Survival Prediction of Heart Failure Patients Using Lasso Algorithm and Gaussian Naive Bayes Classifier

Velmurugan Sathya Narayanan, Kavin N Raj, Kishore Kumar and Manoj Kumar

October 22, 2021

SURVIVAL PREDICTION OF HEART FAILURE PATIENTS USING LASSO ALGORITHM AND GAUSSIAN NAIVE BAYES CLASSIFIER

V.Sathya Narayanan[1], Kavinraj N[2], Kishore Kumar S[3], Manoj Kumar K[4]

[1]Assitant Professor

[2,3,4] U.G students

Electronics and Communication Engineering,

Kongu Engineering College, Erode,

Tamil Nadu.

sathyanarayanan@kongu.ac.in[1],kavinrajn.17ece@kongu.edu[2],

kishorekumars.17ece@kongu.edu[3] , manojkumark.17ece@kongu.edu[4]

ABSTRACT

Cardiovascular diseases kill approximately 17 million people globally per annum , and that they mainly exhibit as myocardial infarctions and heart failures. Heart failure (HF) occurs when heart cannot pump enough blood to satisfy the requirements of the body. Available electronic medical records of patients quantify symptoms, body features, and clinical laboratory test values, which may be wont to perform biostatistics analysis aimed toward highlighting patterns and correlations otherwise undetectable by medical doctors. Health plans must prioritize disease management efforts to scale back hospitalization and mortality rates in heart disease patients. We developed a risk model to predict the 5-year risk of mortality or hospitalization for heart disease among patients at an outsized health maintenance organization. In the existing system, a two-stage classification model to classify patients into lower variability resource user groups by using electronic patient record. There are various statistical tools for classifying patients into lower variability resource user groups. However, the existing system have some limitations. While performing partitioning recursively, it sequences partitioning greedily instead of finding the optimal partitioning sequence. In proposed system, the LASSO algorithm is used to select features and classification using Gaussian Naïve Bayes, and investigate the results. Dimensionality reduction helps in getting much more accurate predictions with the same data set. Lasso and ridge regression with Gaussian Naïve Bayes (GNB) classifiers has given better results in most of the casess

Keywords : Machine Learning, Algorithm, Patient.

1. INTRODUCTION

Surgical patients risk prediction is a routine component in daily care practice in both specific areas (e.g. Approaches wont to determine stroke risk in patients with atrial fibrillation) also as more generally, for instance in identifying patients likely to need hospital admission. Risk stratify-cation of hospital patients for adverse drug events (ADE) can target a population which will ben-edit from interventions aimed to scale back drug- related morbidity, as a sort of personalized medicine. It can support clinicians and hospital pharmacists in patient prioritization to deliver more efficient health care service.

Emphasized that failure to think about risk prediction during a clinical setting may result in poor patient outcomes.

Long waiting lists for elective surgery in Australian hospitals during recent years has driven a nationwide research agenda to enhance the design , management and delivery of health care services. Since operating rooms are the impact on the performance of the hospital, surgery scheduling has been studied by many researchers. The surgery scheduling problem deals with the allocation of operating rooms under uncertain demand in a complex and dynamic hospital environment to optimize the use of resources. Different techniques such as mathematical programming simulation, meta-heuristics and distributed constraint optimization have been proposed to address this problem. However, most current efforts to unravel this problem either make simplifying assumptions (e.g. considering just one department or sort of surgery) or employ simulated data, which make them difficult to use in hospitals.

Hospitals are trying to enhance the use of operating rooms because it affects patient satisfaction, surgery throughput, revenues, and costs. The surgical prediction model which uses post-surgery data often requires high-dimensional data and contains key predictors such as surgical team factors that may not be available during the surgical listing process. The rapid development of medical and surgical subspecialties within the last decade resulted in increasing demands for more ICU beds and provides momentum for its development. In a recent national mortality audit, the shortage of medical care beds has been cited as a serious contributing think about preoperative deaths (mortality in reference to surgery, often defined as death within fortnight of a surgical procedure) within the Ministry of Health hospitals.

A two-phase tree-based classification and ensemble prediction model. The first phase involves feature engineering for predicting the surgeon rank (instead of features of the individual surgeon) by combining operational domain knowledge and a classification method. The second phase encompasses an ensemble approach using Gradient Boosting Machine (GBM). Both techniques are tree-based methodologies to deal with a substantial number of categorical attributes. The modeling approach incorporates key predictors such as patient history, the complexity of the surgery, discipline, surgeon rank, and temporal factors such as moving averages. The advantages of using different predictors while ensuring a parsimonious yet practice feasible prediction model for the operational requirement of a large public hospital. The baseline model of using all predictors (including surgical team, surgical complexity, patient factors, and temporal factors) has a 10% improvement over the Moving Average model. Our final model including only the rank of the primary surgeon and three other groups of things (surgical complexity, patient, and temporal factors) surpasses the other models in terms of prediction performance, and it is viable for both private and subsidized patients at the public hospital.

1.1 Surgical patients risk prediction

Surgical patients risk prediction is a routine component in daily care practice in both specific areas (e.g. Approaches used to determine stroke risk in patients with atrial fibrillation) as well as more generally, for example in identifying patients likely to require hospital admission. Risk stratify-cation of hospital patients for adverse drug events (ADE) can target a population that can ben-edit from interventions aimed to reduce drug-related morbidity, as a form of personalized medicine. It can support clinicians and hospital pharmacists in patient prioritization to deliver more efficient health care service. Emphasized that failure to consider risk prediction in a clinical setting can result in poor patient outcomes.

Long waiting lists for elective surgery in Australian hospitals during recent years has driven a nationwide research agenda to improve the planning, management and delivery of health care services. Since operating rooms are the impact on the performance of the hospital, surgery scheduling has been studied by many researchers. The surgery scheduling problem deals with the allocation of operating rooms under uncertain demand in a complex and dynamic hospital environment to optimize the use of resources. Different techniques such as mathematical programming simulation, meta-heuristics and distributed constraint optimization have been proposed to address this problem.However, most current efforts to solve this problem either make simplifying assumptions (e.g. considering only one department or type of surgery) or employ simulated data, which make them difficult to use in hospitals.

