

Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method - A Comparative Study

Dr. Joy Iong Zong Chen,

Professor,

Department of Electrical Engineering,

Da-Yeh University,

Taiwan.

jchen@mail.dyu.edu.tw

P. Hengjinda,

Department of Industrial Robotics and Automation System,

Phetchaburi Rajabhat University,

Phetchaburi, Thailand.

Abstract- Coronary Artery Disease (CAD) prediction is a very hard and challenging task in the medical field. The early prediction in the medical field especially the cardiovascular sector is one of the virtuosi. The prior studies about the construction of the early prediction model developed an understanding of the recent techniques to find the variation in medical imaging. The prevention of cardiovascular can be fulfilled through a diet chart prepared by the concerned physician after early prediction. Our research paper consists of the prediction of CAD by the proposed algorithm by constructing of pooled area curve (PUC) in the machine learning method. This knowledge-based identification is an important factor for accurate prediction. This significant approach provides a good impact to determine variation in medical images although weak pixels surrounding it. This pooled area construction in our machine learning algorithm is bagging shrinking veins and tissues with the help of clogging and plaque of blood vessels. Besides, the noisy type database is used in this article for better clarity about identifying the classifier. This research article provides the recent adaptive image-based classification techniques and it comparing existing classification methods to predict CAD earlier for a higher accurate value. This proposed method is taking as

evidence to diagnosis any heart disease earlier. The decision-making of classified output provides better accurate results in our proposed algorithm.

Keywords: *Medical Imaging, Machine learning method*

1. INTRODUCTION

One of the dominant diseases in the world is heart disease in the human being. According to the survey of each year, 18M around people is dying in the world due to cardiovascular disease. The heart is functioning in our body to do almost all component activation [1]. Heart disease is dangerous and to be controlled and predicted early. The early prediction can prolong any human dies due to heart disease problem [2]. The different types of heart disease are;

1. Abnormality condition of heart rhythm
2. Tough state of arteries
3. Cardiomyopathy
4. Coronary artery disease
5. Infections attacks heart.
6. Genetic heart disease.

Arrhythmia is one of the heart diseases that generate an abnormal condition in the heart. The tough state is appearing in the wall is called atherosclerosis. The estimation status of heart disease is unpredictable in most of the scenarios [3]. Figure 1 shows coronary artery disease procedure. In the health clinical sector, Recently, health care centers are accelerated to predict heart disease early in the complex and challenging analysis. Data mining techniques are playing important role in the medical imaging and healthcare sector recently [4]. The features are age, sex, blood pressure, chest pain factor, Electro Cardio Gram (ECG) test to measure the activities of the heart function taken for prediction model commonly.

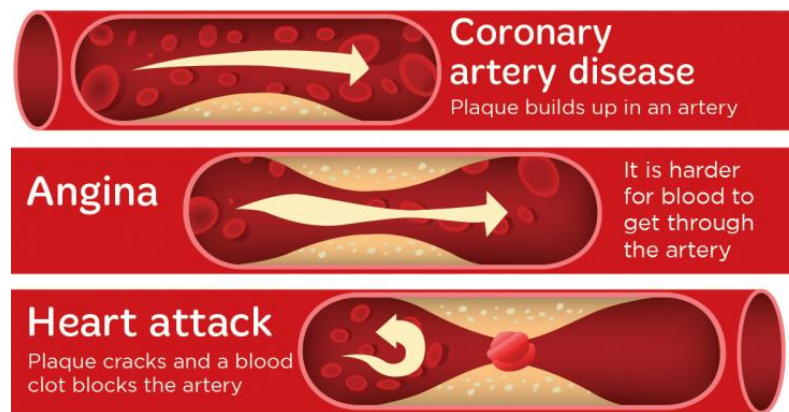


Figure 1 Coronary Artery Disease

The medical centers are considering cholesterol level in serum, peak heart rate, depression level, blood sugar level, smoking habits history, food habit and basic measurements like Body Mass Index (BMI) report for diagnosing heart disease. Arrhythmia type disease can be identifying by improper heartbeat. Generally, sudden changes or loss of heart function are called cardiac arrest disease [5]. High blood pressure is a condition that is raised due to narrowed blood vessels to reduce the flow of blood in the branch. Coronary artery disease will appear based on the damaging of blood vessels. Due to decrease blood circulation to the heart, it receives less energy to function [6]. Therefore, the heart surrounding muscle gets weakens. Finally, it will become arrhythmias disease nothing but failure of heart function. Figure 2 shows a simplified block diagram of the process of the proposed technique.

