Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2022)
IEEE Xplore Part Number: CFP22N67-ART; ISBN: 978-1-6654-9707-7

Heart Attack Prediction using Machine Learning

Janaranjani N[1], Divya P[2], Madhukiruba E[3], Dr. R. Santhosh[4], R. Reshma[5], D. Selvapandian[6]
UG Scholar, Department of Computer Science and Engineering[1, 2, 3]
Professor, Department of Computer Science and Engineering[4]
Assistant Professor, Department of Computer Science and Engineering,[5,6]
Karpagam Academy of Higher Education, Coimbatore. [1, 2,3 ,4,5,6]
Corresponding Mailid: santhoshrd@gmail.com

2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) | 978-1-6654-9707-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIRCA54612.2022.9985736

ABSTRACT

In the recent years, the coronary artery disease (CAD) becomes leading cause of mortality worldwide both in high-income nations and in poor countries to a rising extent. Studies of human genetics can provide useful information for the development of such enhanced techniques by enhancing our ability to identify those who are at a higher risk for CAD. This work explores artificial intelligence technologies that analyze raw clinical data for the study and assessment of heart disease severity. Further utilizes machine learning algorithms to predict various state of CAD in individuals, such as naïve bayes (NB), decision tree (DT), and Weighted Associative Rule Mining (WARM). This research study has tested these algorithms on the clinical data obtained from the Cardiology department, which include highlights and distinct cases. The decision tree achieves the highest prediction accuracy of 99.5 percent with the fastest rate of exactness over other algorithms.

1. CONCEPT OF DATA MINING

Data mining makes use of huge volumes of data to allow substantial information to be recovered through previously unrecognized patterns, has been steadily employed in healthcare to support clinical diagnosis and illness forecasts [1]. Data mining ideas can even analyze and classify massive amounts of information, combine variables with similar behaviors, and predict future occurrences, among other benefits for monitoring and managing health systems that attempt to protect patients' privacy [2]. The insights gained from these approaches can enhance resource management, patient care, infection control, and risk stratification. Numerous digital healthcare researches have addressed data mining approaches to forecast incidence and features of patients in pandemic scenarios, identification of depressive symptoms, prediction of cancer, diabetes, situations in emergency rooms, many others. Data mining in health organizations improves service delivery, decision making, and lowers subjectivity and mistakes.

1.1. MACHINE LEARNING

There is an abundance of data in the digital world today, thanks to the Fourth Industrial Revolution, including Internet of Things (IoT) data, cyber security data, mobile business data, social media health information, and so on [3]. Expertise in artificial intelligence (AI), specifically machine learning (ML) becomes essential for intelligently analyzing these data and developing smart and automated applications. There are many different types of machine learning algorithms, including supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. As a component of the machine learning family, deep learning can effectively evaluate vast amounts of data [4]. Many ML algorithms offers systems with the capacity to learn and improve automatically without even being coded.

1.2. CARDIAC DISEASE

Ailments that fall under the coronary disease umbrella include, among other things, coronary artery pollution, heart musicality troubles (arrhythmias), and heart forsakes you're born with (intrinsic heart deserts). The phrase "coronary ailment" is frequently used in conjunction with the phrase "cardiovascular disease [5]." Cardiovascular disease refers to disorders that include constricted or obstructed veins. Other heart disorders, such as those that affect the muscle, valves, or thump of your heart, are also considered types of coronary sickness. CAD is a condition that affects the veins that provide blood to the heart muscle. Blood cannot flow normally via these veins if they become too small or blocked. Because the cardiac muscle receives less blood.