Hospitals have been trying to improve the utilization of operating rooms as it affects patient satisfaction, surgery throughput, revenues, and costs. The surgical prediction model which uses post-surgery data often requires high-dimensional data and contains key predictors such as surgical team factors that may not be available during the surgical listing process. The rapid development of medical and surgical subspecialties in the last decade resulted in increasing demands for more ICU beds and provides momentum for its development. In a recent national mortality audit, the lack of intensive care beds has been cited as a major contributing factor in preoperative deaths (mortality in relation to surgery, often defined as death within two weeks of a surgical procedure) in the Ministry of Health hospitals.

A two-phase tree-based classification and ensemble prediction model. The first phase involves feature engineering for predicting the surgeon rank (instead of features of the individual surgeon) by combining operational domain knowledge and a classification method. The second phase encompasses an ensemble approach using Gradient Boosting Machine (GBM). Both techniques are tree-based methodologies to deal with a substantial number of categorical attributes. The modeling approach incorporates key predictors such as patient history, the complexity of the surgery, discipline, surgeon rank, and temporal factors such as moving averages. The advantages of using different predictors while ensuring a parsimonious yet practice feasible prediction model for the operational requirement of a large public hospital. The baseline model of using all predictors (including surgical team, surgical complexity, patient factors, and temporal factors) has a 10% improvement over the Moving Average model. Our final model which included only the rank of the first surgeon and three other groups of factors (surgical complexity, patient, and temporal factors) surpasses the other models in terms of prediction performance, and it is viable for both private and subsidized patients at the public hospital.

1.2 Length of stay periodization

Evidence supporting the advantages of improved flow has been mounting. For patients, prolonged hospital stays increase the danger of adverse events, like hospital-acquired infections, adverse drug events, poor nutritional levels, and other complications. For hospitals, economic pressures to deliver efficient and accessible care are at unprecedented highs. Healthcare costs as a percentage of gross domestic product (GDP) (17.9% in 2012) have been rising faster than anticipated, and approximately 30–40% of these expenditures have been attributed to "overuse, underuse, misuse, duplication, system failures, unnecessary repetition, poor communication, and inefficiency.

"These factors impede patient flow, prolong patient stays, and increase the value of care per patient. Patient flow, and by association, bed, and capacity management, may be a common focus area for operations management methods applied to healthcare. Discrete-event simulation, optimization, and Lean Six Sigma approaches are applied successfully in various settings to enhance patient flow by either redesigning healthcare delivery processes or more efficiently matching staff and other resources (e.g., operating rooms, medical equipment) to demand.

More recent advancement in patient flow management alternatively focuses on short- term operational decisions. Real-time demand capacity management (RTDC) may be a new method developed by the Institute for Healthcare Improvement that has shown promising but variable results when pilot tested in hospitals. The RTDC process involves 4 steps: 1) predicting capacity, 2) predicting demand, 3) developing an idea , and 4) evaluating the plan. The RTDC process centers around a morning clinician huddle to predict which and the way many patients are going to be discharged that very same day.

A prediction model could be used to rank patients daily—based on their likelihood of being discharged—in order to prioritize the remaining tasks for the most likely patients. Computed Spearman's rank correlation between the remaining LOS for every patient and their respective RRF model-predicted probabilities for every day (trained on the complete data set). Both plots show that the correlations are almost exclusively positive and moderately large, which suggests that the rank of the model predictions was moderately correlated with the particular discharge order. For some days, the model nearly predicted the precise order during which the patients were discharged.

Improving patient flow continues to be a top priority within the acute-care setting, where patients with longer lengths of stay are less satisfied and exposed to the danger of adverse events (e.g., hospital-acquired infections, complications). Hospitals have aligned incentives to improve patient flow because of the rising demand for services and economic pressures to reduce costs and improve resource utilization. The approaches is meant to empower clinicians and hospital administrators with analytical tools to extend their collective efficiency. Hospital staff could prioritize these patients so as to discharge them as early as possible—without negatively affecting their care—so that other patients are often admitted in their place. Supervised machine learning methods could also be wont to rank patients concurrently during a hospital (or specific unit) consistent with their discharge probabilities. Models perform well for prediction or ranking. Alternatively, patients who are presumably to be discharged might not be impacted significantly by an attempt to prioritize their remaining tasks. Therefore, another approach might be to spot a second tier of patients who are only mildly to moderately likely to be discharged. Prioritizing the remaining tasks for these patients may have a more significant global impact on the amount of patients discharged over a given period of time .

1.3 Health surgery patients

Operative procedures frequently provide dramatic health benefits or essential diagnostic information. While late cancellation of surgery is infrequent, nationally, the absolute number of canceled cases is high, making this a leading source of perioperative wastage. These compensated cases represent unutilized health care resources, valued as high as a dollar per second. Also, the negative impact on patients and families of last-minute cancellation is substantial, leading to a needy patient and family experience. To investigate whether factors can be identified that significantly affect hospital length of stay from those available in an electronic patient record system, using primary total knee replacements as an example. To investigate whether a model can be produced to predict the length of stay based on these factors to help resource planning and patient expectations on their length of stay. Moreover, in the case of elective operations, patients' arrival times are scheduled by the hospital administration. The remaining source of variability in the elective patient flow process is the randomness in LoS. Manage a surgical suite more efficiently can predict patients' accurately.

2.LITERATURE SURVEY

Arpaia, P et al - "A Health 4.0 Integrated System for Monitoring and Predicting Patient's Health during Surgical Procedures " (2020): An innovative system architecture for enhancing patient's health monitoring during surgical procedure is proposed. The integrated system consists of a video see-through (VST) headset (worn by anesthetist or nurses or other member of the surgical team involved in the procedure), which shows in real-time a comprehensive set of information on the patient's health status. In particular, the operator can visualize in real-time the patient's vitals acquired from the operating room (OR) equipment and access the electronic medical records.

Mei, X. - "Predicting five-year overall survival in patients with non-small cell lung cancer by relief algorithm and random forests" (2017): Non-small Cell Lung Cancer (NSCLC) is a leading death disease in many countries. Many studies are focusing on exact surgical approaches to treat the disease. The five-year overall survival rate for NSCLC patients is typically predicted by traditional regression models with small samples and data size. In this paper, we introduce machine learning tools with feature selection algorithms and random forests classifier to predict the five-year overall survival rate based on a large database.

