Analytical Study on Cardiovascular Health Issues Prediction Using Decision Model-Based Predictive Analytic Techniques

Anurag Bhatt, Sanjay Kumar Dubey and Ashutosh Kumar Bhatt

Abstract The Healthcare industry has grown tremendously in recent years providing best possible medical facility in such an effective manner, i.e., in terms of time as well as cost. Healthcare industry needs to define the standard and need to bring the analytical approach to next level. In this Paper, we have gone through various researchers ideas that represent their approach to effectively provide the solutions regarding prediction of various cardiovascular health issues at multiple levels. Predictive analytics differ from the descriptive and prescriptive analytics by utilizing patient's medical records and analyze them with various statistical techniques as well as advanced machine learning algorithms. This paper will present recent research using various predictive analytical tools and techniques in order to analyze the cardiovascular health issues and in predicting the future outcome of the analysis with much efficient and effective way. There is a need to analyze and examine possible future work to have a better understanding of applying more hybrid and effective algorithms.

Keywords Data mining · Naïve Bayes algorithm · Data preprocessing Cardiovascular disease

A. Bhatt · S.K. Dubey (✉)
Amity University, Noida, Uttar Pradesh, India
e-mail: skdubey1@amity.edu

A. Bhatt
e-mail: anurag15bhatt@gmail.com

A.K. Bhatt
Birla Institute of Applied Sciences, Bhimtal, Nainital, Uttarakhand, India
e-mail: ashutoshbhatt123@gmail.com

M. Pant et al. (eds.), *Soft Computing: Theories and Applications*,
Advances in Intelligent Systems and Computing 584,
https://doi.org/10.1007/978-981-10-5699-4_28

1 Introduction

Data extraction and pattern recognition is one of the most relevant fields in knowledge extraction process. In Today's world, loads of data are available from numbers of organization showing the record of each activity or event occurred. This numerous amount of data can be processed and used to extract the actual meaning or knowledge by reducing data sizes and by gracing us with immense knowledge. Medical informatics and clinical research is one of the unplowed fields that can yield valuable information to replenish medical services with knowledge for betterment of medical services. Every year according to WHO (world health organization) annual report, more than 12 million deaths occur worldwide and in USA the death rate due to Cardiovascular diseases are increasing day by day [1]. Previously, data mining was used as to discover patterns through unsupervised learning, i.e., without making assumptions about the structure of the data. Heart diseases including sudden cardiac death (SCD) are one of the fatal diseases in India. Applying intelligent data mining algorithms in cardiovascular disease (CVD) datasets can help to develop a system that can predict the future outcomes of cardiovascular disease and can take precise investigation report values and can predict the degree of heart disease (Fig. 1).

Forecasting of results on the basis of predictive analytics is the part of data mining concerned that gives us insight of various futuristic probabilities and analyzed trends.

There are lots of techniques that are used to determine the accuracy of data mining algorithms in medical datasets, but while working with cardiovascular diseases like heart disease, sudden cardiac arrest we need to work with variable like heart rate variability (HRV) [2] based on Echocardiogram dataset [3]. In this paper, we have introduced the differences between predictive data modeling, descriptive, and prescriptive data modeling. The purpose of descriptive analytics is to summarize what happened from the data given, while predictive analytics focuses on providing the forecasting that "what might happen in the future" rather than what

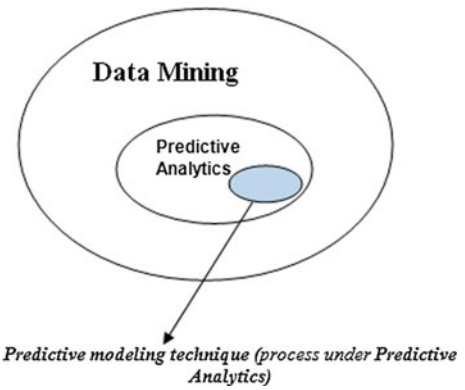

Fig. 1 Relationship between data mining, predictive analytics, and predictive modeling