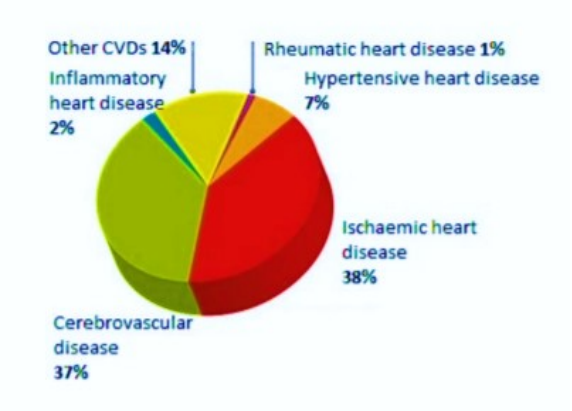

Fig. 1: Cardiac Disease Ratio

These problems harm the veins' coverings, causing them to become constricted or completely blocked. Congestive Heart Failure (CHF) is a condition in which the heart does not pump blood at its normal rate. The two most common reasons are a heart muscle that is either damaged or diseased and faulty heart valves. It is possible that the valves cannot allow enough blood to pass through due to the significant restriction they put on blood flow. The valve, on the other hand, may "open," which would make it possible for the circulation system inside the heart to "switch" (go off track). Whenever the heart and lungs do not function as they should, the heart muscle has been forced to work harder, which can cause it to get tired [6].

1.3. BASICS OF FEATURE SELECTION

Feature selection or choice of features is a method of reducing the number of data variables used in the development of a forecasting model. It is tempting to limit the number of data variables to measure the computing cost of exhibiting and, every now and then, to work on the model's development. AI-based classification techniques entail estimating the relationship between each data variable and the genuine variable and selecting the data factors with the most solid link with the objective variable [7]. These procedures can be quick and efficient, despite the fact that the choice of verifiable metrics is dependent on the data type and yield factors. Regulatory and independent feature decision techniques are the most common, and supervision procedures are classified as covering, channel, and regular. Reliable and accurate measures are put to use in channel-based component assurance techniques to score the correlation or interdependence between input elements. This score is then filtered or choose the characteristics that are the most pertinent to the problem [8]. The data type of the information variable, and the data type of the yield or response variable, should both be taken into consideration when selecting accurate and reliable integrated decision measures.

1.4. MODELS FOR PREDICTION

Prediction models are developed and used to forecast outcomes in various domains based on historical data. Predictive analysis or prediction models are popular statistical techniques for forecasting future behaviour or actions. These models are a type of statistical technology that analyses real - time data to generate a model that helps predict future outcomes. For example, forecasting models are typically used after the wrong behavior has already happened to

recognize infringement and perceive suspects [9]. A significant portion of the time, the model being selected on the basis of area theory to compute the likelihood of a result given a specific quantity of data. An example of this would be analyzing an email to determine the degree to which it is likely to be spam. To determine the likelihood of a large amount of data sharing a location with another, models can employ up to one classifier.

Atherosclerosis is a disease that affects the coronary veins, stroke, and the peripheral supply system [10]. Untreated pharynx could lead to rheumatic coronary illness. Up to 90% of CVD can be prevented through prediction analytics, according to studies. Physically, prevention of CVD includes avoiding tobacco smoke, exercising, and limiting alcohol use. Technically, it can be prevented and diagnosed by data analytics approaches. It is also effective to treat risk factors such as hypertension, blood lipids, and diabetes [11].

2. LITERATURE STUDY

Clinical frameworks in use today, such as medical clinic board frameworks and dynamic frameworks, are focused on collecting and processing all clinical data. Since the variables such as outer environment circumstances, test findings, clinical history, current pandemics, and other elements are required for producing a probabilistic assessment, the clinical information cannot be easily broken down. Authors in [12] recently developed the C4.5 model. An information mining computation was developed based on this new metric to investigate the causal relationship between medications and the risk of coronary artery disease. Statistical calculation has the potential to help disseminate the word that heart disease is among the most common adverse effects in the public health.

Many authors claimed that ML algorithms offers a wide range of potential applications in medicine and can be extremely beneficial [13]. Cardiovascular diseases account for around 33% of all global deaths. Is ML involved in cardiology, and if so, what is the state of the art? To respond to this question, they give a comprehensive assessment that focuses on 1) identifying areas where AI calculations have been used in cardiology, and 2) providing an overview in light of recent work on best-in-class ML calculations used in cardiology. Amin Ul Haq et al. claimed that detecting heart disease (HD) using ML models is very appealing in its early stages [14]. When the disease is diagnosed in its early stages, the HD therapy and recovery are effective. To assist doctors, HD distinguishing proof using ML methodologies was developed. They suggested a cardiac health risk identification framework based on ML models to categorize various health diseases and solid subjects in this focus on. Consecutive in reverse component calculation was used to select more appropriate highlights in order to improve grouping precision and reduce the computing time of the predictive framework. The framework was evaluated using the Cleveland coronary sickness dataset. 70 percent of the dataset was used in the preparation and approval process. The experiment findings reveal that using Sequential Backward Selection (SBS) algorithms and K-Nearest Neighbor controlled AI classifier could pick appropriate highlights and these elements could increase the exactness. This study's high precision suggests that the suggested approach could correctly distinguish between HD and sound participants.

