Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8, July 2021: 5241-5250

Analytic Approach for Heart Disease Prediction using Supervised Machine Learning Algorithms

Charul Sharma [1], Vanita Batra [2], Dr. Meenu Manchanda [3], Dr. Deepak Goyal [4], Dr. Pankaj Gupta[5]

[1]M.Tech Scholar, Department of Electronics and Communication Engineering, Vaish College of Engineering, Rohtak, , sharmacharul53@gmail.com

[2]Assistant Professor, Department of Electronics and Communication Engineering, Vaish College of Engineering, Rohtak, vanita.batra@rediffmail.com

[3]Professor, Department of Electronics and Communication Engineering, Vaish College of Engineering, Rohtak, meenumanchanda73@gmail.com

[4]Professor, Department of Computer Science and Engineering, Vaish College of Engineering, Rohtak, deepakgoyal.vce@gmail.com

[5]Professor, Department of Computer Science and Engineering, Vaish College of Engineering, Rohtak,

pankajgupta.vce@gmail.com

Abstract

According to a recent study of WHO, heart linked ailments are increasing day by day. More than 17.9 million individuals get scummed every year due to this disease. With growth in population, it becomes so much problematic to analyze and start treatments at initial stages of disease. But recent technological advancements have made it possible to accelerate medical diagnosis by early prediction of health-related issues. Therefore, the prime motive of this research is formation of a ML model for heart diseases predictions dependent upon the interrelated parameters. In our research work three machine learning algorithms are implemented, Decision tree, Random Forest and Support Vector Machine (SVM). Out of implemented algorithms SVM give the best accuracy which is 82.01% and followed by decision tree 78.94%. Random forest algorithm gave the minimum accuracy. Accuracy more than 70 % is measured good, but if accuracy is very high than it may be case of over fitting. Therefore, accuracy around 80% is commendable.

Keywords: SVM, Random Forest, Decision Tree, Machine Learning

Introduction

Cardiac disease is a major global health problem in modern medicine. The twenty first-century adage consummate proliferation in life expectancy and a significant transference in the causes of heart disease bereavement throughout the world [1]. Today it is interpreted for approximately thirty percent decrease across the globe including approximately 40 percent in the high-income country and twenty-eight percent in low and middle-income countries. Compelled by economic development, suburbanization and associated with circadian life changes this constant transition is arising around

the world among all races, ethnic groups, and nations at an even faster rate than the last century. Recent developments of modern life style exponentially increase the heart failure rates.

Medical Decision Making

Medical decision making is central to all patient care activities which involves selecting an action from among alternatives. Along with performing procedures and providing comfort, deciding among options for the preventions, diagnosis, and treatment of disease constitutes the critical competencies for the medical profession [4]. Clinicians are often faced with decisions based on the patient's condition. Some decision is made in full partnership with patients. All these decisions have consequences, and some will ultimately determine whether a patient will survive illness and recover without a disability. Banning.M et al. explains about the various models about clinical decisions of current technological perspective Perhaps the biggest challenges of clinical decision making are that decisions are almost always made under circumstances of uncertainty. Another complicating aspect of clinical decision–making is that the physician is not making choices in isolation [2]. The effect of a decision on the outcome of an individual patient can rarely be known. The clinician is more often in partnership with the patient, sharing the decision making and recommending choices that respect the patient's preferences, values, and goals. The true breadth of the science of clinical decision making is enormous, spanning, the disciplines of statistics, sociology, and machine learning techniques [7]. In earlier time clinical decisions help the professionals to reduces the risk factors. Knowledge of medical evidence is necessary for physicians but this knowledge it needs to be appropriately integrated into a valid decision-making strategy that promotes the best interest and confidence of the patients [3]. Clinical decisions are growing increasingly complex. The array of diagnostic and therapeutic options is expanding rapidly. Figure 1 depicts of cardiac conditions in developing countries