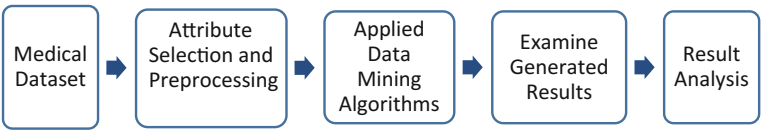

Fig. 2 Steps of medical dataset analysis

will happen in the future because all predictive analytics is based on the probability distribution of events. Prescriptive analytics is the mode of action derived by the results of predictive analytics. Prescriptive analytics basically relies on the "what course of action is required to do the task." The healthcare industry generally relies on doctor's expertise [2, 4] and experience for diagnosis of diseases, while decision support system is required in order to diagnosis of heart disease through prediction. The idea of developing a framework is very simple in order to understand, it proposes the use of sample training set in which we'll run various algorithms of data mining and according to the results shown by it, accuracy of algorithm and predicted value per class will be determined so that we can analyze our results and in the same way, required values will be provided to the decision support system and generated results will determine the possible health vulnerabilities according to provided data (Fig. 2).

2 Literature Review and Problem Discussion

Cardiovascular disease detection is need of the hour as this problem is mushrooming day by day. Data mining application for the detection of these diseases has proved that data mining algorithms can be considered as the most effective approach for these disease detections. Researchers [5] have proposed that using neural network approach for early detection of cardiovascular diseases produces effective results. In [3], we have seen that researcher has shown results of Electro cardiograph (ECG) signal derivative, i.e., HRV and magnetic resonance imaging (MRI) images classification through classification algorithm like support vector machine (SVM) [6, 7] that provides an eye-opening fact about the data and researcher favors statistical analysis over classifiers. Researchers have shown results of ECG signal derivative [8, 9], i.e., HRV that provides an eye-opening fact about the data and researcher favors statistical analysis over classifiers.

Cardiovascular disease detection is divided into various parts through which the results will be generated and these generated results will provide information about data. The whole process of designing optimized framework is divided into several parts. These parts are represented in the form of activity plan which will be followed accordingly.

- **Selection of Dataset**: In this activity phase, we are concerned about selection of dataset that will be using in running results. These datasets must be normalized and organized in such a way that no redundancy is found while running tools into the dataset.
- **Attribute Selection**: Attribute Selection method involves a unique method of selecting attributes that cause variations in the result of data and this variation provides a better approach to deal with results.
- **Data mining algorithm**: In this activity phase, we are concerned about which data mining algorithms we need to apply for better results from data. We need to select the algorithm with high accuracy and precision. Genetic algorithm and associative classification [10] are used by researchers for better results.

Paper [11] has proposed Intelligent Heart Disease Prediction System using data mining techniques using decision tree, Naïve Bayes, and neural network. This paper has demonstrated a prediction system using these classification algorithms (Table 1).

Table 1 Other studies on predicting cardiovascular health issues

Ref.	Database	Features used	Classification method
[4]	Cleveland database	Training the neural network using feed forward neural network	Neural network technique
[11]	Cleveland heart disease database	Data mining extension (DMX) query language	Decision tree, ID3 (iterative dichotomized 3)
[19]	South African medical practitioners	Classification using J48, bayes net	Simple cart and REPTREE (regression tree representative) algorithm
[23]	UCI machine learning repository	Weka tool is used	Naïve Bayes and decision tree
[24]	One minute of ECG Signal	HRV (heart rate variability)	Wigner ville transform
[25]	UCI breast cancer database	Classification, clustering, rule mining	Association rule mining
[26]	Blood-glucose homo monitoring data	Association, classification, subgroup discovery	Best predictor and rules to predict glycemic control
[21]	Physio bank database	Attribute selection and em clustering	Rule base system by clustering techniques

3 Cardiovascular Health Issues Prediction Using Decision Model and Predictive Analytic Approach

Many hospital systems today are designed with hi-tech medical facilities and clinical support stuffed with huge amount of data and these data are required to be analyzed for patterns. In most cases, medical reports of various patients are stored in the database of hospitals that define the various attributes of diagnosis report. Suppose, we have a diagnosis report showing "gender," diagnosis derivatives like "cholesterol," HRV.