Many authors argued that early detection of heart disease is critical for recovery and treatment [15]. Current methods for detecting heart disease in its early phases are under progress in terms of its performance. As a result, health specialists and academics have a tough time understanding the fundamental causes of breast disease. In this work, we proposed an AI-based classification system that will

miraculously sort the harmful and harmless individuals. To risk free life through the design of digital technological strategy, this work has been used Minimal Redundancy Maximal Relevance and Chi-square to develop the strategy's grouping exhibits.

3. PROPOSED SYSTEM

The patient record of coronary illness is utilized as the information/data to be observed. The primary target of our venture is to arrange the informational index utilizing the WARM calculation. The expectation is performed from mining the patient's verifiable informational collection. In Weighted Associative rule mining (WARM), DT, NB various loads are allotted to various characteristics as per their foreseeing ability. It has been demonstrated that the choice tree classifiers are performing great than conventional classifiers approaches. Further from trial results it has been observed that WARM is giving better accuracy as contrast with other previously existing Associative Classifiers. Thus, the framework is utilizing WARM, DT, NB as a procedure to create rule base. The following shows the model design architecture of both user module and admin module of the proposed system.

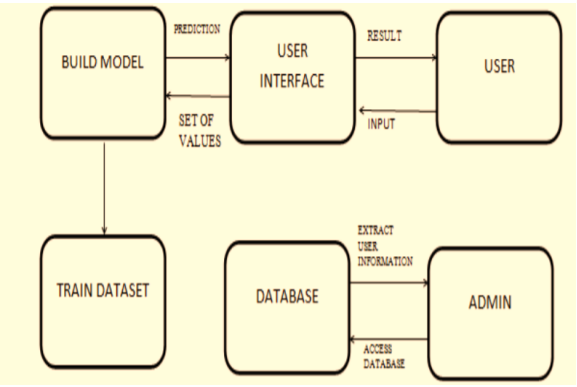

Fig 2: Model Design Architecture

3.1. PRE-PROCESSING

The missing attributes are filled up with appropriate attributes. The performance of the classifier is unaffected by the patient's case ID. After that, it is deleted, and the outcome characteristic is used to characterize the objective or ward variable in this way, lowering the number of capabilities in the list. The algorithmic methodologies utilized for importance examination and characterization are discussed in the following sections.

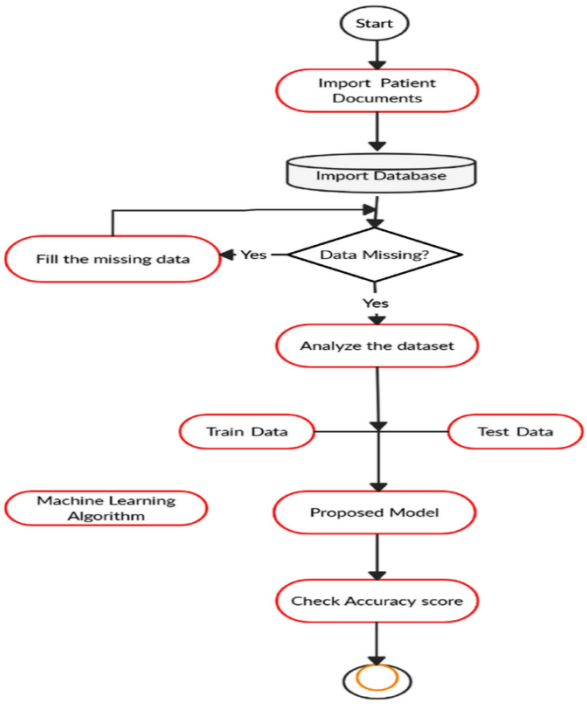

Fig 3: Flow of Model Implementation

3.2. ALGORITHM IMPLEMENTATION

We carry out the new classification approach that utilization affiliation rule mining and characterization which turns into a critical factor for information disclosure. A significant benefit of these classification frameworks is that, hey can look at a few highlights all at once using WARM. Numerous applications can benefit from great classification model in real time scenarios. ML classifiers are particularly fit to applications where the model might help the area specialists in their choices. There are numerous areas, for example, clinical, where the most extreme exactness of the model is wanted and consequently the precision of the respective classifiers. The expectation result is additionally

recognized in the WARM, DT classifier makes the order model by building a decision tree. Every hub in the tree indicates a test on a trait, each branch plummeting from that hub relates to one.