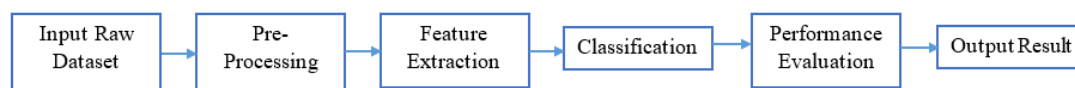


Figure 2 Block diagram of overview of proposed technique

The extracted feature from the given medical image can be classified regression type algorithm to predict Coronary artery disease. The data mining model is very useful to predict the disease in medical imaging. The broad measurement of complex details for the radiologist is challenging for diagnosis [7]. The quality healthcare service is expecting accurate prediction early

at a reasonable cost. Still, there is a huge demand to predict disease early is needed in many developing health sectors. The knowledge-based discovery in the medical field is one of the undeniable progress in the medical field. In the last few years, a huge amount of research papers is published in the interest of medical field diagnosis based on machine learning methods [8]. Heart disease is harmful to affect another part of the human body and it is predicting by studying blood vessels in the surrounding heart's position [9] [10] [11]. The research of cardiovascular disease prediction is providing several methodologies to the diagnosis of it. The decision tree and support vector machine to explore the Coronary artery disease prediction. Many of the research SVM predicts CAD with a minimum error rate and higher precision value [12] [13] [14]. This research article provides different methods to predict CAD and constructing a proposed workflow model based on accurate results.

Research Gap

There is a research gap in the early prediction of coronary artery disease based on clinical parameters. Blood supply reduces to heart function day by day becoming clog with plaque. This is called the early stage of coronary artery disease. Plaque contains cholesterol walls with substances. The oxygen level is decreasing for heart function gradually. During this time, the valve and wall will shrink and gives a strange appearance from the reference image. Our proposed algorithm is used to predict early this coronary artery disease with a higher accuracy level.

2. ORGANIZATION OF THE RESEARCH

The structure of the research article is organized as follows; Section 3 gives related works about prediction methods of heart disease. Section 4 discusses a proposed methodology for early prediction. Section 5 presents the obtained results and a discussion of them. The conclusion and the future direction of the research discussed in section 6.

3. PRELIMINARIES

Many prediction methods are used to determining heart disease in the medical sector by using data mining techniques. Colombet et al investigated ANN and CART techniques in the prediction of heart disease. They succeed in their test with higher accuracy [15]. Yang peilli et al

attempts the survey about determining and predicting coronary disease based on heart signal [16]. G.Parthiban et al investigates the diabetic patient's chances of heart disease and tested their algorithm with a machine learning method. They concluded Naïve Bayes classifier is the suitable method to predict heart disease with a minimum number of the dataset [17]. Jyoti Soni et al classifies and compares the final results of the experiment of the dataset with the Bayesian method to the decision tree method. They were suggesting other methods like KNN and neural networks for their project [18]. Mai showman et al introduces decision tree method classification for diagnosing heart disease patients. Besides, they tested their dataset with various classification techniques. They investigate to improve the accuracy of the system performance for better clinical assessment [19]. Sarath babu et al compares classification methods by bagging algorithm to predict heart disease. They proved bagging algorithm performance rate and accuracy level for the prediction are superior to other methods with the decision tree method [20]. Anbarasi et al present a genetic algorithm to predict heart disease for humans. They were suggesting their algorithm is used to reduce the number of tests that can be reduced to find heart disease problems. They succeed in that to predict any heart disease [21]. Rishabh et al conduct find a precise classification method of predicting heart disease with high accuracy. They have concluded mostly the regression type of classification can be suitable to predict heart disease with higher accuracy than other methods. The proposal of R-C4.5 and discussion of its feature selection for the proposed model. The research is experimented with their device and proved very useful machine to health care sector for prediction of heart diseases [22]. Many research papers are introduced SAS-based software in the computer for the diagnosis of heart disease. The paper incorporated neural network methodology to that [23].