Castaldo, R et al - "Fall Prediction in Hypertensive Patients via Short-Term HRV Analysis" (2017): Falls are a major problem of later life having severe consequences on quality of life and a significant burden in occidental countries. Many technological solutions have been proposed to assess the risk or to predict falls and the majority is based on accelerometers and gyroscopes. However, very little was done for identifying first time fallers, which are very difficult to recognize. This paper presents a met model predicting falls using short term Heart Rate Variability (HRV) analysis acquired at the baseline.

Anjomshoa, H. et al "A Two-Stage Model to Predict Surgical Patients' Lengths of Stay from an Electronic Patient Database" (2018)**:** Soaring healthcare costs and the growing demand for services require us to use healthcare resources more efficiently. Randomness in resource requirements makes the care delivery process less efficient. The existing is to reduce the uncertainty in patients' resource requirements, and we achieve that objective by classifying patients into similar resource user groups.

Davalos, R. V. et al - “Post-treatment analysis of irreversible electroporation waveforms delivered to human pancreatic cancer patient” (2019) : Irreversible electroporation (IRE) is a focal ablation therapy that uses high voltage, short electrical pulses to destroy tumor tissue. The success of treatment directly depends on exposure of the entire tumor to a lethal electric field magnitude. However, this exposure is difficult to predict ahead of time and it is challenging for clinicians to determine optimal treatment parameters. One method clinicians rely upon for the cessation of pulse delivery is to monitor the resistance value of the tissue, as the cells within the tissue will undergo changes during electroporation.

Vetrivelan, P. et al - “Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing” (2019): Heart disease prediction modeled using (Partially Observable Markov Decision Process) POMDP is proposed. In emergency, the patient is alerted through the doctor by fog computing. Ambulance sent to the location of patient at critical situations. The doctor gets the data through fog computing yogism. Fog computing in healthcare is a new area, which gains more attraction in research community.

Hamlin, M. R. A. et al - “Identifying symptoms and treatment for heart disease from biomedical literature using text data mining” (2017): As the heart diseases causing the major problem and they became the main cause of death worldwide, because it is very difficult for identifying the disease, based on symptoms, for that we need lots of experience and also knowledge. For knowing and identifying it is taking a lot of time for doctors also, because they have to observe the health condition and also the food habits of the patients, it is being a very long process, that the patient has to go to doctor and take the test and it is being a long process.

Maalmi, K. et al - “Real-time machine learning for early detection of heart disease using big data approach” (2019): Heart disease is the most common cause of global death. So early detection of heart disease and continuous monitoring can reduce the mortality rate. The exponential growth of data from different sources such as wearable sensor devices used in Internet of Things health monitoring, streaming system and others have been generating an enormous amount of data on a continuous basis. The combination of streaming big data analytics and machine learning is a breakthrough technology that can have a significant impact in healthcare field especially early detection of heart disease. The technology can be more powerful and less expensive.

Deep, V. et al - “Heart Disease Prediction using Evolutionary Rule Learning” (2018): In modern society, Heart disease is the noteworthy reason for short life. Large population of people depends on the healthcare system so that they can get accurate result in less time. Large amount of data is produced and collected by the healthcare organization on the daily basis. To get intriguing knowledge, data innovation permits to extract the data through atomization of processes. Weighted Association Rule is a type of data mining technique used to eliminate the manual task which also helps in extracting the data directly from the electronic records.

Pippal, R. S. - “Design of heart disease diagnosis system using fuzzy logic” (2017) et al: In most of the cases, heart disease results in death. Medical diagnosis is a difficult task and most of the time done by experts in domain. The aim of this work is to develop a fuzzy expert system to identify heart disease risk in the patients. There are several factors to analyses the heart disease in the patient and it is not an easier task, which makes the physician's job difficult.

3.EXISTING SYSTEM

To manage patient flow in a heart disease surgical suite efficiently, need to have effective Strategic patient Flow Management (SFM) and Operational patient Flow Management (OFM) policies. SFM deals with long-term decision making, such as designing a Master Surgery Schedule (MSS) whereas OFM focuses on efficient management of patient flow every day. Can improve both SFM and OFM policies if we can predict patients' resource requirements. This is because can schedule elective patients' arrivals according to resource availability. Modeling and analyzing patient flow helps us design an efficient strategic patient flow management policy. Researchers study the patient flow process either by modeling it as a queuing model or by developing a discrete event. An important input for both models, a queuing model, and a simulation model is the length of stay (LoS) distribution of each patient group. Need to have a manageable set of patient groups to review the patient flow process employing a simulation model or an analytical model. In hospitals, heart patients are grouped together according to their surgery type because they share the same resources after surgery. However, when group patients based entirely on their post-operative wards, the variability in their LoS is large, and therefore the average LoS may be a poor predictor of every individual's LoS.

3.1 Disadvantages:

- In the existing system, a two-stage classification model to classify patients into lower variability resource user groups by using electronic patient record.
- There are various statistical tools for classifying patients into lower variability resource user groups. However, the existing system have some limitations.
- While performing partitioning recursively, it sequences partitioning greedily instead of finding the optimal partitioning sequence.

4.PROPOSED SYSTEMS

This method uses machine learning techniques to predict survival of heart failure patients and explores risk factors among different populations of survivors and non-survivors.

