A Review of Heart Disease Prediction System using Machine Learning Algorithms in Data Mining

S. Janani , Department of Computer Science, KPR College of Arts, Science and Research, Coimbatore, India, janusarguru@gmail.com

P. Indumathi, Department of Computer Science, KPR College of Arts, Science and Research, Coimbatore, India, indumathi.p@kprcas.ac.in

Abstract - Heart disease is considered as a terrible disease since it has the nature of looting lot of lives silently. Heart disease is also termed as "cardiovascular disease" that includes Heart Failure, Stroke, Heart attack, Arrhythmia and so on. In the healthcare industry, data exists in the form of patient's record related to illness, treatment, lab test, clinical data, diagnoses, and medical conditions etc. Proper analysis of this healthcare dataset may help the patients to predict their criticalness at the early stage and thus it leads to reduce the money and time spending for the treatment. Data mining plays a vital role in extracting hidden information from the database. Various data mining techniques and machine learning algorithms are involved in developing decision support system that is often involved in prediction purpose. In this paper, various prediction model developed for predicting heart disease at early stage are compared.

Keyword: Data Mining, Heart Disease, Classification Algorithm, Cardiovascular, Review, Feature Extraction

I. INTRODUCTION

Heart disease generally refers to a condition that affects heart. Conditions may involve blocked or narrowed blood vessels that lead to heart attack, chest pain or stroke. As per WHO report, 17.9 million people die each year from CardioVascularDiseases, approximately 31% of the death occurs worldwide. Some of the reasons behind these heart diseases are unhealthy diet, irregular exercise and change in life style. This life threatening disease can be predicted earlier with the help of some predictive models and this model can be developed by implementing machine learning algorithms using Data Mining. The risk of Heart disease increased by the following factors:

- Family History
- Age
- Smoking
- High Blood Pressure
- High Blood Cholestrol
- Hyper Tension
- Obesity
- Physical inactivity

A. Data Mining in Health Care

Data Mining plays a vital role in extracting knowledge/pattern from the large dataset. Using the techniques available in data mining in proper way ensures extraction of valuable information from the dataset. Applying Data Mining techniques in Health care dataset helps the specialist to automate the diagnosis of disease and helps the patients too. Many researchers work on developing heart disease prediction model by using machine learning algorithms and data mining techniques to identify the heart related diseases at early stage. The heart data set includes important features such as Age, Gender, Chest pain, Blood pressure, Cholestrol, Fasting blood sugar and Electrocardiography. Most popular and frequently used algorithms involved in heart disease predictions are Support Vector Machine, Random Forest, KNN, Multilayer Perceptor, K-Means. This paper attempts to show the performance of several prediction models involved in predicting heart disease collected from different dataset.

1.2 Classification

Classification is a data mining technique which is based on machine learning and it is used to classify every item in a set of data into one of the predefined set of classes or groups. Mathematical techniques such as decision trees, linear programming, neural network and statistics are used in classification techniques.

Heart Disease Prediction model mostly uses Classification algorithms to predict heart disease.

The rest of this paper is presented as follows: Section 2 describes the literature about existing survey of heart diseases and hybrid intelligent techniques. Section 3 describes types of heart diseases. Section 4 and 5 describes various data mining and used for the prediction of heart disease and Section 6 involves conclusion.

II. LITERATURE REVIEW

Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir and Ruinan Sun developed a Hybrid Intelligent

 DOI : 10.35291/2454-9150.2020.0689

System Framework for the prediction of Heart Disease by applying various machine learning algorithms on Cleveland heart disease dataset with 13 features. Feature Selection algorithms were applied on heart dataset to identify the important features. In addition, k-fold cross validation method was used and performance evaluation metric applied to check the classifiers performance. Logistic Regression classifier along with cross validation and Relief Feature Selection showed a performance accuracy of 89%.