Clinical decisions are often based on the doctor's experience and their intuitions that are developed through many years of practice [12, 13]. We want to propose such a framework that helps in defining the accurate prediction based on the clinical data provided to the system [1, 14]. The main concept behind making the framework is to apply data mining algorithms [4, 11], i.e., genetic algorithms, classification algorithms, and techniques like classification based on predictive association rules (CPAR), classification based on multiple association rules (CAMR), C4.5, mesocyclone detection algorithm (MDA), decision tree algorithms, classification and regression tree (CART), but it must have the ability to successful selection of appropriate data mining algorithms. Naïve Bayes algorithm and associative classification can be considered to use datasets to produce effective results [15].

The research will aim at finding solution and proposing a framework that will show the optimal results based on given datasets. In this research, we are focusing on creating a design and framework for cardiovascular disease prediction system [16] that will help heart disease patients or other cardio patients to identify their early symptoms and degree of complexity of the disease [17]. An optimal framework will have certain degree of correctness of data that will define its capability to optimize results from given datasets.

Through this paper, we have used data mining algorithms into consideration for giving better performance and various data analysis tools will be used. Data analysis and pattern finding process will be based on the use of various data mining algorithms as well as predictive analytical techniques, i.e., neural networks, belief networks, cluster analysis, etc. In healthcare delivery environment, predictive analytics can bring surprising results by forecasting various medical outcomes, i.e., cost-effective treatments, doctor's availability, and persistent care facilities (Fig. 3; Table 2).

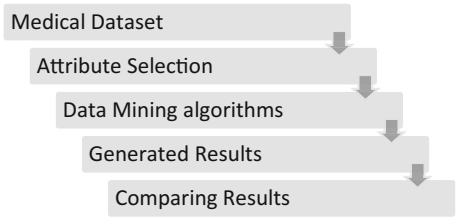

Fig. 3 Steps involved in attribute selection to compare results

Table 2 Data mining technologies for cardiovascular health issues analysis and prediction

Year	Reference	TOOL/technique/model/dataset	Objective	Analysis
2011	[4]	Open source Cleveland dataset	Result oriented paper using neural network approach to analyze heart disease dataset	Experiment showed that neural network performed better when used with multilayer
2008	[11]	Open source Cleveland dataset	To develop an Intelligent Heart Disease Prediction System using data mining algorithms	Naïve Bayes performed better than decision tree by identifying all important medical predictors
2014	[19]	Data collected from South Africa medical practitioners, Naïve Bayes, REPTREE, cart	Predictive analytics is used to analyze the data using various techniques	REPTREE, simple cart and J48 describes the importance of various predictors for accuracy
2014	[23]	UCI (UC irvine) machine learning repository	Result oriented paper using UCI datasets and providing analytical results	Naïve Bayes provided best result as compared to decision tree, while running in weka
2011	[24]	MIT-BIH database and ECG signal using HRV as feature	Sudden cardiac death prediction using HRV features	SCD prediction is done using HRV as a feature of ECG signal, with KNN (*k*-nearest neighbor) classifier as a technique
2008	[26]	Classification techniques like CAMR, CPAR, C4.5, MDA	To use cardiovascular disease features to develop a model for better prediction	CPAR, SVM outperformed of all other techniques and HRV helped in cardiac disease detection
2013	[25]	UCI (University of California, Irwin) machine learning repository, Pima Indian breast cancer data warehouse	To analyze the application of data mining in medical image mining using classifiers and rule mining	PCAR (packet delivery conditions aware routing) algorithm is efficient to predict the cancer levels in breast cancer
2011	[21]	SVM, nearest neighbor technique and clustering approach	To predict cardiovascular events using ECG time series data	Anomaly detection framework is used to identify patients