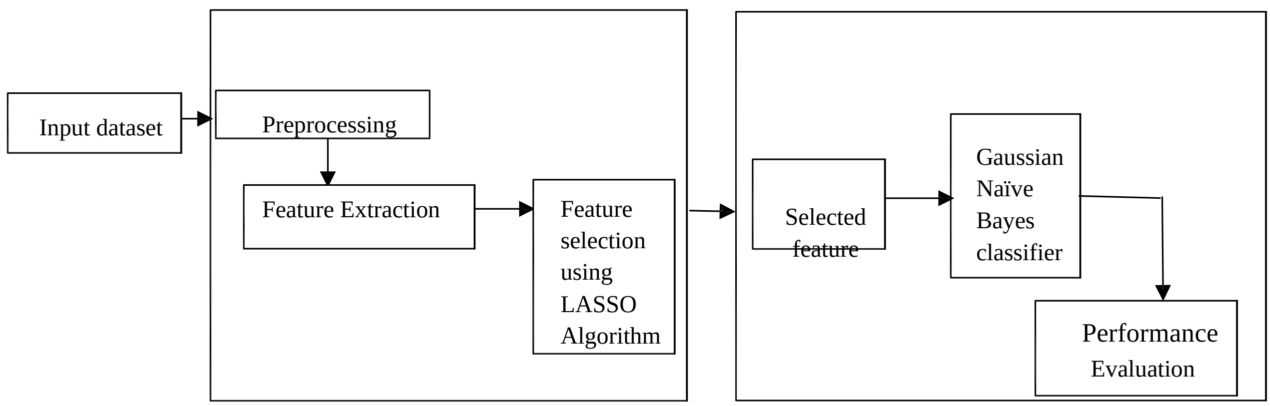

Figure.1. System architecture for Survival prediction

The figure 1 shows the architecture diagram for Survival prediction using Combined Lasso Algorithm with ridge regression. This method consists of input dataset ant its feature extraction and combined lasso algorithm and finally prediction model.

4.1 **Database collection**

The Cleveland heart data set has been used from UCI database. This database contain76 attributes, one of which is predicted. The attributes for patient identification have been dropped (attributes 1, 2, 75 & 76) and 72 attributes have been considered for the survival prediction. The dataset also contains missing values, which have been dropped. After removing the missing values, 115 instances have been considered.

4.2 Feature Extraction Phase

Feature engineer is often mentioned as one of the most important steps in machine learning. It addresses the problem of attaining the most informative and compact set of features to improve the performance of machine learning.

4.2.1 Data Preprocessing

Feature selection is one among the important aspects in data processing . Its necessity is felt thanks to very high dimensionality of knowledge sets and growing computational methodologies of the target problems. Data mining aids in storing huge data and these data is filled with noise i.e. redundant and irrelevant features. Feature selection is the pre-processing step where the noise is filtered, resulting in reducing the dimensionality of the data set and aids in creating computationally effective models with less time and cost.

4.2.2 Survival Prediction using Gaussian Naive Bayes classifier

The Gaussian Naive Bayes classifier in combination with Lasso algorithm for survival of heart disease prediction performs more accurately than the other classifiers like Random Forest, Extra trees, Logistic regression. Dimensionality reduction helps in getting much more accurate predictions with the same data set. Lasso regression with classifiers has given better results in most of the cases.

4.2.3 Advantages

- Lasso regression with classifiers has given better results in most of the cases.
- The noise is filtered, resulting in reducing the dimensionality of the data set.
- This method is computationally effective models with less time and cost.
- Gaussian Naïve Bayes classifier in combination with Lasso algorithm performs more accurately than the other classifiers like Random Forest, Extra trees, Logistic regression.

5.MODULE DESCRIPTION

Modules

1. Database collection details
2. Data preprocessing
3. Feature extraction
4. Feature selection

5.1. DATABASE COLLECTION DETAILS

Dataset collection is a total of 1,214 ICU patients' records and 72 features are extracted from medical records. These features include demographic information, ICU information, surgical information, drug information, and laboratory parameters, and they are selected by experienced doctors and trained nurses.

5.2. DATA PREPROCESSING

In the real world, data is usually incomplete and inconsistent, which cannot be directly used for data mining, or the results are unsatisfactory. In order to enhance the standard of knowledge mining, data preprocessing technology has been developed. There are many methods of knowledge preprocessing: data cleaning, data integration, data transformation, data reduction then on. These processing technologies are used before data processing , which greatly improves the standard of knowledge mining and reduces the time needed for actual mining.

For Preprocessing , we were in need of various medical parameters like Blood Sugar level,Cholesrtol level, Patients' Heartbeat rate and so on as mentioned in Table 1 – Medical Parameters.

1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1

Most studies on ICU LoS prediction have not mentioned the handling of missing values, even though the missing values processing is mentioned, most of them directly exclude patients with missing data.

5.2.2 **Normal distribution transformation**

Most of the numerical ICU data in this study presents skewed distribution, which would affect the final performance of the model. Therefore, Normal distribution technology is once again introduced to this study to correct it. In other words, Normal distribution transformation is performed on all features with skewness coefficient greater than 0.5.

5.2.3 **Normalization**

Considering that the dimension between numerical features (including age, WBC and NEU, etc.) is different and the range of ICU data varies greatly, it is necessary to normalize the original ICU data. The Z-score method is used to normalize the ICU data in this study. In this way, the obtained ICU data conforms to the standard normal distribution.

5.3 FEATURE EXTRACTION

Feature extraction finds new dimensions that are in a combination of original dimensions. This is broadly classified into two techniques. The linear discriminate analysis, etc. for unsupervised learning and mutual information and information theory is used in the compression method for supervised learning. Data mining is broadly classified into two. They are predictive data mining and descriptive data mining. The former constructs one or more models on the data available and develop the prediction model. The latter is used to produce the summary and labels properties of data.

5.4 FEATURE SELECTION

Feature selection retails a subset of best features from an original set of features and the remaining features are discarded. This technique is used to select an optimal feature subset from the original input features according to some criteria. The problem can be solved easily with improved accuracy if the dimension is reduced from the higher dimension to the lower dimension. The feature subset selection techniques are divided into the filter method, wrapper method, and embedded methods. The filter method doesn't use any learning algorithm. It filters out the in-significant features that have little chance to be useful in the analysis of data. It is fast and less expensive. The wrapper method used the learning algorithm, which can be a classifier or clustering algorithm to evaluate the subset of features. Since it utilizes the learning algorithm the accuracy of the method in selecting the relevant features is improved.

6.SYSTEM TESTING

6.1 TESTING PRINCIPLES

Before applying method to design effective test cases, a software engineer must understand the basic principles that guide software testing. Davis (DAV95) suggests a set of testing principles which have been adapted for use in this book.

- All tests should be traceable to customer requirements.
- Test should be planned long before testing begins.
- Test pare to principle applets to software testing. Testing should begin “in the small” and progress towards testing “in the page”.
- Exhaustive testing is not possible.

6.1.1 Unit Testing

Unit testing focuses on verification errors on the smallest unit of software design- the module. Using the procedural design description as a guide, important control paths are tested to uncover errors within the boundary of the module.The module interface is tested to ensure that the information properly flows into and out of the program unit under test. Boundaries conditions are tested to ensure that the module operates properly at the boundaries established to limit of restrict processing.