3.3. TREE VIEW AND ACCURACY

The pace of proper commands, either for a free test set or using some variation of the cross-approval concept, is what characterization precision is all about. The characterized exactness result is determined, and the output is displayed in the tree design with the quantity of significant vessels, segmenting into present or absent occurrences, and distinguishing between accurately and incorrectly ordered.

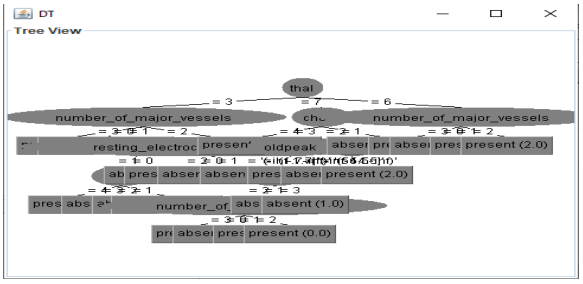

Fig 3.1: Flow of Model Implementation

4. RESULT AND DISCUSSIONS

Supervised learning algorithms such as Naive bayes, Decision tree, WARM (weighted associated rule mining) are then used to model cad cases.

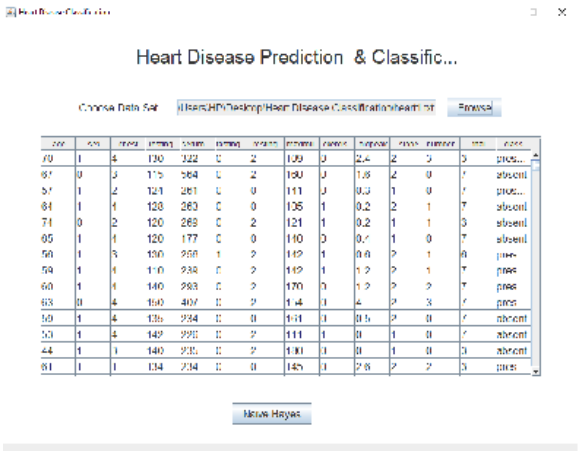

Fig 4: Choosing data sets

All of the patient's information is put into the system for effective and fast monitoring. An analysis of the medical data becomes difficult because a probabilistic rating requires information which could not be included in the report, including but not limited to symptoms and test results and information on current epidemics and past medical history and the state of the external climate.

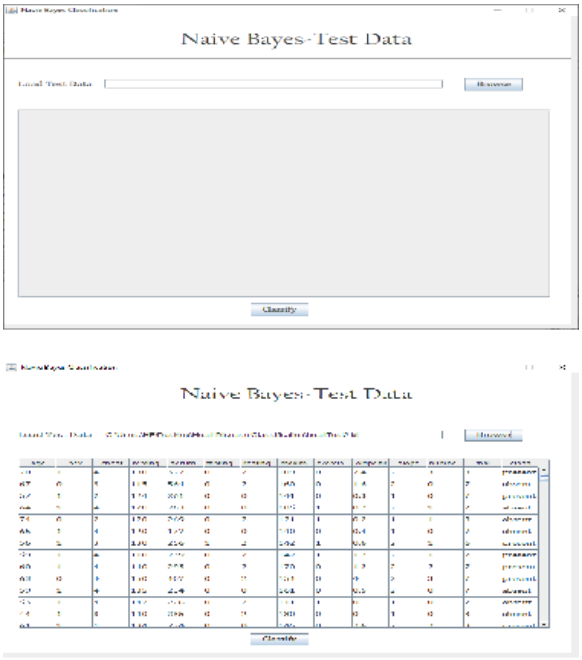

Fig 4.1: Naive Bayes Classification

The Naive Bayes classification algorithm can be

used to provide decision assistance in heart disease prediction systems due to its highest performance compared to the other algorithms. From a historical database of cardiac illness, the system retrieves valuable information. Patients with heart disease can be diagnosed using this predictive approach which lead to healthy life style. A high-dimensional training dataset is the primary use case for this algorithm.