S.Nandhini , Monojit Debnath , Anurag Sharma and Pushkar proposed a system for monitoring and diagnosing heart disease. This proposed system notifies the user and the Bluetooth attached in it sends a message to the mobile application. The pulse rate sensor (AMPED) attached in it is used for predicting heart rate of a user. Random Forest algorithm outperformed and achieved the accuracy of 89. Diagnosis system predicts the heart disease by using ML algorithm and for monitoring system 20 persons with good health conditions were experimented and 20 individuals with cardiac problem were also experimented. The system showed a good performance and produced accuracy of 100% for both cases.

Elham Nikookar and Ebrahim Naderi developed the model to help physician to diagnose the heart disease by processing SPECT images of patients. This model has three modules namely Partitioning, Inner Classifiers and Fuser. In the first module, 80 samples were assigned to train set and 187 samples were assigned to test set. In the Inner Classifiers module, five base classifiers and 10-fold cross validation were involved and applied on test data and in the Fusion module, a new feature vector generated from the second module is built with 27 features. Newly generated dataset best fits for finding optimal fuser classifier. This artificial Intelligence based model predicts the heart disease and increases the effectiveness of the CAD system in heart disease thus leads to better clinical decision making.

Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava suggested a technique named Hybrid Random Forest with Linear Model (HRFLM). In this model, without any restrictions all features are taken into account. Cleveland Dataset from UCI repository were involved to classify heart disease. Several classifiers are included in predicting the heart disease includes Support Vector Machine, Linear Regression, Naïve Bayes, KNN, Random Forest, Neural Network, Decision Tree, Language Model. It has been found that combining Random Forest with Linear Model classifiers provides best result for early stage heart disease prediction.

Youness Khourdifi and Mohamed Bahaj utilized the Fast Correlation-Based Feature Selection (FCBC) method to improve the diagnosis accuracy and also reduce the redundancy in features. For optimization of feature, Particle Swarm Optimization (PSO) combined with Ant Colony Optimization(ACO) is used. Experimental results showed that the hybrid approach produces better optimization with respect to features. Different classifiers were compared to predict the best classifiers for predicting heart disease. This proposed model optimized by FCBF, PSO and ACO provide an accuracy of 99.65% with K-Nearest Neighbour and 99.6% with Random Forest.

K.Gomathi and Dr. Shanmugapriyaa discussed some effective techniques for heart disease prediction. In this analysis, three classifiers were involved. They are Naive Bayes, Artificial Neural Network and Decision Tree(J48). Using selected classifier algorithm, the accuracy of classifiers is evaluated. The analysis showed that of these three classification techniques Naive Bayes outperformed and predicts heart disease with higher accuracy.

M. Marimuthu, S.Deivarani and Gayathri.R used various classification algorithms to predict the patient with possibility of coronary heart disease in next 10 years. Training dataset with 16 attributes is taken from Kaggle and it is evaluated using K-Nearest Neighbours(KNN), Support Vector Machine, Decision Trees and Naive Bayes classifiers. AUC-Receiver Operating Characteristic Metric is used to predict the accuracy of the classifiers. Compared to all classification techniques used in the model, the accuracy of K-Nearest Neighbours(KNN) is good with 83.60%.

P. Suresh and M.D. Ananda Raj studied various diagnosis model for the heart disease and selected important heart disease feature by applying genetic algorithm. Four classifiers were applied to compare and find the best classifiers which suits well for diagnosis of heart disease. They are Random Forest, NaiveBayes, Decision Tree, Support Vector Machine. For the above purpose, various datasets were experimented where each dataset contains 16 attribute. As per observation, in most of the cases Naive bayes classifier outperformed by having more accuracy compared to other classification algorithms in respect of all the dataset and also observed that Cleveland dataset has less missing values and has all 16 features. Genetic algorithm suggested some important features for accurate diagnosis.

G. Nagrajan, A.P Mohan Raju, V. Logeshwaran, K. Nandhakumar and S. Naveenkumar developed a model for prediction using various classifiers named Support Vector Machine, Random Forest, Naive Bayes, Multi Layer Perceptor, and j48. Particle Swarm Optimization is an advanced approach of feature extraction that extracts important attributes from the dataset and ensures the improvement of accuracy. The results of various classification algorithms were analyzed using validation metrics including MAE, RMSE, RAE and RRSE. From the point of observation, compared to other classifiers, j48 combined with PSO feature extraction approach outperformed with an accuracy of 94.40%.