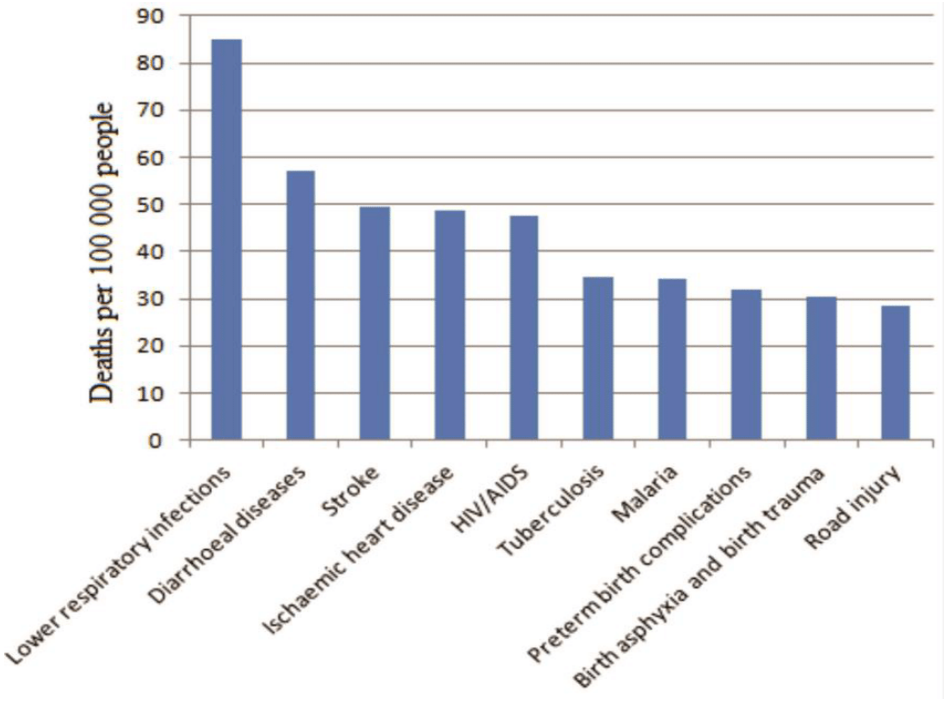

Fig.1 cardiac conditions in developing countries

The cost of care is escalating, which creates the potentially competing pressure to consider value in selecting among clinical strategies. The time available to make decisions is often brief, particularly with the shortening of the average patient visit. Figure 2 depicts of cardiac conditions in developed countries.

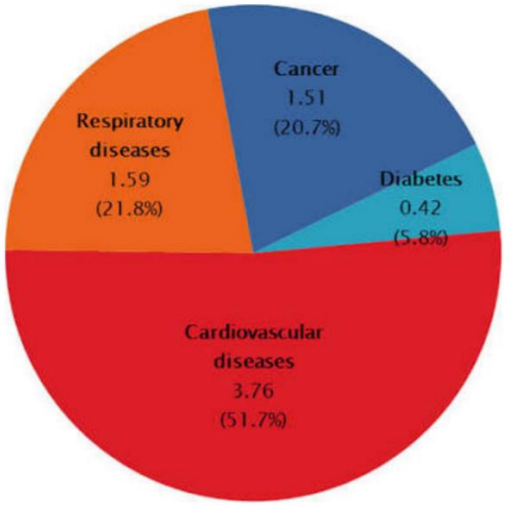

Fig. 2 Cause of death in developed countries, 2015

1. Types of Cardiac Disease

There are several categories of heart diseases. Figure 3 shows the various types of heart disease based on clinical conditions. These categories are broadly classified as myocardial infarction, heart failure, heart arrhythmia, angina pectoris, cardiomyopathy, atrial fibrillation based on their clinical evidence. Heart disease has many features, which affect the function or structure of the heart.