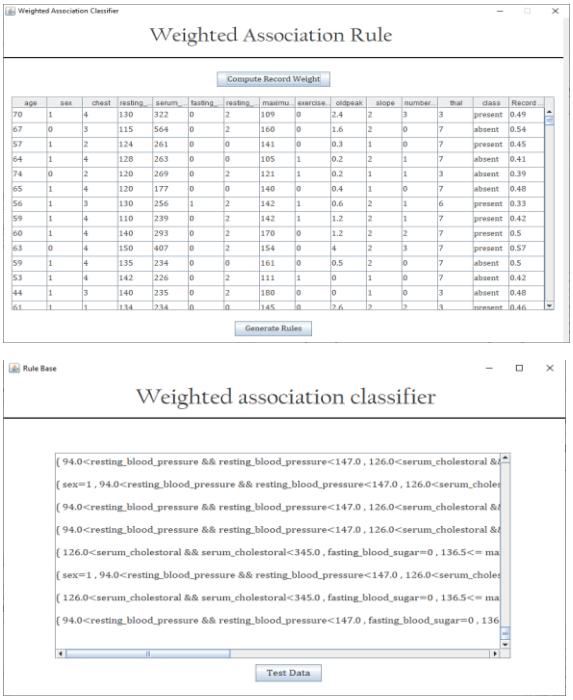

Fig 4.2: Weighted Association Classification

The prediction is performed from mining the patient's historical data set. In Weighted Associative rule mining (WARM), DT, NB different weights are assigned to different attributes according to their predicting capability. Hence the system is using WARM, DT, NB as a technique to generate rule base. The system is expandable for the heart disease data set.

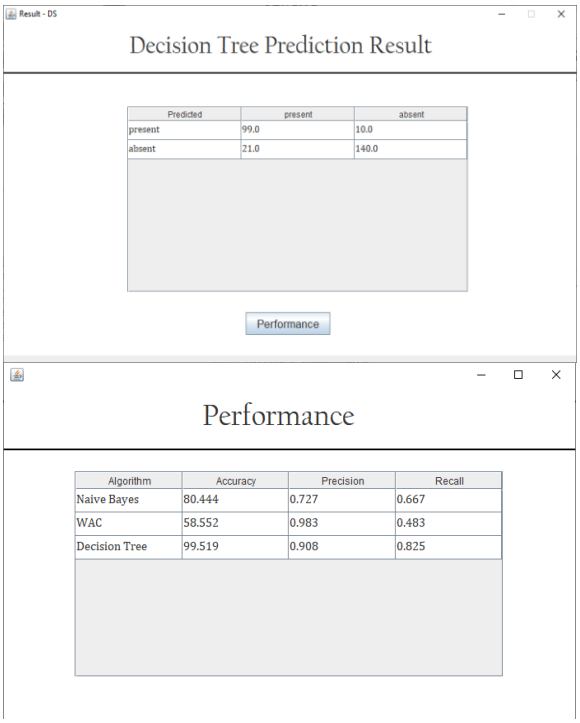

Fig 4.3: Decision Tree Classification

Decision tree produces the highest rate of accuracy achieves highest prediction accuracy of 99.5%. We tested this approach on bench marked heart disease data set as well. In this case also, decision tree algorithm outperforms other techniques. Decision tree frame work provides the best accuracy among all the others. Using decision trees for predictive modeling seems to be a popular choice since they are easy to comprehend and powerful. A decision tree's primary objective would be to divide a large set of data into smaller groups.

5. CONCLUSION AND FUTURE WORK

Clinical analytics is a large field of study that aids in detecting the onset of a cardiac illness. The work revealed the fundamental coronary ailment close to the course of action of most probable heart diseases have identical incidental effects. This research examines predictive analytics that evaluate raw clinical data for the investigation and evaluation of heart disease severity using machine learning algorithms. We employed machine learning techniques like naïve bayes (NB), decision tree (DT), and

Weighted Associative Rule Mining (WARM) to predict different stage of CAD in people. We evaluated these algorithms using clinical data from the Cardiology department, which contained highlights and unique instances. This research concluded that the decision tree could ensure the highest prediction accuracy of 99.5 percent with the fastest rate of exactness over other algorithms.

