dimensionality of the risk factors different methods like CFS Subset, Chi-Squared, Consistency subset, Filtered attribute, Gain Ratio can be used. Accuracy was predicted for different algorithms as the ratio of total number of records being correctly classified by the classifier to that of the total number of records taken. Overall Accuracy was calculated as for Naive Bayes Technique, it was 83%, for Neural Network it was 78%. It was observed that for KNN accuracy was 75%, for Decision Tree, it was 77% and for Logistic Regression, the accuracy of 77% was predicted.

## Heart Failure Diagnosis Using Predictive Modelling

According to World Health Organization (WHO) Statistics, cardiovascular disease has become the major reason of death in recent times. It has been estimated that around 17.9 million people die every year because of the heart issues and stroke problems. The most common risk factors are unhealthy lifestyle and eating habits. However, this can be prevented if the patients are diagnosed with the symptoms at an early stage and if proper treatment is given to patients well in time then their lives can be saved. The diseases can be completely cured if they are diagnosed properly at the first stage. Heart and stroke problems are one of the major concerns by which many people die annually. There can be many heart issues like coronary heart disease, heart attack, blockage in the heart, arrhythmia or heart failure [2].

Heart issues can be easily diagnosed at an early stage using the concept of data mining algorithms. It follows an approach method that can analyse huge volumes of data. The methodology involves the building up of the proposed model architecture. The first step is to identify the target data set and then perform pre-processing of the data i.e. the data is prepared. This involves cleaning the data, checking for the missing values and inconsistencies, and pruning the irrelevant attributes [2]. The data needs to be converted into a standardised form before it can be used for analysis. The data can be collected from any medical or clinical source. The next step is to

identify the risk factors that can lead to heart problems like blood pressure, diabetes, obesity, stress level, etc. These attributes are very important to analyse the risk factors that can predict the heart problem. More the number of attributes better is the accuracy rate in the prediction of heart diseases. It is been observed that often the unhealthy lifestyle of people includes habits like smoking, drinking alcohol, no physical activity or exercising, and eating oily food can lead to heart issues. After this, the predictive model can be built using different data mining algorithms like classification techniques viz. Decision tree, random forest, SVM, K-NN, neural networks and their hybridization [10]. It is been observed in various research papers that there is no single classifier technique that gives the best accuracy rate, in fact, hybrid models i.e. the combination of one or more classification algorithms give better results and accuracy. The classifier model is built initially using the training data set [6, 13]. Supervised and unsupervised learning can be used to determine the output of heart related issues. These data mining techniques can be implemented using tools like WEKA, Python, R language, MATLAB, etc. The next step is to calculate the accuracy of each technique implemented. The accuracy helps to find the best technique that can be used for the diagnosis of the heart problem at an early stage. [2, 5, 17]

# Applications of Data Mining Techniques Used in Heart Attack Prediction

Data mining techniques can be extensively used in the medical field to uncover valuable unknown information from huge amounts of health data. This information can be used to extract hidden patterns and knowledge in order to make the best decision and treat the diseases at an early stage [11].

## Applying Classification Process in Diagnosis of Heart related Issues

Classification is a very popular mining technique that is used today in the analysis of the data. It can be used to diagnose the early symptoms of heart problem by classifying the patient's symptoms with the stored dataset and finding the cause of the disease. Classification is a two step process that involves the Learning or Training phase and Classification phase. In the learning phase, a classification model is constructed by using different algorithms. Classification rules are used to represent the learned model. The accuracy of the classifier model can be calculated using the test data in the classification phase.