Mohammad Shafenoor Amin et al identified the significant feature from the dataset using naïve Bayes algorithm for logical regression. This is used to predict heart disease with higher accuracy. They believed a maximum number of features can give higher accuracy in the prediction [24]. K.Srinivas et al presented much application of data mining techniques in the health care sector. They developed a proposed technique with a decision tree and the KNN method. They conducted data analysis with the Tanagra tool for machine learning purposes. They tested with around 3000 training datasets. They were splitting data set into training and testing as 70 and 30 percentages. They achieved high accuracy of their proposed algorithm after compared existing

methods [25]. Milli N et al used a backpropagation algorithm for prediction purposes to improve neural network performance in the supervised learning method. This algorithm is based on the correction of the error in exists one. They compared and discussed the merits and limitations of the prediction-based classification algorithm [26].

4. METHODOLOGIES

Figure 3 shows the first phase of proposed architecture.

4.1 First phase of Proposed Algorithm

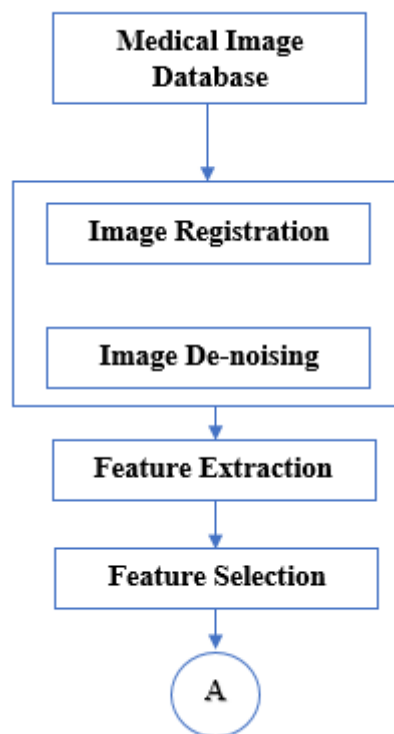


Figure 3 First phase of proposed architecture

Preprocessing

The image registration from the medical raw image database is categorizing under the preprocessing unit. This registration performs on a raw image database and it aligns the image to de-noising processing. Generally, the images will be raw and dusty due to speckle noises that affected a medical image. Therefore, the de-speckling of medical images through filtering approach technique to get a better result for classification.

Feature Extraction

This modern construction is consisting of the work to select variable subset attributes for our proposed algorithm. This is associated with many conditions of the data mining techniques. The selection can be identifying from the huge extracted features [27]. This association idea will not be interfering prediction and classification problems.

4.2 Second phase of Proposed Algorithm

Classification of the extracted features is in the second stage of our proposed architecture. In the data mining problem, classification is one of the virtuosi. The extracted data should be classified into different classes and sections. This separation is classified based on the content and size of the file. The data object to the subset of input class-oriented set. The classification is finding and categorizing the class for better prediction for unknown given data class objects with higher accuracy [28] [29]. The prediction of each class model will be worked out after data class objects separation for each class model.

Feature Database

The datasets are splitting as 70% and 30% into two sections for training and testing dataset. From the feature database, the datasets are getting trained for the classification of a dataset. At the same time, one part of 30% splitting datasets going testing dataset to improve the speed and accuracy of the proposed model. The further iteration of the process can be more accurate than the previous one due to implementation in the error minimization technique in classifiers [30] [31].

