Journal of Computational Science and Intelligent Technologies
eISSN: 2582-9041
DOI: 10.53409/MNAA/JCSIT/e202203034249
Type: Review Article

A Review of Different Methods for Detecting Cardiac Disease

Anand K[1] Yuvaraj D[2], Sudhakaran P[2] and Hariharan Shanmugasundaram[3]

[1]Associate Professor, Chennai Institute of Technology, Chennai, Tamilnadu, India.
[2]Assistant Professor, Department of Computer Science, Cihan University, Duhok, Iraq.
[3]Dept of Computer Science and Engineering, SRM TRP Engineering College, Trichy, Tamilnadu, India.
[4]Professor, Dept of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, India.
[*]Corresponding Author: Yuvaraj D. Email: yuvaraj@duhokcihan.edu.krd
Received: 3 April 2022; Accepted: 10 August 2022

Abstract: Disease prediction has emerged as a widespread issue in public health. The most common is cardiac disease, which is caused by unhealthy food habits, poor lifestyle choices, and a lack of health awareness. As a result, diagnosis becomes more difficult, creating an active study topic. Conversely, the quantity of knowledge is thought to deal with making the right choices. The necessity for medical parameters and analysis results in accurate risk analysis prediction. This study examines several well-exposed factors and algorithms in this exploration. Here, we give the performance metrics from a study we conducted using accepted methodologies. The research discusses how well popular algorithms for detecting cardiac attacks may be predicted using reliable criteria.

Keywords: Disease Prediction; Machine Learning Algorithms; Heart Disease; Poor Lifestyle; Health Awareness.

1 Introduction

Health Care System has increased human life expectancy with the advancement of drugs, improved medical services and, more importantly, the analysis of patient's health records. Despite these modern developments, there are significant challenges. This has been mainly identified with several factors, including lack of adequate information, unidentified errors, data threats, delayed reporting and others. Conversely, the amount of available information for disease identification is vast, with inaccurate decision-making [21]. To alleviate this issue, computerized medical information systems have been designed for the support system process, focusing on Health Records analysis and disease prediction process.

A more accurate clinical diagnosis is possible, and the success rate could be high only when the patient remembers all the symptoms and reports to the physician when he approaches for the treatment history. Henceforth patient history is recorded using medical management procedures. As a result, clinical decision support systems have gained unforeseen attention due to the advancement of technologies. Recently, many data mining techniques have been widely employed to predict patient disease and are deemed more successful in reducing the death rate [3, 6].

The motivation to do this problem comes from the estimation predicted by World Health Organization, which has predicted that by 2030 nearly 23.6 million individuals will pass on because of cardiac problems. We have presented a study using Machine Learning (ML) techniques for the efficient detection, analysis and prediction of cardiac disease. A major goal is to efficiently identify the diseases, reduce the mortality due to the heart-related disease, and provide preventive measures. Machine learning is used for knowledge extraction.

The overall organization of the paper goes in this order. The basic need for the study is presented in section I. The related study and analysis are presented in section II. The basic idea of the study process is presented in sections III and IV, which focus on experimental analysis. Section V describes the evaluation metrics, and finally, Section VI gives the conclusions and future work.

2 Related Works

Much research has been developed for a better and more precise model for the Heart Disease Dataset. A rule-Based approach has been presented for efficient Cardiac Disease prediction, focusing on several known parameters like age, sex, chest pain, blood pressure, cholesterol levels etc. [1]. In the earlier study, an average accuracy of 86.7% was found using the 10-fold method. The accuracy produced in the testing phase is 86.3 % and in the training phase is 87.3%, respectively [2]. Additional parameters can also be used for the study, like Cardiographic results, heart rate achieved, depression analysis, etc.

Research studies with heart disease classification using a forward Heart Network Algorithm with the clever land dataset have obtained a good accuracy outcome of 98% [4]. A more generalized CVD algorithm with the incorporation of age, high-density lipoprotein cholesterol, blood pressure, hypertension and diabetes have been evaluated in one another study with 1174 participations, including 456 females and 718. Hybrid Reasoning-based using Fuzzy set theory, case-based reasoning (CBR), and K-NN have yielded better prediction outcomes [6].