## Using Decision Tree to project early symptoms of Heart Disease

Decision tree based on supervised learning technique, is used for both classification and regression. It uses a greedy approach in which the tree is built in a top down manner using the divide and conquer technique. It forms a tree where the dataset used is shown by the internal nodes, classification rules are represented by branches and each leaf node shows the output or the class label. Bandarage Shehani et al. [5] discussed risk factors Gender, Blood Pressure, Cholesterol Level, Diabetes, Diet, Tobacco intake, Stress, Age, etc. The author implemented the algorithm by combining it with other data mining techniques and found that the decision tree gave a 77% accuracy rate whereas authors J. Thomas and R. T. Princy [6] also implemented ID3 with KNN together and predicted the accuracy level was increased up to 80.6%. Priyanka N et al. [7] executed the algorithm using MSQL database as a backend server, front end as .NET framework and predicted that decision tree showed the highest accuracy. The authors Srabanti Maji and Srishti Arora [9] built the Hybrid model of Decision Tree and ANN using WEKA Tool and stated Accuracy (%) - 78.14, Sensitivity (%) – 78, Specificity (%) - 22.9. Ritu Chauhan et al. [13] collected the dataset from the online repository of the University of Lyon. 209 tuples were used to analyze the data using the rapid miner Tool and determined the accuracy of 75.10%.

## Applying Naive Bayes in Prediction of Heart Disease

The Naive Bayes Data Mining Technique is another very important algorithm that uses the concept of Bayes Theorem. It helps in predicting the future class label on the basis of the probability of an object. The name "Naive" is given because of the assumption that the occurrence of any feature is independent of other features. The given dataset can be converted into a frequency table that can be further used to calculate the occurrence of the probability of the features. [11] Bandarage Shehani et al. [5] predicted that Naive Bayes - 83% and determined that the algorithm gives the maximum accuracy. The authors Priyanka N.; Pushpa Ravi Kumar [7] used the dataset from the UCI laboratory. It contained 13 attributes like sex, serum cholesterol level, resting ECG etc., and implemented the algorithm using MSQL database as a backend server, front end as .NET framework.

## Implementing Random Forest Technique in early Diagnosis of Heart Problem

Random Forest is the extension of decision tree algorithm based on supervised learning technique only. It uses ensemble learning in which multiple classifiers can be combined to solve the related issues by improving the performance of

the proposed model. It is similar to decision tree algorithm wherein multiple decision trees are constructed in the training phase and final outcome of the Class is predicted by voting of individual trees. The authors Hnin Ei Ei Cho and Nan Yu Hlaing [15] determined an accuracy of 88.52% using the Random Forest with 5-fold cross-validation. The dataset was collected from Cleveland dataset provided by DHDSP (303 samples) and National Cardiovascular Disease Surveillance System and Cardio train dataset from Kaggle (700,000 samples). Ritu Chauhan et al. [13] implemented random forest using rapid miner which is coded in java programming language and determined the accuracy of 75.60%. The authors Mamatha Alex P and Shaicy P Shaji [2] used the dataset from Jubilee Mission Medical College and Research Institute Thrissur and also from the discharge summary of the respective patients. Total 2200 samples were collected with 20 attributes and analyzed using open CV in the platform of QT creator. The accuracy estimated for Random forest was 85.88%.

## Using K Nearest Neighbour (K-NN) for Predicting Heart Problems

K Nearest Neighbour (K-NN) is the Lazy learner data mining technique by comparing the test data to the nearest training set that is similar to it. For an unknown record, the K-NN model finds those k-nearest tuples that are similar to the given unknown record. The concept of Euclidean Distance is used to calculate the similarity between the given tuples. The authors Mamatha Alex P and Shaicy P Shaji [2] implemented K-NN to identify the heart issues with the data set containing 2200 samples of 20 attributes and determined the accuracy of KNN as 83.21%. Similarly, Bandarage Shehani Sanketha Rathnayakc et al. [5] predicted the diagnosis of a heart problem with an accuracy rate of 75%. J. Thomas and R. T. Princy [6] built the Prediction model using KNN and ID3 together with an accuracy of 80.6% to diagnose the heart disease effectively.

## Applying Artificial Neural Network (ANN) in Prediction of Heart Disease

ANN can be considered as the Computer modelled algorithm and is based on the biological nervous system. It consists of multiple layers neural network architecture with back-propagation forward feed network [10]. There are many interrelated elements called neurons that work together to solve a particular problem. Neural Network behaves like an artificial human brain that works on these neurons, weights and biases. It consists of many hidden layers and the output of one layer determines the input of the other layer. The authors Mamatha Alex P and Shaicy P Shaji [2] implemented the ANN algorithm and found that ANN gave the best accuracy rate of 92.21% and it can be further increased using more no. of attributes. Srabanti Maji and Srishti Arora [9] built the Hybrid model using the decision tree technique combined with ANN and concluded that the hybrid model showed the best accuracy rate with data that was used from the UCI repository having 13 attributes and 270 instances. Also, the authors Hnin Ei Ei Cho and Nan Yu Hlaing [15] used data mining algorithms to diagnose the heart issues and determined that Neural Network with 72.99% accuracy was preferable amongst other classification techniques for the larger dataset.