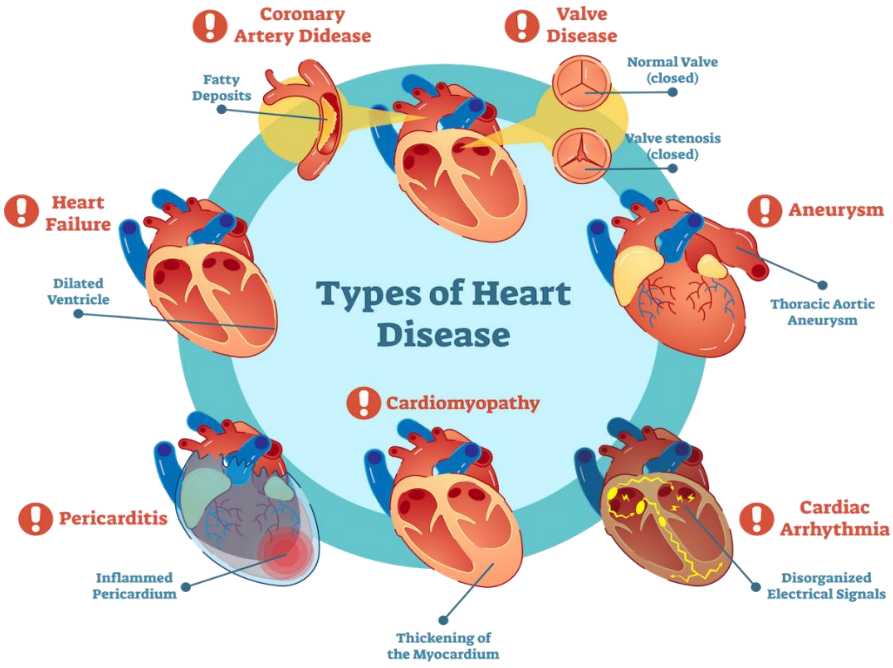

Fig. 3 Types of Cardiac Disease

Coronary Artery Disease:

The coronary artery disease is discomfort induce by depleted circulation of blood. The depletion supply in arteries will damage the vein and produce the discomfort to the regular systolic and diastolic function of the heart.

Acute myocardial infarction

Clinical name for a cardiac arrest is acute myocardial infarction. A cardiac arrest is a condition that fatty substance present in the blood value affect the rate of flow which results tissue damage on arteries [5-6]. The blockage arteries may not be able to supply the oxygenated blood supply to the body which will result in the dysfunction to other organs. Figure 4 explains a type of heart arrest caused by intense pressure.

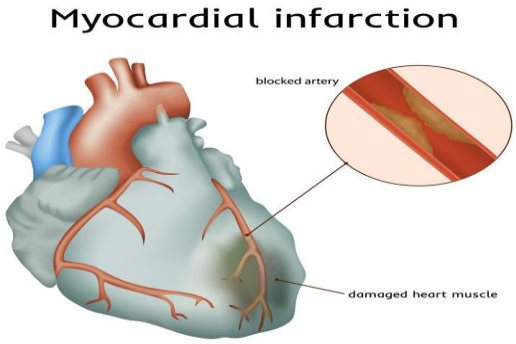

Fig. 4 Acute Myocardial Infarction

Chest Pain (Angina)

Clinical name of chest pressure is Angina. It is predominant medical attention need emergency treatment for the patients. Patients has to treated with ventilators immediately if we experience this type of discomfort. Due to the poor supply of blood flow will cause the pressure on the blood walls and affect the blood vessels [10]. Which will create pressure on the blood vessels results chest pain. Figure 5 shows typical angina caused in the coronary vessel.

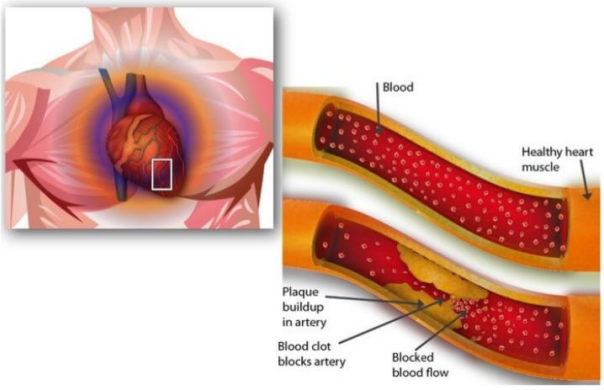

Fig. 5 Angina

Stable angina is the condition causes in peritoneum. Irregular blood flow between the peritonitis walls. The main reasons of unstable angina are lifestyle modification, behavioral habits.