Similar works in the proposed development of cardiac disease prediction and validation are done using NIPPON DATA. The work has also arrived at an experimental cardiovascular prediction tool which considers the standard factors as discussed previously. For the study, data was collected from 2742 residents of Japan [7]. The validation of the model is predicted based on a rules-based procedure [8].

Nahar et al. [9] have studied the prediction work using a knowledge-driven approach. It is found that the Logistic Regression mechanism has yielded results with 77% accuracy [10]. At the same time, an analysis with Naive Bayes algorithms has reported a classification accuracy of 81.48% [11].

Investigation using different data mining approaches for diagnosing heart disease involving naïve Bayes, decision trees and neural networks have yielded promising results [12]. Results illustrated the proficiency of naïve Bayes compared to the other two approaches. Also, the K-means clustering approach is investigated in this research process to find the result outcomes, which is also proven to be one of the most popular schemes [13]. Here the authors have analyzed different methods using centroid selection, inlier, outlier, random attribute selection and k-means clustering technique.

Probabilistic artificial neural networks BF networks were used for the study [14], which reveals a class of radial basis function for pattern recognition, nonlinear mapping and proper estimation of class membership probabilities with likelihood ratios. The experiments reveal the significance of the proposed approach. A total of 576 records and 13 medical attributes were chosen for the study process, and the best accuracy performance was obtained at 94.60%.

K-Nearest Neighbour (KNN) is a classical classification problem for a research study [15] due to its simplicity and high convergence speed. Though it is successive in many cases, the disadvantage is the large memory requirement needed to store the whole sample [16]. Currently, most of the work relies strongly on cholesterol level analysis. In cases where a fasting blood sample was unavailable, alternatively, non-HDL cholesterols were chosen for the comparison instead of LDL cholesterol levels [17-18].

Decision curve analysis and prediction models were another procedure to access the clinical value [19]. Also, there exists a study adopting the evaluation of prediction models with the calculation of the difference between true positive values and false positives [20].

3 Methodology

We describe the implementation system procedure and the dataset used for the study in this cardiac prediction system in detail in this section. The major modules focus on three broad phases elaborated below in Fig 1.

3.1. Medical Data Collection

In this phase, the patient approaches the doctor for a clinical checkup of possible cardiac disease. The Doctor collects all the information from the patient and from the ECG and other scans. The information includes the data such as Age, Gender, Chest Pain, Cholesterol, Resting ECG, Resting Blood Force, Heartbeat Rate etc. [21].

3.2. Data Processing and Classification

The study involves UCI Machine Learning Repository, a more popular database set used by the machine learning community. This could be an interesting empirical analysis procedure involving different machine-learning algorithms.[22] Throughout the study in this paper, we have utilized Python libraries like scikit learn, numpy, pandas and matplotlib. The work presented here uses both test train split idea and cross-validation for optimal parameters selection choice, which has revealed some promising results (discussed in the later part of section V).

3.3. Disease Detection Workflow Process

As we found several algorithms have led to some promising results using machine learning procedures, we focus on the study using training and test capability with raw data [23]. We have chosen five learning algorithms for our study purpose and analyzed the results as categorized into three groups (refer to Table 1). The study reveals the behaviour of analyzing various algorithms in the subsequent section using a well-known dataset.