6.1.2 Integration Testing

Integration testing is a systematic technique for constructing the program structure while conducting test to uncover errors associated with interfacing. The objective is to take unit tested modules and build a program structure that has been dictated by design.

6.1.3 White Box Testing

White box testing is some time is called glass box testing, is a test case design that uses a control structure of the procedural design to drive the test cases. Using white-box testing methods, the software engineer can drive test cases that

- Guarantee that logical decisions are on the true and false sides
- Exercise all logical decisions are on the true and false sides
- Execute all loops at their boundaries and within their operational bounds

- Exercise internal data structure to assure the validity

6.1.4 Acceptance Testing

Finally when the software is completely built, a series of acceptance tests are conducted to enable the client to validate all requirements. The user conducts these tests rather than the system developer, which can range from informal test drive to a planned and systematical executed series of tests. These acceptance tests are conducted over a period of weeks or months, there by uncovering cumulative errors that might degrade the system order time. In this process alpha testing and beta testing are used to uncover the errors that only the end user seems able to find.

6.1.5 Alpha Testing

The customer conducts the alpha test at the developer’s site. The client notes the errors and usage problems and gives report to the developer. Alpha tests are conducted in a control environment.

6.1.6 Beta Testing

The beta testing is conducted at one or more customer’s sites by the end users of the software. Unlike the alpha testing, the developer is not present. Therefore a beta test is a “live” application of the software in the environment that cannot be developed by the developer. The customer records all the problems encountered during the beta testing and reports these to the developers at regular intervals.

6.1.7 Black Box Testing

Black box testing focuses on the functional requirements of the software. That is black box testing enables the software engineer to drive a set of input conditions that will fully exercise the requirements for a program.

Black box testing is not an alternative for white box testing techniques. Rather, it is a complementary approach that is likely to uncover different class of errors. Black box testing attempts to find errors in the following categories:

- Interface errors.
- Performances in data structures or external database access.
- Performance errors.
- Initialization and termination errors.

- Incorrect or missing functions.

All the above-mentioned errors were checked in the process of black box testing and the bugs Found Were Fixed.

6.1.8 Test Cases

Once source code file has been generated, software must be tested to uncover (and correct) as many errors as possible before delivery to your customer. The goal is to style a series of test cases that have a high likelihood of finding errors. To do so we've techniques provide systematic guidance for designing tests that: (1) exercise the interior logic of software components, and (2) exercise the input and output domains of the program to uncover errors in program function, behavior, and performance. Resource presented during this section address the subsequent topic categories.Software Testing is that the process of confirming the functionality and correctness of software by running it. Software testing is typically performed for one among two reasons:

- Defect detection
- Reliability estimation.

The problem of applying software testing to defect detection is that software can only suggest the presence of flaws, not their absence (unless the testing is exhaustive). The problem of applying software testing to reliability estimation is that the input distribution used for choosing test cases could also be flawed. In both of those cases, the mechanism wont to determine whether program output is correct is usually impossible to develop. Obviously the advantage of the whole software testing process is very hooked in to many various pieces. If any of those parts is faulty, the whole process is compromised.Software is now unique unlike other physical processes where inputs are received and outputs are produced. Where software differs is within the manner during which it fails. Most physical systems fail during a fixed (and reasonably small) set of the way . By contrast, software can fail in many bizarre ways. Detecting all of the various failure modes for software is usually infeasible.

The key to software testing is trying to seek out the myriad of failure modes – something that needs exhaustively testing the code on all possible inputs. For most programs, this is computationally infeasible. It is commonplace to aim to check as many of the syntactic features of the code as possible (within some set of resource constraints) are called white box software testing technique. Techniques that don't consider the code’s structure when test cases are selected are called recorder technique.Functional testing may be a testing process that's recorder in nature. It is aimed toward examine the general functionality of the merchandise . It usually includes testing of all the interfaces and will therefore involve the clients within the process. Final stage of the testing process should be System Testing. This type of test involves examination of the entire computing system , all the software components, all the hard ware components and any interfaces. The whole computer based system is checked not just for validity but also to satisfy the objectives.

6.1.9 System Implementation

Implementation includes all those activities that take place to convert from the old system to the new. The new system may be totally new, replacing an existing system or it may be major modification to the system currently put into use. This system “Access Point Selection for Improving the Voice Quality and Overall Throughput in Wireless LANs” is a new system. Implementation as a whole involves all those tasks that we do for successfully replacing the existing or introduce new software to satisfy the requirement. The test case has performed in all aspect and the system has given correct result in all the cases.

The System implementation phase consists of the following steps:

- Testing the developed software with sample data.
- Correction of any errors if identified.
- Creating the files of the system with actual data.
- Making necessary changes to the system to find out errors.
- Training of user personnel.

The system has been tested with sample data, changes are made to the user requirements and run in parallel with the prevailing system to seek out out the discrepancies. The user has also been appraised the way to run the system during the training period.This phase is primarily concerned with user training, site preparation and file conversions. During the ultimate testing, user acceptance is tested, followed by user training. Depending within the nature of the extensive user training could also be required.

After development and testing has been completed, implementation of the knowledge system can begin. During system implementation, the project team should be brought back to full strength. During software development stage, project teams end to play passive role because the technical steps of program development and testing evolve. However, broad organizational representation, accomplished through the project team, is required to finish the system development cycle.NET Framework has offer very efficient yet simple implementation techniques for development of the project.

6.1.10 Implementation plan

Implementation is that the stage, which is crucial within the life cycle of the new system designed. Implementation means converting a replacement or revised system design into an operational one. This is the stage of the project where the theoretical design is turned into a working system. In this project “Access Point Selection” implementation includes all those activities that happen to convert from the old system to the new one. The important phase of implementation plan is change over.

The implementation phase’s construction, installation and operations lie on the new system. The most crucial and really important stage in achieving a replacement successful system and in giving confidence on the new system for the user that it'll work efficiently and effectively.