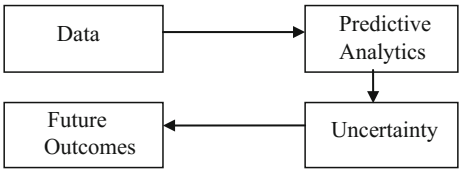

Fig. 4 Steps involved in data analytics

4 Heart Disease Prediction Using Predictive Analytics

Cardiovascular diseases are the leading cause of death in both young and old age group people, leaving behind the traces of events that caused this disease. In this modern society, changing lifestyle, improper nutrition, and mental stress are considered as one of the most effective reasons behind having these cardiovascular diseases, even in young people. We are basically suggesting heart disease prediction system using predictive analytics on the basis of decision based model. In order to predict, we need sufficient amount of knowledge and test cases required to calculate the probability of occurring of these events, which are based on data mining algorithms. Knowledge extraction is one of the clear approaches that allow us to identify those hidden patterns required for finding patterns in the cardiovascular disease data. There are multiple approach to use decision-based model like researcher used genetic algorithm and associative classification in [10], Naïve Bayes in [18]. In order to get the desired result, we also need to consider the performance of the algorithms, on the basis of which efficient algorithm will be associated with particular dataset. Researcher Abhisek Taneja [1] has proposed J48 pruned algorithm and Naïve Bayes algorithm for the datasets taken from UCI Machine Learning Repository. In this paper, researcher has proposed the methodology in which data analysis has been done in two ways: using selected attributes and using all attributes. First, attribute selection method is taken as in data preprocessing phase and attributes are then run through J48 and Naïve Bayes Algorithm with taking both as selected attributes and all attributes (Fig. 4).

5 Heart Disease Prediction Using Attribute Selection

5.1 Using Selected Attributes

Attribute selection is one of the data preprocessing techniques used to generate results. Data preprocessing techniques work in defining the attributes required in the most effective way. Attribute selection process comes under preprocessing where some relevant attributes are considered as efficient as compared to others on the basis of the algorithm applied for preprocessing (Fig. 5).

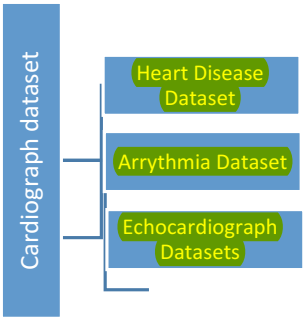

Fig. 5 Categories of datasets

5.2 *Using All Attributes*

The experimental analysis of the paper can be concluded with both the techniques: with selected attributes and all attributes. Results are generated totally different from each other when compared while executing with other algorithms. Attributes selection plays an important role in generating results, as every attribute plays an important role in result gathering and analysis. Results in [1] are found different when attributes run through data preprocessing as the overall performance of the classifiers are different. Attribute selection algorithms select attributes according to the effectiveness of the overall attributes in the given subset. Results were found numerically different when run through selected attributes as compared to all attributes (Fig. 6).

Naïve Bayes algorithm and J48 algorithm in [1] are compared with the help of bar representation. In this way, we can easily compute the algorithmic performance when provided with selected attributes and with all attributes, respectively. Naive Bayes algorithm outperformed J48 in both the cases, i.e., with selected attributes and all attributes.

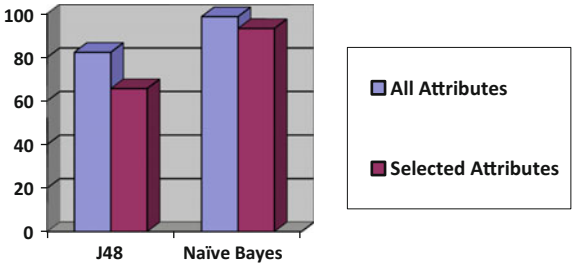

Fig. 6 Result analysis using all attributes and selected attributes

6 Decision Model Based Predictive Analytics

Cardiovascular health issues are constantly affected by the multivariate predictors used in order to predict the cardiovascular disease in general public which is based on their provided data. Predictive analytics in this paper is discussed on the basis of considering multivariate prediction, i.e., multivariate predictive models, tools, and scores are used in order to increase the precision rate of predicting cardiovascular health issue. In the way to multivariate predictive analytics, various factors play an important role, i.e., multivariate predicts help in better analysis of the predictive outcome as it helps in testing the predictive outcome generated by model through local environment as well as external environment. Updating procedures should be used in order to improve the performance of the models generated by considering various predictive factors and the updates in the predictive model must be cross validated with the external environment variables as well as development samples by monitoring the performance of the model generated.