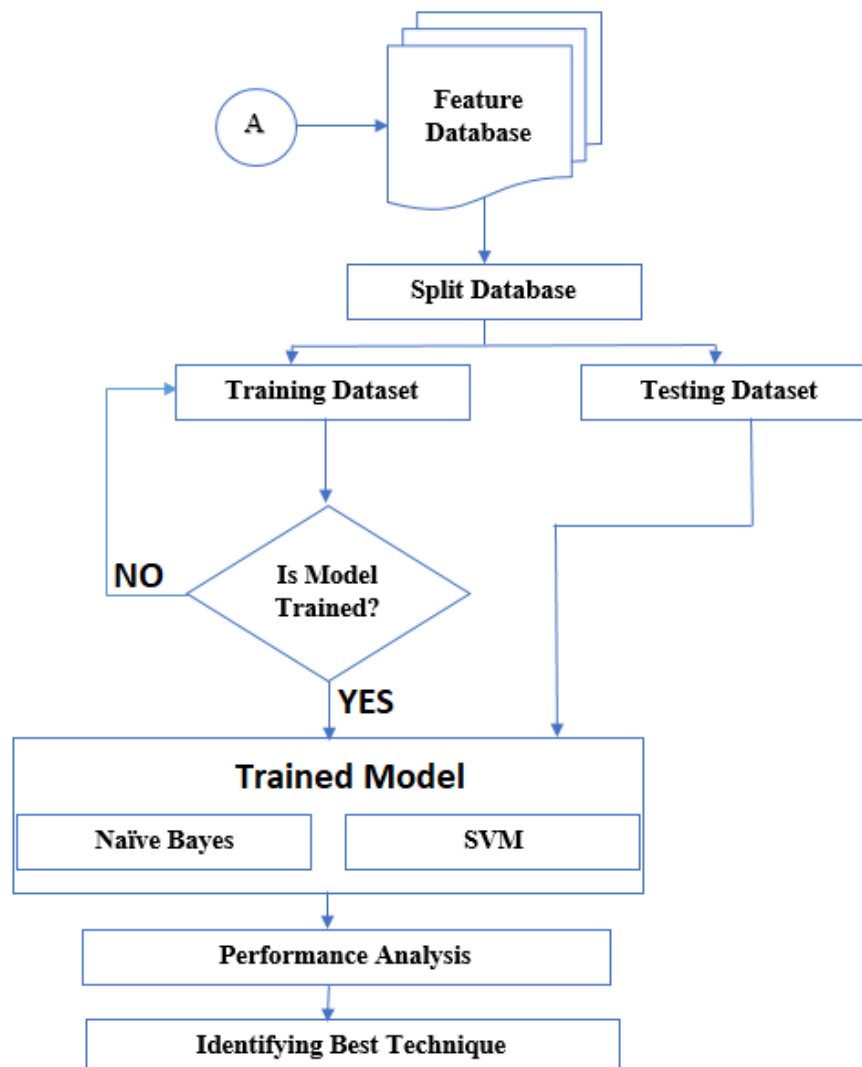


Figure 4 Second phase of proposed architecture

Trained Model

Here the two models have taken which is most efficient classifying based on medical image data. One classifier is simple based on probabilistic method called Naïve Bayes classifier. Another is the Support Vector Machine (SVM) method to recognize patterns and separates in a dataset. The Bayesian network model is constructed with kernel estimation for higher accuracy prediction [32]. The training of the classifier is combining with the probability model. Figure 5 shows structure of Naïve Bayes classifier.

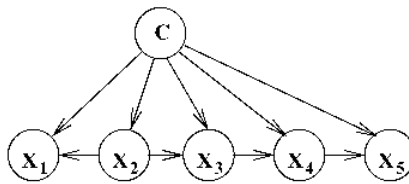


Figure 5 Simplified structure of Naïve Bayes Classifier

$$\hat{u} = \underset{n \in \{1, \dots, N\}}{\operatorname{argmax}} P(C_n) \prod_{i=1}^n P(x_i | C_n)$$

Where C is class variable, x is mutually independent.

This continuous data experiment is associated with variable class and distributed to normal distribution. During classification, the segmentation and computing mean and variance of each class. The average value of the class and corrected variance can be updated for prediction based on the Bessel function. The normal distribution is taking place in class after updating,

$$P(x = v/c_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\left[\frac{(v-\mu_n^2)}{2\sigma_n^2}\right]}$$

The estimation of the class probability can be calculated by a normal distribution which is written in the equation. This event model is calculating discretized feature values for the continuous values with distributed features for prediction.

Support Vector Machine

The preparing set for SVM is important for any classification. The collected data points are separated according to training the classifier based on an algorithm. Figure 6 shows the slope line between two isolated data. The data are classifying linearly through a hyperplane support-vector. The estimating kernel is extending the dataset based on the database [33].