Akshay Jayraj Suvarna, Arvind Kumar M, Ajay Billav, Muthamma K M and Asst. Prof. Gadug Sudhamsu

 DOI : 10.35291/2454-9150.2020.0689

proposed a methodology which contains Admin side and User's side. In the Admin side, the system is trained using Support Vector Machine and Random Forest Algorithm. On the other hand, user side gets attribute detail from the user and manipulates using the algorithms and predicts the heart disease.

Ravindhar NV et.al performed a work on predicting Cardiac disease using Machine learning techniques. In the UCI repository database, five algorithms including Logistic Regression, Naive Bayes, Fuzzy KNN, K-Means and BP-Neural Network were trained to predict the probability of cardiac disease. In the Neural Network Back Propagation classifier, 50% of the dataset is taken for training and testing the data in each iteration. In this study, it is identified that Neural back propagation outperforms and predicts the cardiac disease accurately with the accuracy of 98.2% compared to other classifiers included.

III. CONCLUSION

In this paper, the various models for the prediction of heart diseases are discussed. Using feature extraction techniques and machine learning algorithms, the prediction can be effective and accurate. These models can be implemented to predict the disease at the early stage to avoid spending money and importantly to save the life. In the future work, a process for the development of new model that outperforms all the above mentioned algorithms in predicting heart diseases accurately may be proposed.

REFERENCES

[1] Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir and Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Hindawi, Volume 2018 |Article ID 3860146.

[2] S.Nandhini , Monojit Debnath , Anurag Sharma and Pushkar, "International Journal of Recent Engineering Research and Development (IJRERD)", ISSN: 2455-8761, Volume 03 – Issue 10, October 2018, PP. 39-46.

[3] Elham Nikookar and Ebrahim Naderi," Hybrid Ensemble Framework for Heart Disease Detection and Prediction", "International Journal of Advanced Computer Science and Applications", Vol. 9, No. 5, 2018, PP. 243-248.

[4] Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", "IEEE Access", VOLUME 7, 2019, PP. 81542-81554.

[5] Youness Khourdifi and Mohamed Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", "International Journal of Intelligent Engineering & Systems", Vol.12, No.1, 2019.

[6] K.Gomathi and Dr. Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification", "International Journal for Research in Applied Science & Engineering Technology (IJRASET)", Volume 4 Issue II, February 2016, ISSN: 2321-9653

[7] M. Marimuthu, S.Deivarani and Gayathri.R , "Analysis of Heart Disease Prediction using Various Machine Learning Techniques", "International Conference on Artificial Intelligence, Smart Grid and Smart City Applications (AISGSC 2019)".

[8] P. Suresh and M.D. Ananda Raj, "Study and Analysis of Prediction Model for Heart Disease: An Optimization Approach using Genetic Algorithm"," International Journal of Pure and Applied Mathematics", Volume 119 No. 16 2018, 5323-5336, ISSN: 1314-3395.

[9] G. Nagrajan, A.P Mohan Raju, V. Logeshwaran, K. Nandhakumar and S. Naveenkumar, "Effective Prediction Model for Heart Disease Using Machine Learning Algorithm", "International Journal of Engineering Research & Technology", RTICCT – 2019 (Volume 7 – Issue 01), ISSN: 2278-0181.

[10] Akshay Jayraj Suvarna, Arvind Kumar M, Ajay Billav, Muthamma K M and Asst. Prof. Gadug Sudhamsu, "Predicting the Presence of Heart Disease Using Machine Learning", "International Journal of Computer Science and Mobile Computing", Vol. 8, Issue. 5, May 2019, pg.119 – 125.

[11] Ravindhar NV, Anand, Hariharan Shanmugasundaram, Ragavendran and Godfrey Winster ," Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE)", ISSN: 2278-3075, Volume-8 Issue-11, September 2019.

 DOI : 10.35291/2454-9150.2020.0689