2. Methodology

First of all, we will train different classification models on the training set & will analyze the optimized accuracy**.** In our implemented research work there will be three different models will be implemented which are listed below:

- SVM (Support Vector Machine)
- Decision Trees
- Random Forest

Decision Tree:

This type of learning method mostly used across healthcare industries. Figure shows the model of decision tree types. Each decision tree contains a root node, branches, and leaf nodes. The topmost node is the root node, and all the other node are the leaf or the branch nodes. The internal node imposes the decision rule or a test on one or more attributes of the given data. Whereas the branch node defines the output. Decision trees are well-known method for classification, as it does not require any prior knowledge of data distribution

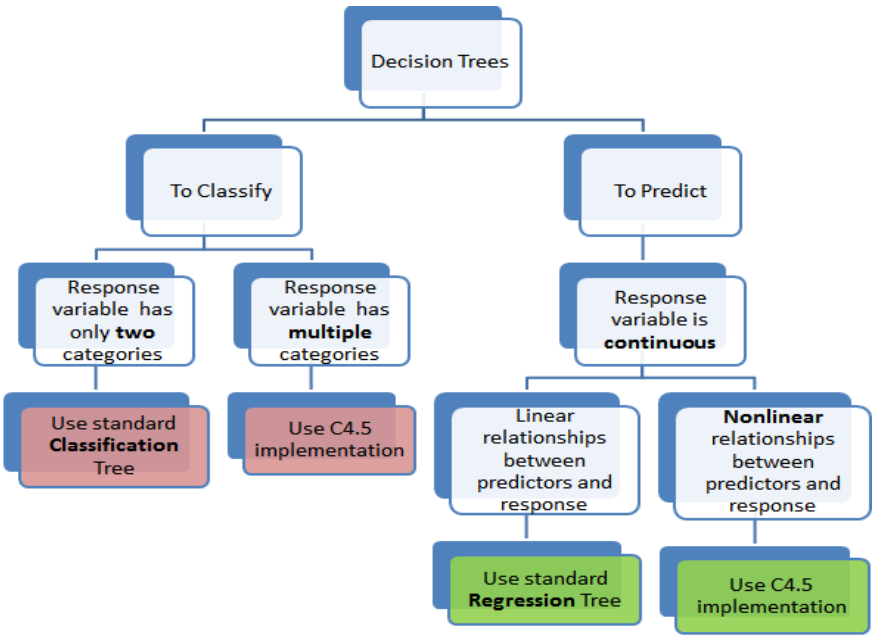

Fig.6 Decision Tree

Classification and Regression Tree

Rutkowski et al. explain in his article that classification and regression tree is a well-known method of decision tree-based classifier. Every base nodule depicts a unique input and a separate base point over the variable. It assumes the input attribute value to be numeric one. The leaf node represents the output variable which is used for prediction purpose [11]. It is based on discriminant analysis and builds a statistical model to classify the dataset with higher accuracy measures. It works well on both the categorical and continuous attributes.

Support Vector Machine:

Support Vector Machine is an overseen AI figuring which can be utilized for used for regress and gathering. In any case, it is commonly used in game plan issue. There is various way through which anyone can depict data via n-dimensional space with estimating every aspect of component being inference of a specific arrangement. Support vector machine technique is based upon the concept of machine learning. In machine learning there are three types of learning exist, supervised learning, unsupervised learning and reinforced learning. SVM technique falls under supervised learning [12]. Every feature is identified by support vector machine and well in advance SVM technique is edified by known samples.