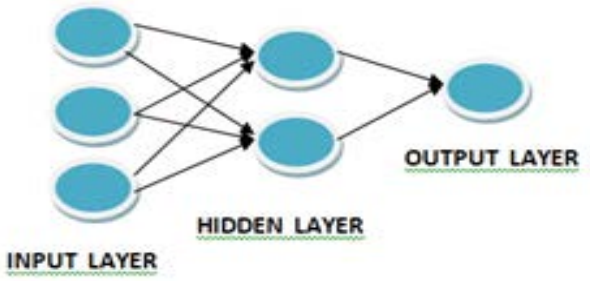


**Fig. 1.** Forward Feed Neural Network

## Implementing Support Vector Machine (SVM) in Diagnosis of Heart related Issues

SVM is the algorithm that creates a decision boundary between the different classes and this boundary is the hyper plane. Distance between classes is called margin and max the margin, more is the accuracy. Various kernel functions like polynomial, Gaussian; sigmoid, etc. can be used to map the training tuples non-linearly [10]. Amani Yahyaoui et al. [3] used the Support Vector Machine algorithm to determine the heart problem at an early stage with an accuracy rate of 65.38%. Similarly, the authors Mamatha Alex P and Shaicy P Shaji [2] determined 85.88% accuracy for SVM on the dataset of 2200 samples of 20 attributes. Hnin Ei Ei Cho and Nan Yu Hlaing [15] predicted 65% accuracy for SVM by implementing the algorithm in the Python programming language.

# RESULT ANALYSIS

Different data mining techniques have been used by the different authors in their research work for the prediction of the heart attack and then their accuracy is calculated using various tools [10]. These data mining algorithms like SVM, random forest, decision tree, K-NN, Neural networks, Naive Bayes were implemented on different datasets using various tools like WEKA, Python, R Interface, MATLAB, etc. It was

determined that Artificial Neural Network (ANN) showed the best accuracy rate and SVM gave the lowest accuracy rate amongst the different data mining algorithms that were implemented. It was further observed that the single classifier model could not give the best results as compared to the hybrid models in which the mining algorithms were combined together to build the proposed classifier model with better accuracy rates and efficiency.

While predicting the heart attack, the most important aspect is the determination of the accuracy. The accuracy of different data mining techniques helps to determine which technique is the most efficient technique to predict heart problems. The accuracy can be defined as the ratio of a total number of records correctly classified to that of the total records taken into consideration [17]. The accuracy of the data mining techniques is calculated using the following formula (1):-

$$\text{Accuracy} = (tp + tn) / (tp + tn + fp + fn) \quad (1)$$

where tp is the true positive, tn is the true negative, fp is the false positive and fn is the false negative [18]. The accuracy also depends on the type of risk factors chosen for the determination of the heart attack. Further, it has also been observed that the authors have obtained high accuracy results if the number of attributes taken is increased.

Different research papers from different authors were studied and their results are tabulated (Table 1).

| Author | Publication Year | Mining Techniques Used | Observed Accuracy |
|---|---|---|---|
| Amani Yahyaoui et al. | 2019 | Support Vector Machine (SVM) | 65.38% |
| | | Random Forest(RF) | 83.67% |
| Mamatha Alex P and Shaicy P Shaji | 2019 | Support Vector Machine (SVM) | 85.88% |
| | | Random Forest(RF) | 85.88% |
| | | K Nearest Neighbour (K-NN) | 83.21% |
| | | Artificial Neural Network (ANN) | 92.21% |
| Bandarage Shehani Sanketha Rathnayakc et al. | 2018 | K Nearest Neighbour (K-NN) | 75.00% |
| | | Decision Trees | 77.00% |
| | | Artificial Neural Network (ANN) | 78.00% |
| | | Naïve Bayes | 83.00% |
| J. Thomas and R. T. Princy | 2016 | Hybrid model using K Nearest Neighbour (K-NN) & Decision Tree | 80.60% |
| Srabanti Maji and Srishti Arora | 2019 | Hybrid-DT | 78.00% |
| | | Decision Trees | 76.66% |
| | | Artificial Neural Network (ANN) | 77.00% |
| Hnin Ei Ei Cho and Nan Yu Hlaing | 2019 | Random Forest(RF) | 70% |
| | | Artificial Neural Network (ANN) | 73% |
| | | Support Vector Machine (SVM) | 65% |
| Ritu Chauhan et al. | 2018 | Random Forest(RF) | 75.60% |
| | | Decision Trees | 75.10% |