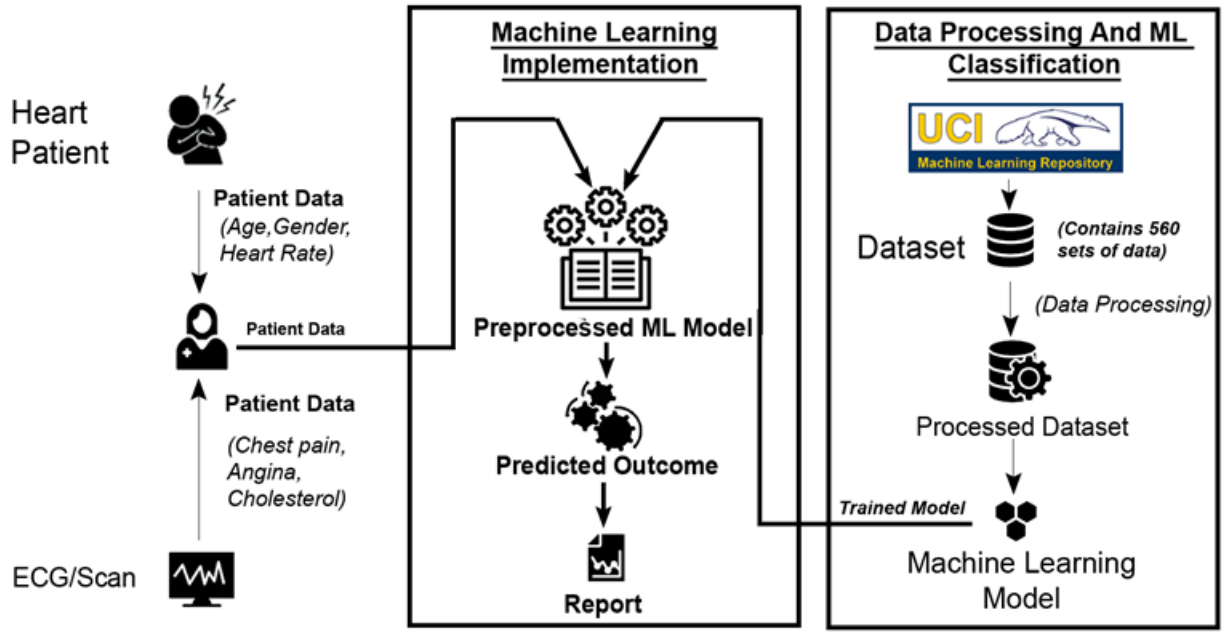

Figure 1: Proposed Model for the Cardiac Disease Prediction

Table 1: Algorithms used to perform disease detection

Categories	Algorithm
Classification	Fuzzy K-Nearest Neighbour (Fuzzy-K-NN) Logistic Regression, NB (Naïve Bayes)
Clustering	K-Means
Neural Networks	Back Propagation Network (BPN)

4 Results and Discussion

A research study based on heart disease analysis for detection is analyzed in this section. The study has discussed the prediction of cardiac vascular diseases and performance measures using machine learning techniques as outlined in Table 1. The pre-processing, training and testing, and feature selection were performed with the state and availability of details concerning the dataset associated with Cardiac Disease. The methodology focused on here (presented in Fig 1) has certain limitations and is implemented with limited algorithms.

Generally, the dataset has been collected from various repositories, including the UCI repository and Kaggle. Some users from worldwide voluntarily publish their data through clinical agencies. The environment chosen for the study was built on a windows platform. It uses a Jupyter notebook with an anaconda which incorporates the necessary packages. These packages include a matplot library for plotting the points; numpy, pandas, pickle and sklearn are used to perform. All dataset documents are stored by importing the dataset according to a scheme containing a unique numeric identifier, timestamp, label for each data, and a numerical field with corresponding attributes.

The dataset chosen in the study process here contains 600 patterns. From the patterns revealed, pieces of information set are analyzed, those which are proper and noisy are filtered, and redundancies in the data are removed. This filtering is done by applying a standard pre-processing technique. Machine Learning and Neural network techniques are applied during the training stage to extract useful information from raw data. In pre-processing, a certain ratio has been taken for training and testing the data set. Since the dataset contains a large number of samples, we use 10-fold cross-validation in the study process. With the same hardware and software configuration, the evaluation of performance measures of each algorithm under the same set of datasets is evaluated. The 2:3 ratio means 40% of samples are used for testing, and the remaining 60% are used for training the model.

Based on the risk factor analysis is proceeded. Fig. 2 indicates the relation between the risk factors Age and cholesterol, which determines the risk of cardiac disease based on this relation. Resting blood Pressure places a vital role in cardiac disease. Fig 3 indicates the risk, and safe resting pressure ranges from a set of patients chosen for the study process.

4.1. Risk Factor Analysis

Table 2: Risk Factors

Risk Factors
Age
Cholesterol Levels
Chest Pain
Resting Blood Pressure
Maximum Heart Rate Reached

Fig 3 indicates the percent of the chest pain type in which the presence of cardiac disease is determined; thus, Chest pain is one of the risk factors. Fig 4 indicates the range of safe and risk based on the cholesterol and the resting blood pressure, where resting blood pressure is plotted on the x-axis and the cholesterol is plotted on the y-axis (presented in Fig 5).