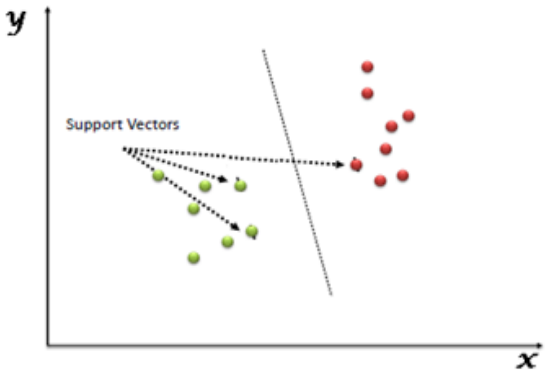

Fig.7 SVM differentiate two classes

Random Forest:

One of the keys, machine learning algorithm, random forest falls under the category of supervised learning. Before going to deep insight of random forest algorithm, first of all try to identify with fundamental blocks of random forest, which is a decision tree. Decision tree is also a supervised machine learning algorithm, in which data is classified into subsets. The classification procedure begins with binary divide and it goes until divide can be possible. By doing this, different branches are created of different length. The main focus of decision tree algorithm is to put training data in a nutshell so that generated tree would be nominal feasible [13]. In this method, to keep tree size as minimum as is the main logic rule as compared to other methods. If there would be small tree, then they give faster result with respect to larger tree, besides these small trees are easy to understand and look cool.

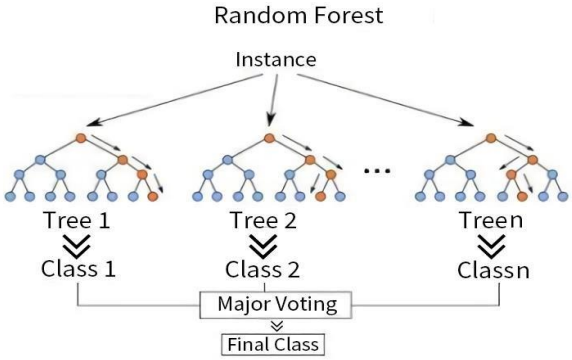

Fig.8 Plot of Simplified Random Forest

3. Result

In our implemented work, three types of supervised machine algorithms are used which are listed below:

- Decision Tree
- Support Vector Machine
- Random Forest

Reseach work carrioed out using google colab using python langauge. In this section results of all implemented supervised machine algorithms are depicted

Decision Tree:

```
[41] accuracy_score(y_test,yhat)

     0.7894736842105263

[42] accuracy_score(y_train,obj.predict(x_train))

     0.8898678414096917
```