**TABLE 1:** Comparison of Data Mining Technique

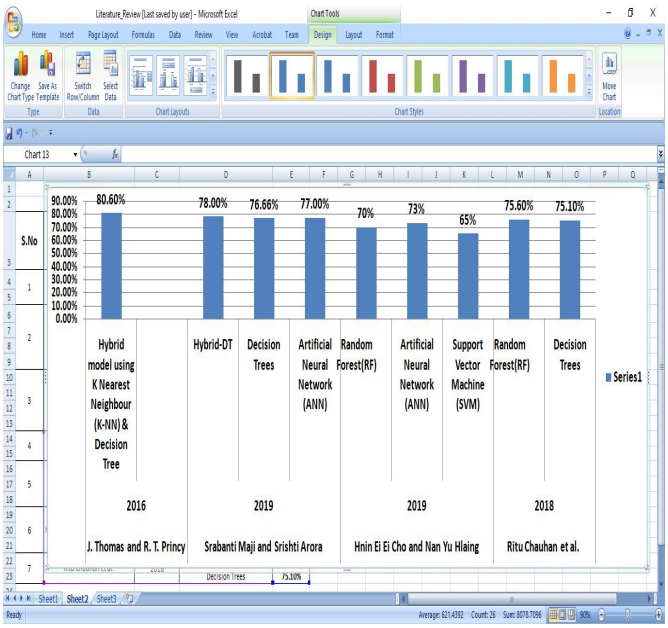


**Fig 2.** Comparison between data mining techniques

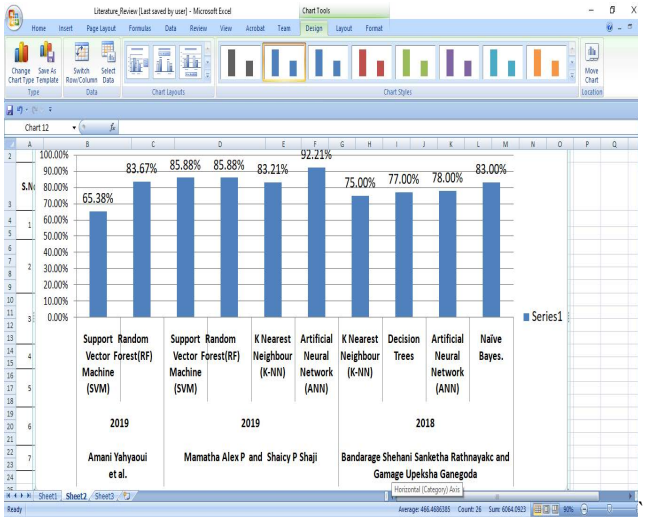


**Fig 3.** Comparison between data mining techniques

# Limitations Identified

1. Data Collection: Capturing data is a huge task as complete, accurate and formatted data from multiple platforms in the healthcare industry is not possible.
2. Storage and Cleaning: It is very difficult to manage and store such huge volumes of data because of its cost, security and performance issues. Cloud storage is now an alternative option but that also has its own pros and cons.
3. Data Analysis: Data in medicine does not conform to the patterns and trends hence data mining becomes a task. In the healthcare industry, a slight change in the associated patterns can affect the balance between life and death completely.
4. Privacy: Faulty results and compromised privacy are the major concerns for the security issues in the implementation of the data mining algorithms. Taking statistical decisions about the diagnosis and treatment may raise questions on personal autonomy.
5. Unstructured Clinical notes: The biggest challenge is text mining the information from the clinical notes like extraction of name, entity, biomedical report or symptom.