7 Conclusion

In this Paper, this survey has discussed various recent studies that relate the various techniques to identify the hidden pattern form scratch through comparing various predictive analytical methods. Research on using these tools based on predictive analytics and generated decision model is critical and will help for forecasting future outcomes and results for the patients about their cardiovascular health issues, because it requires a great deal of testing and conformation before generating real-world predictions. This whole new dimension will identify the root cause of the cardiovascular disease and will help us to analyze new datasets using results produced by training datasets [19, 20]. Predictive analytical technique will help to generate decision model approach used in order to get the deep insight into cardiovascular health issues and helps in notifying the expected outcome based on the probabilistic aspect of that data arrived from model. Researchers [21, 22, 17] have introduced an effective technique for faster ECG data analysis when provided in compressed form. Predictive analysis plays a vital role in anamoly detection and ensures the effective probability distribution, so that it does not affect the future outcome of the results.

But, on the other hand getting optimal results based on the figures or data provided by us will be highly sensitive to standard environment variables. Changes in these values will cause effects in the integrity and accuracy of the result because we cannot predict the exact result based on the training data [10, 15]. This paper has witnessed various scenarios show promises and provide inspiration about future work, and show the importance of using all accessible levels of data in cardiovascular issues prediction, describing best possible probability of developing a framework that will help to analyze the patient's data and will help develop us with

the simple GUI that can determine the prediction using forecasting based on patient's data. This framework will be decision model-based deploying facility that can generate great probabilistic results.

In a nutshell, we can conclude that there is a lot more work is required in the field of machine learning and data mining having specialized branch of predictive analytics which will enable to develop a system that will work in the cellular level to diagnose human body and have a better understanding of applying more hybrid and effective algorithms.

References

1. Taneja, A.: Heart disease prediction system using data mining techniques. Orient. J. Comput. Sci. Technol. **6**(4), 457–466 (2013)
2. Soni, J., Ansari, U., Sharma, D., Soni, S.: Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int. J. Comput. Appl. (0975–8887), **17**(8), 43–48 (2011)
3. Murukesan, L., Murugappan, M., Iqbal, M.: Sudden cardiac death prediction using ECG signal derivative (heart rate variability): a review. In: 2013 IEEE 9th International Colloqium on Signal Processing and Its Application, 8–10 Mac. Kuala Lumpur, Malaysia, pp. 269–274 (2013)
4. Rani, K.U.: Analysis of heart disease dataset using neural network approach. Int. J. Data Min. Knowl. Manage. Process (IJDKP) **1**(5), 1–8 (2011)
5. Chitra, R., Seenivasagan, V.: Review of heart disease prediction system using data mining and hybrid intelligent techniques, ICTACT J. Soft Comput. **03**(04), 605–609 (2013). ISSN 2229-6956
6. Vatankhah, M., Attarzadeh, I.: Proposing an efficient method to classify mri images based on data mining techniques. Int. J. Comput. Sci. Netw. Solut. **2**(8), 38 (2014). ISSN 2345-3397
7. Mao, Y., Chen, Y., Hackmann, G., Chen, M., Lu, C., Kollef, M., Bailey, T.C.: Medical data mining for early deterioration warning in general hospital wards. In: IEEE International Conference on Data Mining Workshops, p. 1042 (2011)
8. Azhim, A., Yamaguchi1, J., Hirao, Y., Kinouchi1, Y., Yamaguchi, H., Yoshizaki, K., Ito, S., Nomura, M.: Monitoring carotid blood flow and ECG for cardiovascular disease in elder subjects. IEEE Eng. Med. Biol. 5496 (2005)
9. Islam, M.R., Ahmad, S., Hirose, K., Molla, M.K.I.: Data adaptive analysis of ECG signals for cardiovascular disease diagnosis, IEEE, 2246 (2010)
10. Jabbar, M.A., Chandra, P., Deekshatulu, B.L.: Heart disease prediction system using associative classification and genetic algorithm. In: International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT 2012
11. Palaniappan, S., Awang, R.: Intelligent heart disease prediction system using data mining techniques, IEEE, pp. 108–115 (2008)
12. Nayak, S.R., Dash, B.K., Mishra, S., Bhutada, T., Jena, M.K.: Sudden cardiac death in young adults. J Indian Acad Forensic Med. **37**(4), 438–440 (2015). ISSN 0971-0973
13. Langara, B., Georgieva, S., Khan, W.A., Bhatia, P., Abdelaziz, M.: Sudden cardiac death in young man. In: Case Report, Dept of Respiratory Medicine, Tameside General Hospital, Ashton Under Lyne, UK, vol. 11. March 2015
14. Sudhakar, K., Manimekalai, M.: Study of heart disease prediction using data mining. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **4**(1), 1157–1160 (2014). ISSN 2277-128X
15. Kaur, B., Singh, W.: Review on heart disease prediction system using data mining techniques. Int. J. Recent Innov. Trends Comput. Commun. **2**(10), 3003–3008 (2014). ISSN 2321-8169