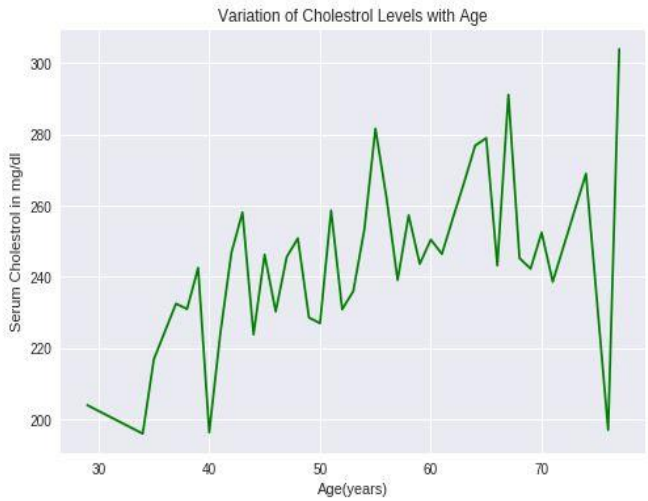

Figure 2: Plotting the variation of cholesterol levels with Age

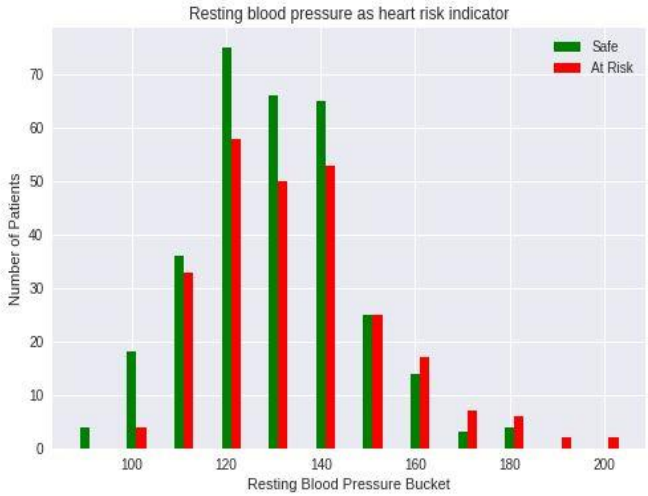

Figure 3: Chest Pain Types

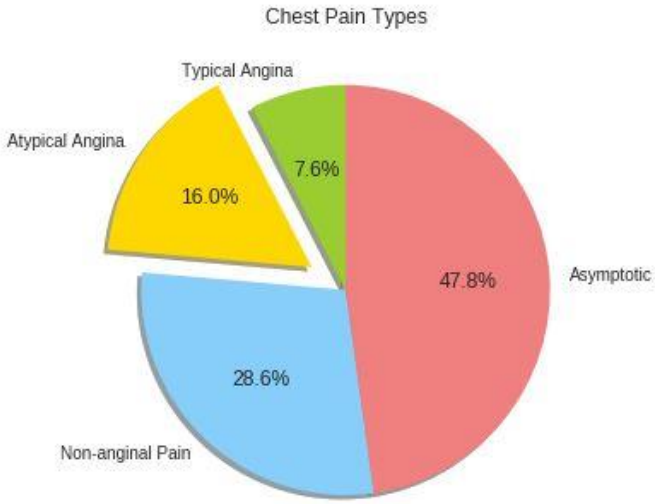

Figure 4: Resting blood pressure as a heart disease risk indicator

4.2. Evaluation Metrics

Here the section focuses its attention on metrics for performance, stability and scalability evaluation. The study uses six different algorithms, as presented in Table 1.