REFERENCES

1. S. I. Ansarullah and P. Kumar, "A precise writing survey on cardiovascular confusion ID utilizing information mining and AI strategy," Int. J. Ongoing Technol. Eng., vol. 7, no. 6S, pp. 1009-1015, 2019.
2. Batalik, Ladislav, Filip Dosbaba, Martin Hartman, Vladimir Konecny, Katerina Batalikova, and Jindrich Spinar. "Long-term exercise effects after cardiac telerehabilitation in patients with coronary artery disease: 1-year follow-up results of the randomized study." Eur. j. phys. rehabil. med.(Testo stamp.) (2021).
3. Chen, Joy Iong Zong, and P. Hengjinda. "Early prediction of coronary artery disease (cad) by machine learning method-a comparative study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.
4. Kim, Hyun Su, and Hee Joung Choi. "N-terminal pro-B-type natriuretic peptide levels in children: comparison in cardiac and non-cardiac diseases." Cardiology in the Young 30, no. 4 (2020): 500-504.
5. Ueda, Yu, JoAnn L. Yee, Amber Williams, Jeffrey A. Roberts, Kari L. Christe, and Joshua A. Stern. "Identifying cardiac diseases using cardiac biomarkers in rhesus macaques (Macaca mulatta)." Comparative medicine 70, no. 5 (2020): 348-357.
6. Jung, Sunhee, Eunyong Ahn, Sang Baek Koh, Sang-Hak Lee, and Geum-Sook Hwang. "Purine metabolite-based machine learning models for risk prediction, prognosis, and diagnosis of coronary artery disease." Biomedicine & Pharmacotherapy 139 (2021): 111621.
7. Kamiutsuri, K., A. Tsujikawa, and S. Kobayashi. "Cesarean delivery complicated by acute heart failure: myotonic dystrophy, peripartum cardiomyopathy or cardiac disease associated with myotonic dystrophy?." International Journal of Obstetric Anesthesia 46 (2021).
8. A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Ali, "Intelligent AI approach for viable acknowledgment of diabetes in E-medical care utilizing clinical information," Sensors, vol. 20, no. 9, p. 2649, May 2020
9. U. Haq, J. Li, M. H. Memon, J. Khan, and S. U. Commotion, "An original incorporated conclusion strategy for bosom malignant growth location," J. Intell. Fluffy Syst., vol. 38, no. 2, pp. 2383-2398, 2020.
10. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective coronary illness expectation utilizing mixture AI methods," IEEE Access, vol. 7, pp. 81542-81554, 2019.
11. G. G. N. Geweid and M. A. Abdallah, "another programmed ID strategy for cardiovascular breakdown utilizing further developed help vector machine in light of duality improvement procedure," IEEE Access, vol. 7, pp. 149595-149611, 2019.
12. A. Ul Haq, J. Li, Z. Ali, J. Khan, M. H. Memon, M. Abbas, and S. Nazir, "Recognition of the Parkinson's illness utilizing a mixture highlight choice methodology," J. Intell. Fluffy Syst., vol. 39, pp. 1-21, May 2020, doi: 10.3233/JIFS-200075
13. R. Sivaranjani, V. S. Naresh, and N. V. Murthy, "4 coronary illness forecast utilizing hereditary calculation based choice tree," Intell. Decis. Support Syst., Appl. Signal Process., vol. 4, p. 71, Oct. 2019.
14. A. U. Haq, J. Li, J. Khan, M. H. Memon, S. Parveen, M. F. Raji, W. Akbar, T. Ahmad, S. Ullah, L. Shoista, and H. N. Monday, "Identifying the prescient capacity of AI classifiers for planning coronary illness recognition framework," in Proc. sixteenth Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process., Dec. 2019, pp. 130-138.
15. Li, Jian Ping, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. "Heart disease identification method using machine learning classification in e-healthcare." IEEE Access 8 (2020): 107562-107582.