# Conclusion

The prediction of the various diseases with the help of different data mining algorithms at an early stage involves many challenges but minimizes human efforts and saves time; thereby saving the lives of people. This paper aimed to compare various mining algorithms that could be implemented in the health care industry to predict the risk factors causing heart diseases. It is observed that only one single model cannot produce good results with the different datasets. Further, it is observed that a combination of many techniques can be used in the early diagnosis and prediction of the disease with better accuracy results. Related work was reviewed and the specific data mining algorithms used in the healthcare sector were mentioned. The accuracy of the data mining algorithms can be increased by having more attributes in the dataset [10]. Hence, we can say that there is a vital requirement of the mining algorithms to analyse the data in the healthcare sector and thereby diagnose the disease at an early stage.

# References

Deshpande, S. P., & Thakare, V. M. Data Mining System and Applications: A Review. International Journal of Distributed and Parallel Systems, 1(1), 32–44. (2010)

Alex P, Mamatha; Shaji, Shaicy P: Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique, IEEE 2019 International Conference on Communication and Signal Processing (ICCSP) - Chennai, India (2019.4.4-2019.4.6)] (2019)

Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. In 1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings. Institute of Electrical and Electronics Engineers Inc. (2019)

Ogundele I.O, Popoola O.L, Oyesola O.O, Orija K.T. A Review on Data Mining in Healthcare, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 9, September 2018, ISSN: 2278 – 1323(2018)

Singh, P., Singh, S., & Pandi-Jain, G. S. Effective heart disease prediction system using data mining techniques. International Journal of Nanomedicine, 13, 121–124. (2018)

Thomas, J., & Princy, R. T. Human heart disease prediction system using data mining techniques. In Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016. Institute of Electrical and Electronics Engineers Inc. (2016)

Priyanka, N., & Ravikumar, P. Usage of data mining techniques in predicting the heart diseases - Naïve Bayes & decision tree. In Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2017. Institute of Electrical and Electronics Engineers Inc. (2017).

JI-WON BAEK AND KYUNGYONG CHUNG, "Context Deep Neural Network Model for Predicting Depression Risk Using Multiple Regression", SPECIAL SECTION ON MACHINE LEARNING DESIGNS, IMPLEMENTATIONS AND TECHNIQUES, IEEE Access 10.1109/ACCESS.2020.2968393, January 21, 2020.

Maji, S., & Arora, S. Decision Tree Algorithms for Prediction of Heart Disease. In Lecture Notes in Networks and Systems (Vol. 40, pp. 447–454). Springer. (2019).

Tomar, D., & Agarwal, S. A survey on data mining approaches for healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241–266. (2013).

Ahmed Zriqat, I., Mousa Altamimi, A., & Azzeh, M. A comparative study for predicting heart diseases using data mining classification methods, International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 12(2017, April 10)

Soni, J., Ansari, U., Sharma, D., & Soni, S. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications, 17(8), 43–48. (2011).

Chauhan, R., Jangade, R., & Rekapally, R. Classification model for prediction of heart disease. In Advances in Intelligent Systems and Computing (Vol. 584, pp. 707–714). Springer Verlag. (2018)

Isra'a Ahmed Zriqat, Ahmad Mousa Altamimi, Mohammad Azzeh, "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods", International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 12, December 2016.

Hnin Ei Ei Cho, Nan Yu Hlaing , DATA MINING CLASSIFICATION TECHNIQUES FOR CARDIOVASCULAR DISEASE DIAGNOSIS.

Hazra, Animesh & Mandal, Subrata & Gupta, Amit & Mukherjee, Arkomita & Mukherjee, Asmita. (2017). Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. Advances in Computational Sciences and Technology. 10. 2137-2159.

B. S. S. Rathnayakc and G. U. Ganegoda, "Heart Diseases Prediction with Data Mining and Neural Network Techniques," *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1-6, doi: 10.1109/I2CT.2018.8529532.

Das T.R., Hasan S., Sarwar S.M., Das J.K., Rahman M.A. (2021) Facial Spoof Detection Using Support Vector Machine. In: Kaiser M.S., Bandyopadhyay A., Mahmud M., Ray K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4_50.

Poonia, P. R. (2003). Fundamentals of Data Warehousing.