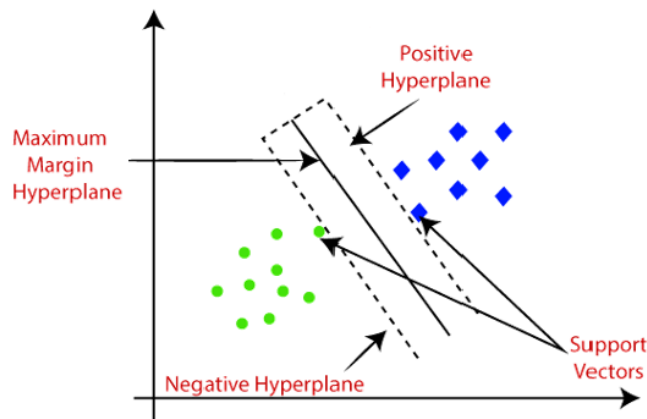


Figure 6 Categorizing isolating data

The mapping technique is completing by support vector or kernel estimation. The mapping is incorporating both linear and nonlinear data information. Figure 7 shows the pooling layer presents in the classification.

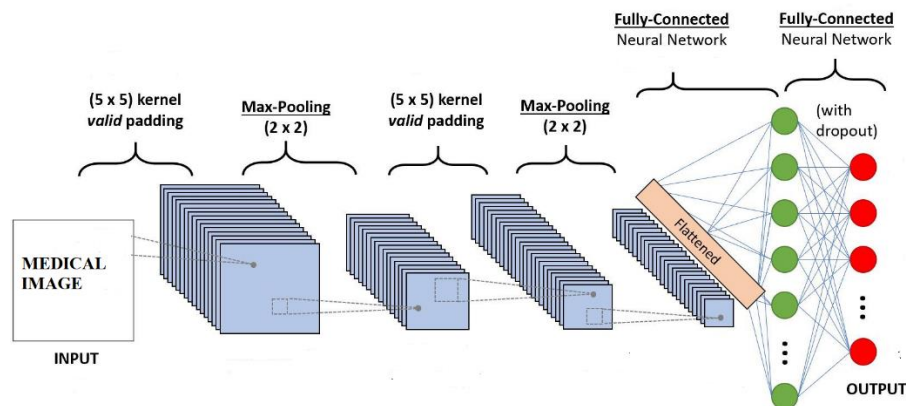


Figure 7 Proposed SVM classifier with Pooling layer

The observation from magnifying the attributes with the pooling layer in the SVM classifier provides good accurate results. Considering primal minimize,

$$= \frac{1}{n} \sum_{i=1}^n \varphi_i + \lambda \|w\|^2$$

According to the kernel risk factor, the nonlinear classification rule is transforming from the mapping of linear classification. The strategy has chosen to minimize the error during classification is derived for solving the problem as an equation;

$$\varepsilon(f) = \frac{1}{n} \sum_{k=1}^n l(\hat{u}, f(\varepsilon_k))$$

Where ε_k is a random variable of empirical

5. RESULTS DISCUSSION

The Naïve Bayes is simple and it shows good efficient performance. This feature classification is processing based on prior probability in the class attribute. This is the statistical regression analysis method to predict the diagnosis with an approximation of dependent attributes. These attributes are categorizing based on data description which has shown in Table 1. But we have achieved good accurate results through the SVM method and it is a good tool for medical diagnosis. The high accuracy is derived from the SVM method. Besides the processing is done by many iterations and replacing and updating the missing attributes at each iteration. This column process is based on arithmetic mean and data replaced with the normal mode of the dataset. Due to splitting the dataset, the processing computation speed has increased overall. This prediction of the coronary artery disease dataset can be comprised of around 80 features. Based on the performance of the classifier the next iteration decides the classification method for the procedure. Their outcomes will be compared and evaluated SVM is a good tool to identify or predict the CAD early based on the image. But naïve Bayes algorithm is a bit lower accuracy compared to the SVM method. It will be evaluated in the final stage of our proposed algorithm and it is tested and proved. An error minimization in the classifier result has shown in figure 8.