There are several activities involved while implementing a project:

- Careful planning
- Investigation current system and its constraints on implementation
- Design of methods to achieve the change over
- Training of the staff in the changeover procedure and evaluation of change over method

The implementation is that the end and it's a crucial phase. It involves the individual programming system testing, user training and therefore the operational running of developed proposed system that constitutes the appliance subsystems. On major task of preparing for implementation is education of users, which might really have taken place much earlier within the project when we’re being involved within the investigation and style work. The implementation phase of software development cares with translating design specifications into ASCII text file . The user tests the developed system and changes are made consistent with their needs.

6.1.11 Change over

The implementation is to be done step by step since testing with dummy data won't always reveal the faults. The system are going to be subjected to the workers to figure . If such error or failure is found, the system are often corrected before it's implemented fully stretch.

The trail should be done as long because the system is formed bound to function with none failure or errors. Precautions should be taken in order that any error if occurred shouldn't totally make the method to a halt. Such a care should be taken. The system are often fully established if it doesn't create any error during the testing period.

6.1.12 Education and user training

Well-designed and technically elegant systems can succeed or fail due to the way they're operated and used. Therefore the standard of the training received by the personnel involved the systems help or hinder, and should even prevent, the successful completion of the system.An analysis of user training focuses on user capabilities and therefore the nature of the system being installed. Those users are verifying type and nature. a number of them might not have any knowledge about the computers and therefore the others could also be very intelligent. the wants of the system also range from simple to complex tasks. therefore the training has got to be generated to the precise user supported his/her capabilities and system’s complexity.

User training must instruct individuals in trouble shooting the system, determining whether a drag that arises is caused by hardware or software. an honest or perfect documentation which instructs the user on the way to start the system and therefore the various functions and meanings of varied codes must be prepared which will help the user to know the system during a better manner. Through the training demonstration with personnel contact also, the user are often trained. This training demonstration will help the users to know the system in some ways .

By this the user receives encouragement and a spotlight . Another rapid way of coaching the user is by resident experts. Several user training aids are provided like user manual.

7.SIMULATION RESULT

7.1 PATTERN RECOGNIZTION NEURAL NETWORK :

There are two states available , one is training state and another is testing state. Figure 7.1.1 shows in hidden layer part Training state is being executed.Here Testing state is much more efficient than the training algorithm due to the implementation of LASSO Algorithm in it.

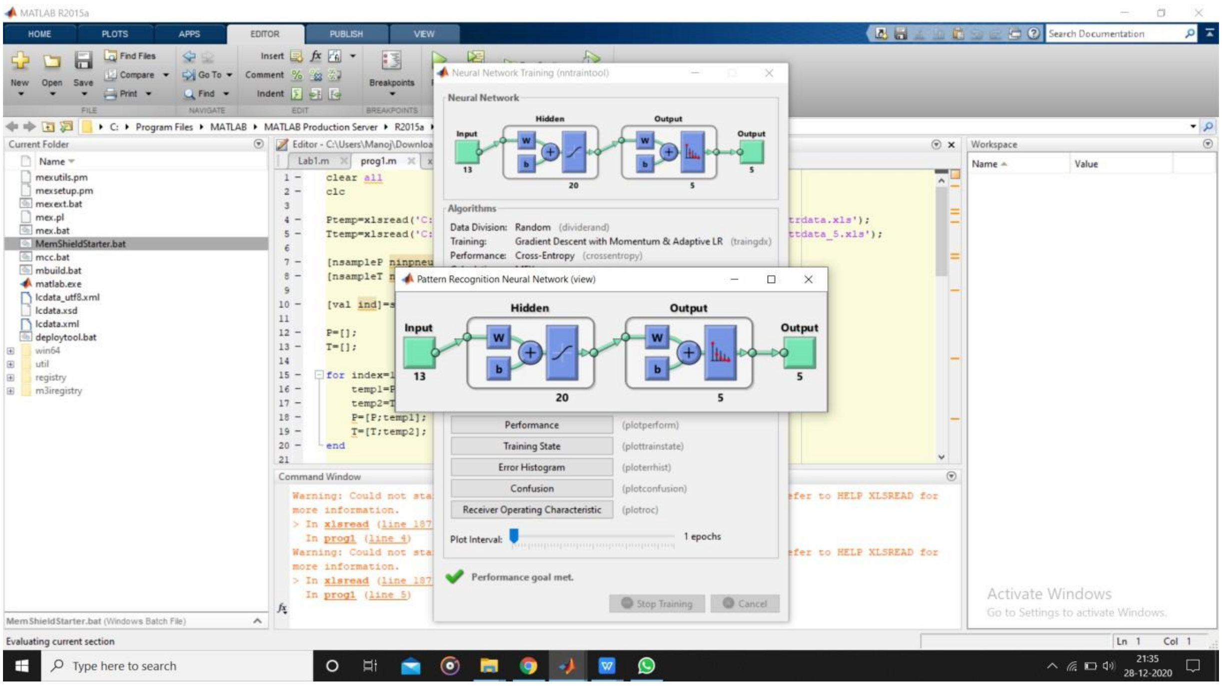

Fig 7.1.1 Pattern Recognizition Neural Network

Predicted class is class i, divided by the amount of outputs whose predicted class is class i. FPR is that the number of outputs whose actual class isn't class i, but predicted class is class i, divided by the amount of outputs whose predicted class is not class i..

7.2 NEURAL NETWORK TRAINING STATE :

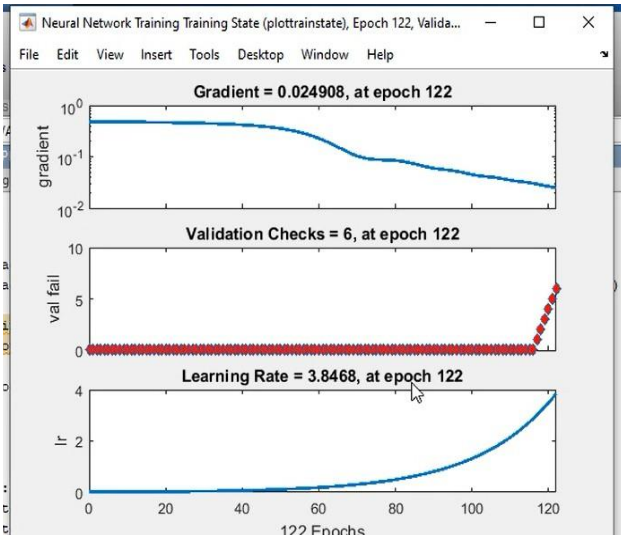

Fig 7.2.1 : Neural Network Training State

Figure 7.2.1 shows the variation in the Gradient,validation check and the learning rate with respect to Epoch

Figure 7.2.2 shows the receiver operating characteristic .The receiver operating characteristic may be a metric wont to check the standard of classifiers. For each class of a classifier, roc applies threshold values across the interval [0, 1] to outputs. For each threshold, two values are calculated, truth Positive Ratio (TPR) and therefore the False Positive Ratio (FPR).For a specific class I,TPR is that the number of

Outputs.