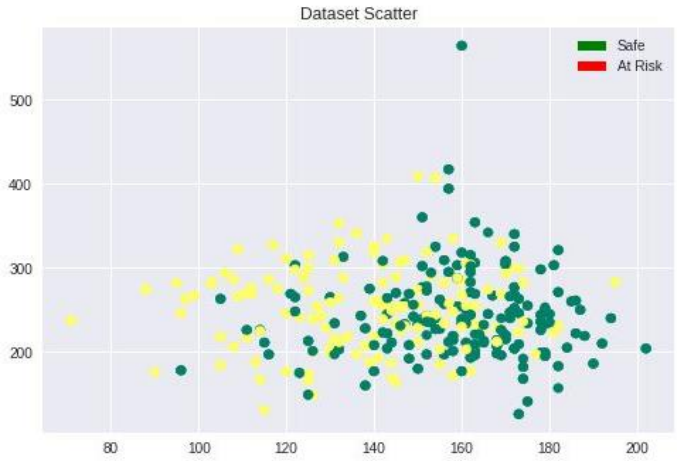

Figure 5: Scatter plot of Resting Blood pressure and cholesterol levels

The performance measure is used to calculate the accuracy of the data set with the help of standard metrics, namely True Positive (TP), True Negative (TN), False positive (FP) and False Negative (FN) values.

The diagrammatic representation of the study algorithms is shown in Fig 6 using the algorithms presented in Table 1. Results were analyzed using a 10-fold cross-validation technique, as mentioned earlier. The study results infer that the best performance measures with diseased and not as 1 and 0, respectively. For each algorithm implementation, cross-validation has been done to gain accuracy. Neural network shows the highest accuracy among the entire five algorithms. The diagrammatic representation of all five algorithms' accuracy is shown in Fig 6.

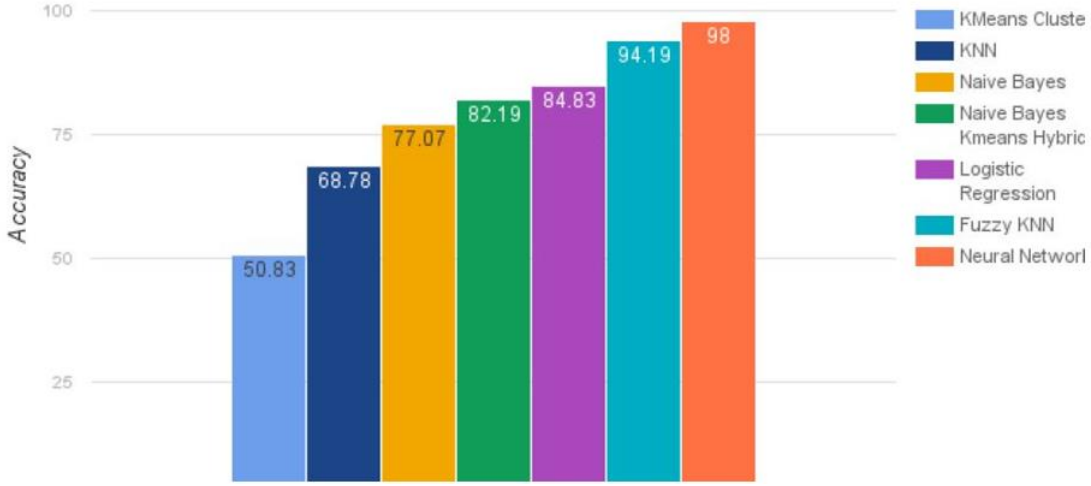

Figure 6: Comparison of various ML approaches

5 Conclusion

This paper analyzed the cardiac disease detection model using a well-known machine learning approach with a learning model based on regularized logistic regression, NB, KNN, K-Means Clustering and BPN. The evaluation compares several machine learning approaches. The results infer that Neural Networks algorithms were found to be a strong competitor as compared to others. It is also inferred that the results

obtained are higher since the dataset chosen is minimal. However, it has led to lower results when the dataset has a higher volume of data. We focus on identifying the optimal point that would lead to better average results, which is focused on our future work.

Acknowledgements: The authors thank their families and colleagues for their continued support.

Funding Statement: The author(s) received no specific funding for this study.

Availability of Data and Materials: The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Purushottama, Kanak Saxenab and Richa Sharmac, "Efficient Heart Disease Prediction System", Vol: 85, pp: 962 – 969, (2018).

[2] Somayeh Nazari, Mohammad Fallah, Hamed Kazemipoor and Amir Salehipour, "A fuzzy inference- fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases", pp: 261–271, (2018).

[3] Chen M, Hao Y, Hwang K and Wang L, "Disease prediction by machine learning over big data from healthcare communities". IEEE Access, Vol:88, pp: 69–79, (2017).