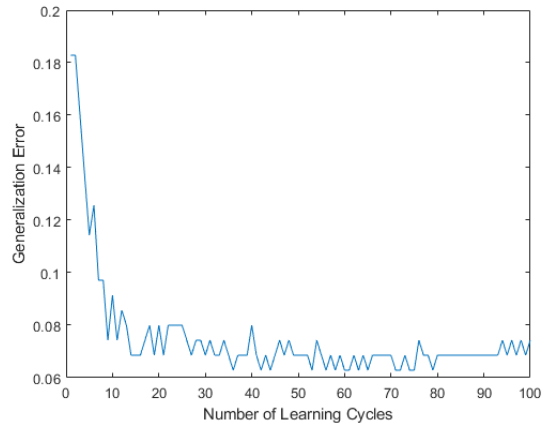


Figure 8 Error Minimization in classifier

Our dataset is comprised of the following description as shown in Table 1. Two types of database structure are used in this research article. Noisier and less noisy database images named database 2 and database 1 respectively.

Table 1 Dataset Description with their value

Feature Name	Description	Value
Age	AGE factor	In Year
Sex	SEX determination	Male, Female, Transgender
Chest	Chest pain type	Low, Normal, tolerable, Intolerable
Blood Pressure	Blood pressure at normal condition	High and normal
Serum cholesterol	Serum Cholesterol in mg/dl	Serum cholesterol level
Resting Electro cardio Graphic Result	Resting electrocardiographic Results	60 to 100 bpm

The following parameter has tested for the class balancing and proposed architecture performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FP}$$



Figure 9 Performance chart of proposed framework for database 1 & 2

In this proposed architecture data set description, only sex determination is language attributes and other features are numeric values. The SVM technique is consisting the favorable prediction estimation parameter by using a fewer number of features in the training dataset. The graph in figure 9 is showing that SVM is an appropriate tool for the prediction of CAD with higher accuracy.

6. CONCLUSION

The machine learning algorithm for early prediction procedure has discussed here. Thus our proposed algorithm has concluded that the overall performance of SVM is high to get accurate prediction results for CAD. The appropriate evaluation is needed to determine the prediction results of the proposed architecture. In our proposed architecture, the most relevant feature has been used and tested for the prediction; besides that dataset is very clear due to noise remove the filter. The diagnosis is very simpler and can be used in the cardiovascular medical field. The irrelevant dataset should be introduced and tested for the overall performance of the proposed architecture. Every medical field needs automatic early prediction of any diseases, especially cardio problem disease for protecting human life in nature. Generally, CAD is a deadly disease by its nature. In future work, our proposed framework database should be updated with more descriptions of patients. To obtain better efficiency, the SVM method will be validated for the

huge amount of database. The proposed algorithm should be extended till Graphical User Interface (GUI) to the medical center of the cardiovascular sector. Besides, the probabilistic approach examination will be conducted for more accurate prediction for CAD further.

REFERENCES

- [1] M. Akhil Jabbar, Priti Chandrab and B.L Deekshatuluc, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", *ICECIT*, 2012.
- [2] Syed Umar Amin, Kavita Agarwal and Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", *ICT*, 2013.
- [3] K. Mathan, Priyan Malarvizhi Kumar, Parthasarathy Panchatcharam, Gunasekaran Manogaran and R. Varadharajan, "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease", *Kluwer Academic Publishers Norwell*, vol. 22, 2018.
- [4] Aditi Gavhane, Gouthami Kokkula, Isha Pandya and Kailas Devadkar, "Prediction of Heart Disease using Machine Learning", *ICECA*, 2019.
- [5] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 8, 2017.
- [6] H. Jayasree, D. S. S.K.R. T. Naren, K. Sai Sankeerth and T. Kumidini, *HEART DISEASE PREDICTION SYSTEM*, July 2019, [online] Available: www.researchgate.net/publication/334612815.
- [7] Jaekwon Kim, Jongsik Lee and Youngho Lee, Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree, *Health Inform Research*, July 2015.
- [8] Priyan Malarvizhi Kumar, S. Lokesh, R. Varatharajan, C. Gokulnath and P. Parthasarathy, "Cloud and IoT based Disease Prediction and Diagnosis System for Healthcare using Fuzzy Neural Classifier", *Future Generation Computer Systems*, vol. 86, 2018.
- [9] Krishnan. J Santhana and S. Geetha, Prediction of Heart Disease Using Machine Learning Algorithms, *IEEE*, 2019.
- [10] C. Kalaiselvi, Diagnosing of heart diseases using average k-nearest neighbor algorithm of data mining, *IEEE*, 2016.