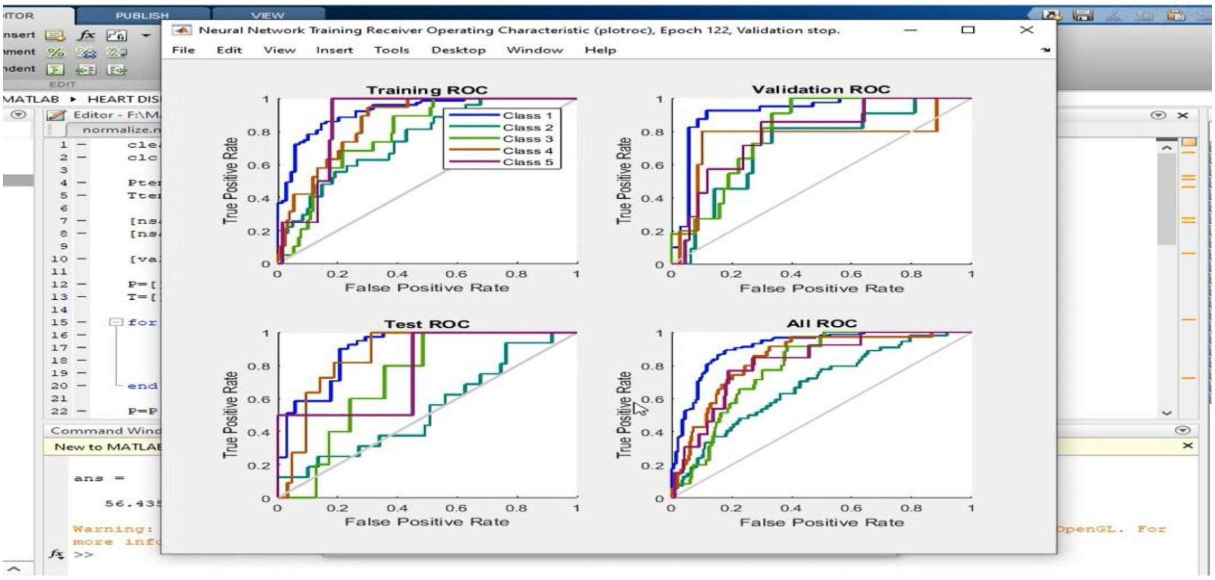

Fig 7.2.2 : Neural Network Training Receiver Operating Characteristic

A true positive is an outcome where the model correctly predicts the positive

class. Similarly, a real negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class

After observing all the five stages, output will be available at workspace

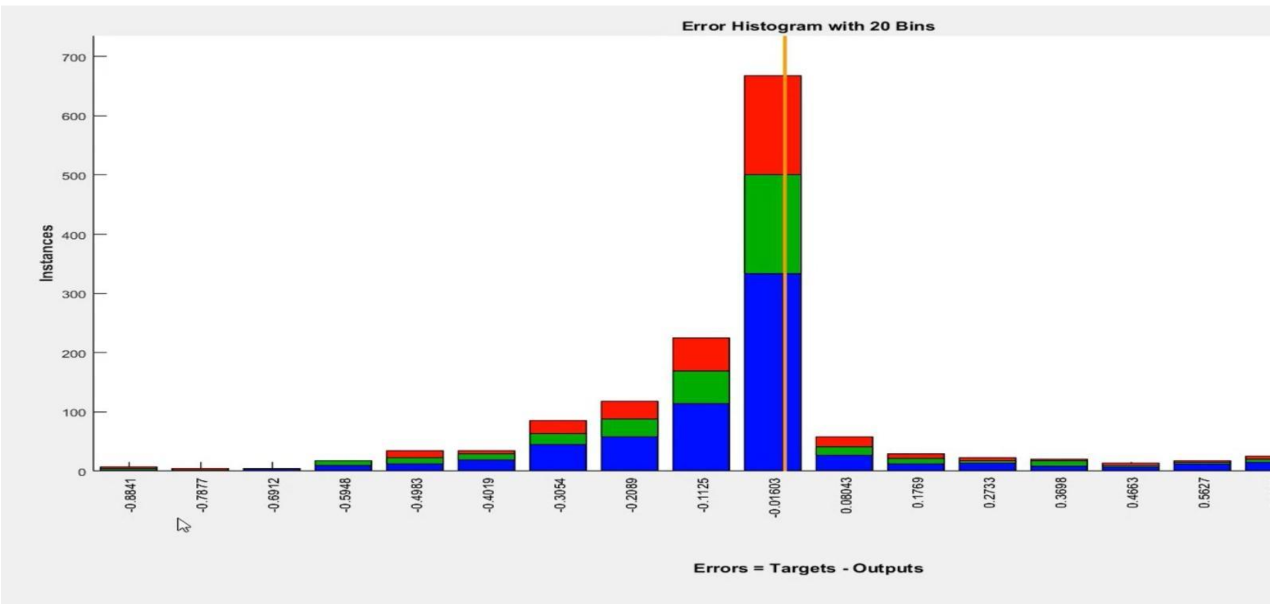

Fig 7.2.3 :Error Histogram

The Figure 7.2.3 depicts histogram of the error occurs between training and predicted value after training.As the error value indicates how predicted values differ from target values.