16. Somanchi, S., Adhikari, S., Lin, A., Eneva, E., Ghani, R.: Early prediction of cardiac arrest (code blue) using electronic medical records, ACM, 2119–2126, 11–14 August (2015). ISBN 978-1-4503-3664
17. Herland, M., Khoshgoftaar, T.M., Wald, R.: A review of data mining using big data in health informatics. J. Big Data, p. 35
18. Patil, R.R.: Heart disease prediction system using Naïve Bayes and Jelinek-mercer smoothing. Int. J. Adv. Res. Comput. Commun. Eng. **3**(5), 6787–6792 (2014). ISSN 2278-1021
19. Masethe, H.D., Masethe, M.A.: Prediction of heart disease using classification algorithms, In: Proceedings of the World Congress on Engineering and Computer Science. San Francisco, USA, WCECS 2014, pp. 22–24 October (2014). ISSN 2078-0966
20. Manikantan, V., Latha, S.: Predicting the analysis of heart disease symptoms using medical data mining methods. Int. J. Adv. Comput. Theor. Eng. (IJACTE) **2**(2), 5–10 (2013). ISSN 2319-2526
21. Syed, Z., Guttag, J.: Unsupervised similarity-based risk stratification for cardiovascular events using long-term time-series data. J. Mach. Learn. Res. 1002 (2011)
22. Sudhir, R.: A survey on image mining techniques: theory and applications. Comput. Eng. Intell. Syst. (Paper), (Online), **2**(6), 45–46 (2011). ISSN 2222-1719, ISSN 2222-2863
23. Venkatalakshmi, B., Shivsankar, M.V.: Heart disease diagnosis using predictive data mining. Int. J. Innov. Res. Sci. Eng. Technol. **3**(3), 1873–1877 (2014). ISSN 2319-8753
24. Ebrahimzadeh, E., Pooyan, M.: Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals. J. Biomed. Sci. Eng. (JBiSE), 699–706 (2011)
25. Kavipriya1, A., Gomathy, B.: Data mining applications in medical image mining: an analysis of breast cancer using weighted rule mining and classifiers, IOSR J. Comput. Eng. (IOSRJCE), **8**(4), 18–23 (2013). ISSN 2278-0661, ISBN 2278-8727
26. Marinov, M., Mosa, A.S.M., Yoo, I., Boren, S.A.: Data-mining technologies for diabetes: a systematic review. J. Diab. Sci. Technol. **5**(6), 1550–1551 (2011)