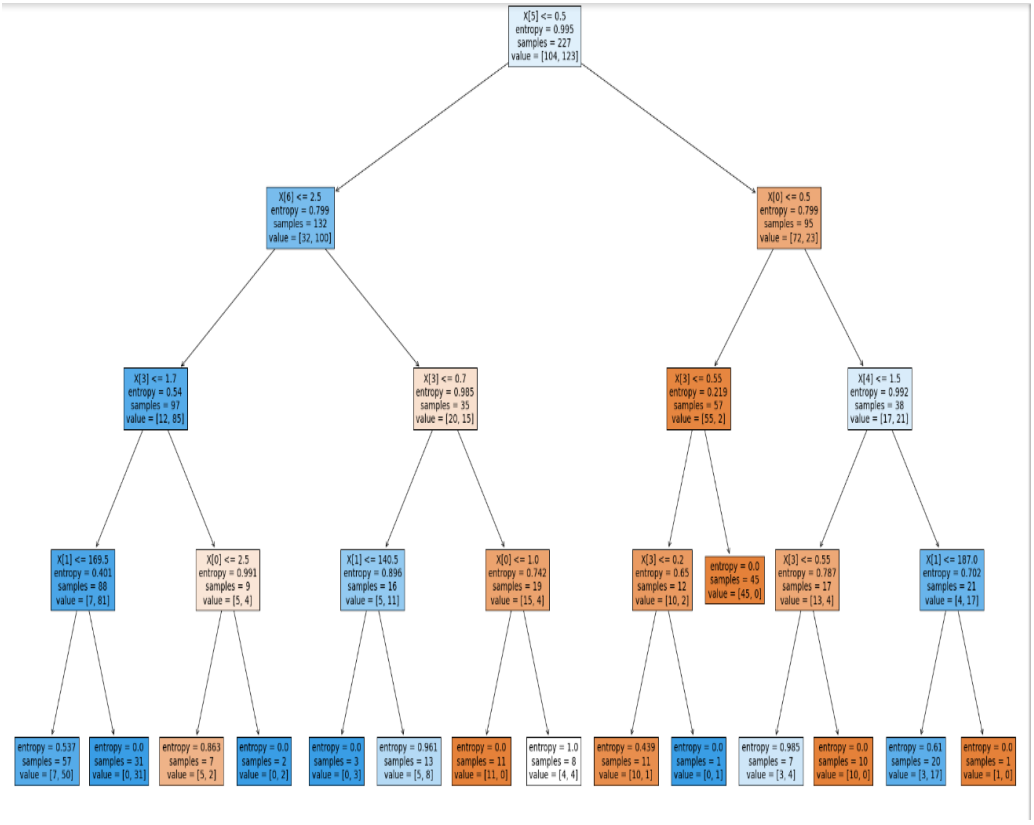

Fig.9 Decision tree formation after implementing algorithm

Support Vector Machine Algorithm Result:

```
(37+44)/(37+6+4+44)
```

```
0.8901098901098901
```

```
[ ] cnf_matrix = confusion_matrix(y_train,clf.predict(X_train))
    print(cnf_matrix)
```

```
[[ 68  27]
 [ 11 106]]
```

```
(68+106)/(68+27+11+106)
```

```
0.8207547169811321
```

Random Forest Algorithm Result:

```
[ ] m.score(x_test, y_test)
```

```
0.5789473684210527
```

```
[ ] m.score(x_train, y_train)
```

```
0.9074889867841409
```

Table 1 Comparative analysis of Decision Tree, SVM and Random Forest Algorithms

Accuracy	Decision Tree	SVM	Random Forest
Training	**88.98**	**89.09**	**90.74**
Testing	**78.94**	**82.01**	**57.89**

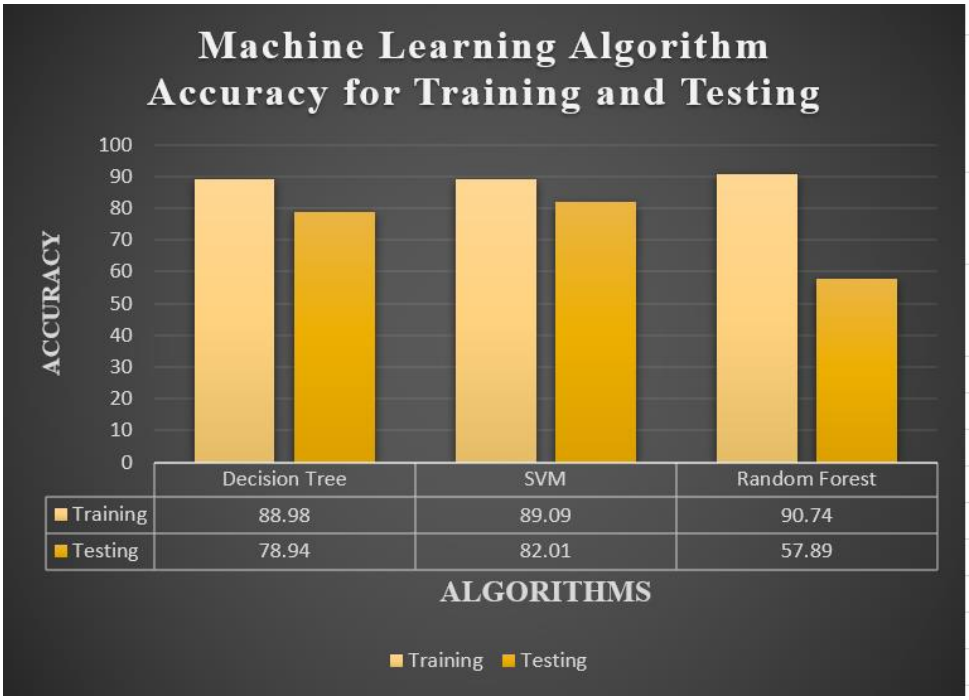

Fig.10 Accuracy analysis of implemented algorithm for training data set and testing data set