[4] Manjusha B. Wadhonkar, P. A. Tijare and S. N. Sawalkar, Classification of Heart Disease Dataset using Multilayer Feed forward back propagation Algorithm, Vol:4, (2013).

[5] Ralph B. D, Agostino, Ramachandran S. Vasan, Michael J. Pencina, Philip A. Wolf, Mark Cobain, Joseph M. Massaro and William B. Kannel, "General Cardiovascular Risk Profile for Use in Primary Care", Vol:117, pp: 743–753, (2008).

[6] Malathi D, Logesh R., Subramaniyaswamy V, Vijayakumar and Arun Kumar Sangaiah, "Hybrid Reasoning-based Privacy-Aware Disease Prediction Support System, Vol: 73, pp: 114–127, (2019).

[7] H. Arima, K. Yonemoto, Y. Doi, T. Ninomiya and J. Hata, "Development and validation of a Cardiovascular risk prediction model for Japanese: The Hisayama Study", Hypertens. Res. Vol:32, pp:1119–1122, (2009).

[8] Takanori Honda, Daigo Yoshida, Jun Hata, Yoichiro Hirakawa, Yuki Ishidaa, Mao Shibata, Satoko Sakata, Takanari Kitazono, Toshiharu Ninomiya, "Development and validation of modified risk Prediction models for cardiovascular disease and its subtypes". The Hisayama Study, Vol:279, pp:38–44, (2018).

[9] J. Nahar, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," Expert Systems with Applications, Vol:4, pp: 96-104, (2013).

[10] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. Schmid, and S. Sandhu, "International application of a new probability algorithm for the diagnosis of coronary artery disease," American Journal of Cardiology, Vol. 6, pp. 304-310, (1989).

[11] N. Cheung, "Machine learning techniques for medical analysis," B.Sc. Thesis, School of Information Technology and Electrical Engineering, University of Queensland, (2001).

[12] Palaniappan, S. and R. Awang, "Web Based Heart Disease Decision Support System using Data Mining Classification Modeling Techniques", Proceedings of iiWAS, (2007).

[13] Bramer, M., Principles of data mining. 2007: Springer.

[14] I. S. F. Dessai, "Intelligent Heart Disease Prediction System Using Probabilistic Neural Network," Karnataka, (2013).

[15] F. Moreno Seco, L. Mico, and J.A. Oncina, "Modification of the LAESA Algorithm for Approximated kNN Classification," Pattern Recognition Letters, Vol:24, pp. 47–53, (2003).

[16] E. Alpaydin, “Voting over Multiple Condensed Nearest Neighbors. Artificial Intelligence Review”, Vol:11, pp. 115–132. (1997).

[17] H.E. Bays, “A lipidologist perspective of global lipid guidelines and recommendations, part 1: lipid treatment targets and risk assessment”, J. Clin. Lipidol. Vol:10, pp: 228–239, (2016).

[18] T. Imamura, Y. Doi, T. Ninomiya, J. Hata and M. Nagata, “Non-high-density lipoprotein cholesterol and the development of coronary heart disease and stroke subtypes in a general Japanese population: the Hisayama study”, Atherosclerosis, Vol. 233, pp: 343–348, (2014).

[19] A.J. Vickers and E.B. Elkin, “Decision curve analysis: a novel method for evaluating prediction models”, Med. Decis. Mak. Vol: 26, pp 565–574 (2008).

[20] A.J. Vickers, B. Van Calster and E.W. Steyerberg, “Net benefit approaches to the evaluation of prediction models”, molecular markers, and diagnostic tests, BMJ 352 (2016).

[21] 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, Vol. 8, Issue-11, September 2019.

[22] Sundara Vadivel, P., Yuvaraj, D., Navaneetha Krishnan, S., & Mathusudhanan, S. R. (2019). An efficient CBIR system based on color histogram, edge, and texture features. Concurrency and Computation: Practice and Experience, 31(12), e4994.

[23] Yuvaraj, D., Sivaram, M., Karthikeyan, B., & Abdulazeez, J. (2019). Shape, color and texture based CBIR system using fuzzy logic classifier. CMC-COMPUTERS MATERIALS & CONTINUA, 59(3), 729-739.