- [11] Han C. W. Hsiao, Sean H. F. Chen and J. P. Jeffrey, Deep Learning for Risk Analysis of Specific Cardiovascular Diseases Using Environmental Data and Outpatient Records, IEEE, 2016.
- [12] Georgia Spiridon Karanasiou, Evanthia Eleftherios Tripoliti, Theofilos Grigorios Papadopoulos, Fanis Georgios Kalatzis, Yorgos Goletsis, Katerina Kyriakos Naka, et al., "Predicting adherence of patients with HF through machine learning techniques", *IET*, vol. 3, no. 3, 2016.
- [13] Gaurav Dangi, Tanupriya Choudhary and Praveen Kumar, A smart approach to diagnose Heart disease through machine learning and Springleaf Marketing Response, IEEE, 2016.
- [14] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia and Juan Gutierrez, A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease, IEEE, 2017.
- [15] Colombet, I.; Ruelland, A.; Chatellier, G.; Gueyffier, F.(2000). "Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression". Proceedings of AMIA Symp 2000, p 156-160.
- [16] Yang Peilli, Yin Xuezhen, Ye Jian, Yang Lingfeng, Zhao Hui and Liang Jimin, Deep learning model management for coronary heart disease early warning research, IEEE, 2018.
- [17] Parthiban, G., Rajesh, A., & Srivatsa, S. K. (2011). Diagnosis of heart disease for diabetic patients using naive bayes method. International Journal of Computer Applications, 24(3), 7-11.
- [18] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43- 48.
- [19] Mai Shouman, Tim Turner, Rob Stocker(2011) "Using Decision Tree for Diagnosing Heart Disease Patients. Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia CRPIT Volume 121 - Data Mining and Analytics,2011.
- [20] Sarath Babu, Vivek EM, Famina KP, Fida K, Aswathi P, Shanid M, Hena M, "Heart Disease Diagnosis Using Data Mining Technique" 2017 IEEE International Conference on Electronics, Communication and Aerospace Technology(ICECA)

- [21] Anbarasi, M., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370-5376.
- [22] Rishabh Wadhawan, "Prediction of Coronary Heart Disease Using Apriori Algorithm with Data Mining Classification"
- [23] Bandarage Shehani Sankheta Rathnayake, Gamage Upeksha Ganegoda, "Heart disease prediction with Data Mining and Neural Network Techniques" 2017-IEEE 4th International Conference For Convergence in Technology
- [24] Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan, "Identification of Significant Features and data mining techniques in predicting heart disease" 2018 ELSEVIER, telematics and Informatics.
- [25] Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)*, 2(02), 250-255
- [26] A1-Milli N, (2013) "Back Propagation algorithm neural network for prediction of heart disease". *Journal of theoretical and applied information technology* 56(1).
- [27] NImai Chand Das Adhikari, Arpana Alka, and rajat Garg, "HPPS: Heart Problem Prediction System using Machine Learning"
- [28] Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan, "Identification of Significant Features and data mining techniques in predicting heart disease" 2018 ELSEVIER, telematics and Informatics
- [29] Meenal Saini, Niyati Baliyan, Vineeta Bassi, "Prediction of Heart Disease Severity with Hybrid Data Mining" *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)-IEEE2017*
- [30] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System" 2016 ELSEVIER *Procedia Computer Science*
- [31] Marjia Sultana, Afrin Haider, "Heart Disease Prediction Using WEKA tool and 10-Fold cross-validation", *The Institute of Electrical and Electronics Engineers*, March 2017.

[32] UCI repository accessed on 20 March 2017, available online at <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>.

[33] D.Karthick, B.Priyadarshini, “predicting the chances of occurrence of Cardio Vascular Disease(CVD) in people using Classification Techniques within fifty years of age” Proceedings of the Second International Conference on Inventive Systems and Control(ICISC 2018)-IEEE.