4. Conclusion and Future Scope

In these days machine learning playing a vital role in healthcare system to diagnose the various severe diseases precisely. There is various machine learning algorithm available under the classification of supervised, reinforcement and unsupervised. In our research work, three algorithms are implemented which are listed below:

- SVM
- Decision Tree
- Random Forest

First of all, a reliable data set taken and that data set classified into training data set (70%) and testing data set (30%). After that some prime steps taken out to clean the data so that implemented algorithm can give optimized accuracy. In our data set out of 13 prime features there are 4 most noteworthy features that helped us to predict between a positive & negative diagnosis. These features were ST depression induced by exercise relative to rest (oldpeak), chest pain type (cp), number of major vessels (ca), maximum heart rate achieved (thalach). Out of implemented algorithms SVM give the best accuracy which is 82.01% and followed by decision tree 78.94%. Random forest algorithm gave the minimum accuracy. Accuracy more than 70 % is measured good, but if accuracy is very high than it may be case of over fitting. Therefore, accuracy around 80% is commendable. To prevent heart illness, we can take subsequent activities as listed below:

- Eat a well diet
- Take consistent physical action
- Evade tobacco use
- Check and control your overall cardiovascular risk.

References

1. Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath, "Heart disease prediction using machine learning algorithms", ICCRDA 2020, IOP Conf. Series: Materials Science and Engineering, 1022 (2021) 012072 IOP Publishing DOI:10.1088/1757-899X/1022/1/012072
2. Pranav Motarwar, Ankita Duraphe, G Suganya, M Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning", 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE, 2020, DOI: 10.1109/ic-ETITE47903.2020.242
3. Vijeta Sharma; Shrinkhala Yadav; Manjari Gupta, "Heart Disease Prediction using Machine Learning Techniques", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, 18-19 Dec. 2020, DOI: 10.1109/ICACCCN51052.2020.9362842
4. Atharv Nikam; Sanket Bhandari; Aditya Mhaske; Shamla Mantri, "Cardiovascular Disease Prediction Using Machine Learning Models" 2020 IEEE Pune Section International Conference (PuneCon), IEEE, 16-18 Dec. 2020, DOI: 10.1109/PuneCon50868.2020.9362367
5. Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, Special Section On Smart Caching, Communications, Computing And Cybersecurity For Information-Centric Internet Of Things, DOI: 10.1109/ACCESS.2019.2923707
6. Dinesh Kumar G, Santhosh Kumar D, Arumugaraj K, Mareeswari V, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms", Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India
7. Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar, "Prediction of Heart Disease Using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), IEEE Conference, IEEE Xplore ISBN:978-1-5386-0965-1
8. Zhang, Shuai, Y.-L. S. A. (2017), 'Deep learning based recommender system: a survey and new perspectives', Journal of ACM Computing Surveys 1(1), 1–35.
9. AminKhatami, AbbasKhosravi, C. L. (2017), 'Medical image analysis using wavelet transform and deep belief networks', Journal of Expert Systems With Applications 3(4), 190–198.
10. AshwiniShetty, Naik, C. (2016), 'Different data mining approaches for predicting heart disease', International journal of innovative research in science, engineering and technology 3(2), 277–281.
11. Aydin, S. (2016), 'Comparison and evaluation data mining techniques in the diagnosis of heart disease', Indian journal of science and technology 6(1), 420–423.
12. Bayasi, N. and Tekeste (2016), 'Low-power ECG-based processor for predicting ventricular arrhythmia', Journal of IEEE transactions on very large scale integration systems 24(5), 1962–1974.
13. Berikol, B. and Yildiz (2016), 'Diagnosis of acute coronary syndrome with a support vector machine', Journal of Medical System 40(4), 11–18.
14. Zhiyong Wang, Xinfeng Liu, J. G. (2016), 'Identification of metabolic biomarkers in patients with type-2 diabetic coronary heart diseases based on metabolomic approach', 6(30), 435–439.
15. Singh, M. and Martins (2016), 'Building a cardiovascular disease predictive model using structural equation model and fuzzy cognitive map', Journal of fuzzy system 02(6), 1377–1382.