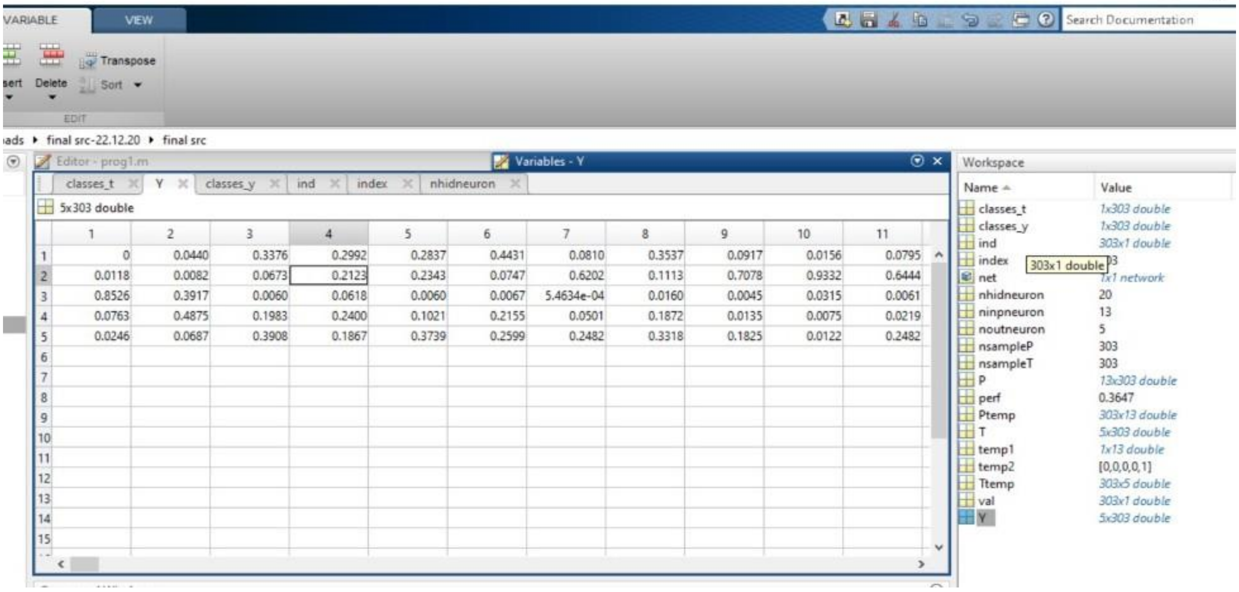

Fig7.2.4

Figure 7.2.4 shows the final results for the entire patients depicted which has an efficiency of 94.92% processed with the help of LASSO Algorithm and Gaussian Naive Bayes Classifier

8.CONCLUSION

The Gaussian Naive Bayes Classifier used in the proposed system shows remarkable results with an accuracy score of 94.92% after using feature selection methods like Lasso and Ridge regression. Dimensionality reduction helps in getting much more accurate predictions with the same data set. Also, we have observed that Lasso regression with classifiers has given better results in most of the cases. The embedded feature selection methods using Lasso regression involves using two sorts of penalty functions. L1 and L2 are loss functions, which help in minimizing the error in regression. In Lasso, the sum of the absolute values of the coefficients is considered as the penalty that is L1. Ridge regression uses L2 wherein the penalty is sum of squares of coefficients of the variables. Lasso basic objective is shrinkage of an absolute value (L1 penalty) towards zero rather than a using sum of squares (L2 penalty). Finally, with extensive experiments and comparisons on the Cleveland dataset, we demonstrated that our proposed Gaussian Naive Bayes classifier has a superior performance in survival prediction of heart disease. Future work can be carried out in combining optimization techniques with feature selection to study the impact on various classifiers and providing a cost and time effective approach for prediction of heart disease using non- invasive methods

9. REFERENCE

Reference

1. Kumar, A., & Anjomshoa, H. (2018). A Two-Stage Model to Predict Surgical Patients' Lengths of Stay from an Electronic Patient Database. IEEE Journal of Biomedical and Health Informatics, 1–1.
2. Arpaia, P., Cicatiello, M., Benedetto, E. D., Anna Dodaro, C., Duraccio, L., Servillo, G., & Vargas, M. (2020). A Health 4.0 Integrated System for Monitoring and Predicting Patient's Health During Surgical Procedures. 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).
3. Mei, X. (2017). Predicting five-year overall survival in patients with non-small cell lung cancer by reliefF algorithm and random forests. 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).
4. Castaldo, R., Melillo, P., Izzo, R., De Luca, N., & Pecchia, L. (2017). Fall Prediction in Hypertensive Patients via Short-Term HRV Analysis. IEEE Journal of Biomedical and Health Informatics, 21(2), 399–406.
5. Beitel-White, N., Martin, R. C. G., & Davalos, R. V. (2019). Post-treatment analysis of irreversible electroporation waveforms delivered to human pancreatic cancer patients. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
6. Latha, R., & Vetrivelan, P. (2019). Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing. 2019 11th International Conference on Communication Systems & Networks (COMSNETS).
7. Sudeshna, P., Bhanumathi, S., & Hamlin, M. R. A. (2017). Identifying symptoms and treatment for heart disease from biomedical literature using text data mining. 2017International Conference on Computation of Power, Energy Information andCommunication (ICCPEIC).
8. Ed-Daoudy, A., & Maalmi, K. (2019). Real-time machine learning for early detection of heart disease using big data approach. 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS).
9. Chauhan, A., Jain, A., Sharma, P., & Deep, V. (2018). Heart Disease Prediction using Evolutionary Rule Learning. 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT).
10. Qidwai, U. (2018). Fuzzy Data to Crisp Estimates: Helping the Neurosurgeon Making Better Treatment Choices for Stroke Patients. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES).
11. Kasbe, T., & Pippal, R. S. (2017). Design of heart disease diagnosis system using fuzzy logic. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).

12. Meena, G., Chauhan, P. S., & Choudhary, R. R. (2017). Empirical Study on Classification of Heart Disease Dataset-its Prediction and Mining. 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC).

13. Pawlovsky, A. P. (2018). An ensemble based on distances for a kNN method for heart disease diagnosis. 2018 International Conference on Electronics, Information, and Communication (ICEIC).

14. Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, P. K. (2018). Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).

15. Raju, C., Philipsy, E., Chacko, S., Padma Suresh, L., & Deepa Rajan, S. (2018). A Survey on Predicting Heart Disease using Data Mining Techniques. 2018 Conference on Emerging Devices and Smart Systems (ICEDSS).