IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Prediction of Heart Disease Using Machine Learning Algorithms

Nukala V V Pravallika[1*], P Suresh Varma[2]

[1,2]Department of Computer Science and Technology, Adikavi Nannaya University College of Engineering, Rajahmundry, India

Corresponding Author: nvvpravallika@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i3.149155 | Available online at: www.ijcseonline.org

Accepted: 26/Mar/2019, Published: 31/Mar/2019

***Abstract*—** Heart-related disease or cardiovascular disease is the main reason for a huge number of deaths in the world. Machine learning techniques help health care professionals in the diagnosis of heart disease. This aims a better understanding and application of machine learning in the medical domain, an automated system in medical diagnosis would enhance medical efficiency and reduce costs. In order to decrease the number of deaths by heart diseases there have to be a quick and efficient detection technique, the use of multiple machine learning algorithms for heart disease, models based on supervised learning algorithms such as: Decision tree, Neural Networks, Logistic Regression, and then implement them to predict heart disease based on patients' medical records. Find the accuracy of the models, Choose the best output with the highest accuracy. These machine learning algorithms and techniques have been applied to various medical data sets. The implementation of work is done on heart disease data set from the University of California Irvine (UCI) machine learning repository, it contains several instances and attributes. By using the data set we test on different machine learning techniques and predict the best model which is computationally efficient as well as accurate for the prediction of heart disease.

***Keywords*—**Machine Learning, Supervised Learning, Classification Techiniques.

I. INTRODUCTION

Heart diseases have emerged as one of the most prominent causes of death all around the world. According to the latest survey conducted by WHO (World Health Organization), the medical professional able to correctly predicted only 67% of heart disease, so there is a vast scope of research in the area of predicting heart disease in human. May people were killed by heart disease, according to the Global Burden of Disease Report. An estimation carried out relating to the cost of therapy for cardiovascular disease would be $1,044 billion by 2030.

The heart disease occurs in several ways such as if Heart muscles do not have enough oxygenating blood the muscle die leads to heart disease, If the body that does not receive required oxygenating blood leads to heart disease, If heart receives less oxygenating blood content due to plaque deposits that contain cholesterol leads to heart disease, The electrical impulses that trigger an individual heartbeat that does not coordinate then leads to heart disease. This different type of heart disease is named as Coronary heart disease, Cardiomyopathy, Cardiovascular disease, Ischemic heart disease, Heart failure, Hypertensive heart disease, inflammatory heart disease, Valve heart disease. The common risk factors are used in data set of heart disease for the prediction of heart disease: high blood pressure, abnormal blood lipids, use of tobacco, obesity, physical inactivity, diabetes, old age, family generations, smoking, alcohol intake, hypertension, high cholesterol.

There are two major problems in the prediction of heart disease, the first one is much expensive and the second is inefficient to calculation the chance of heart disease in human. There are several tools, software's available which use the prediction algorithms, but they have few flaws, most of the tools cannot handle big data. The hospitals and healthcare industries collect a large amount of heart-related patient data which had become difficult to handle with the currently existing system. An automated system in medical diagnosis would enhance medical efficiency and reduce costs. The prediction of heart disease requires a huge size of data which is too complex and massive to process and analyse by conventional techniques.

From last few years, Machine Learning has an impact full in the Healthcare industry. Some of its applications in health care areas include medical imaging diagnosis, which focuses on computer vision and deep learning. It is a widely used artificial intelligence tool in all major sectors of applications, with advancement in processing the power of machine learning. By this the prediction of diseases has been broadening significantly, the algorithms and the process used in machine learning techniques that improves accuracy in heart disease risk prediction which helps for the patients

finding the heart disease in the initial stage rather than mistreating the patients which would lead to the physical and financial burden for them.

The supervised learning algorithms play a vital role in analysing and deriving hidden knowledge and information from the data sets of heart disease. It improves accuracy, speed and reduces cost. It is one such process widely utilized in different domains because it doesn't require different algorithms for different data set we can use one data set for different algorithms and find the efficient algorithm that fulfils the task given to it. The main objective is to find the suitable machine learning technique that is computationally efficient as well as accurate for the prediction of heart disease and to discuss various supervised learning algorithms for heart disease prediction and their relative comparison on the various parameters. The analysis of data prediction is done well-using machine learning algorithms, which leads to efficient results and an analytical comparison has been done for finding out the best available algorithm for the medical dataset. This prediction system for heart disease helps doctors to predict heart disease, with high processing speed. The main goal of this project was to determine which model worked the best from the three algorithms taken, based on different evaluation criteria.

Rest of the paper is organized as follows, Section I contained the introduction of heart disease types and how it occurrences, Section II contains the related work of previous papers and their algorithms and predictions, Section III contains the data required for the prediction of heart disease, Section IV contain the architecture, methodology and essential steps of execution of the prediction of heart disease using data set taken section V explain the algorithms taken for the classification of heart disease, Section VI contains the comparative study with step by step process undergone in methodology, Section VII contains the results and discussion, Section VIII contains the conclusion and future scope, Section VIII concludes research work with future directions.

II. RELATED WORK

Predicting and diagnosing of heart disease using machine learning algorithms [1] the main objective of this paper is the prediction of heart disease of a patient using machine learning models. This is a comparative study of the various performances of machine learning models is done through a graphical representation of results. Machine learning algorithm plays a crucial role in analysing and deriving hidden knowledge, information from the data sets. It improves accuracy, speed. The algorithms are applied directly to a dataset. WEKA implements algorithms for data pre-processing, feature reduction. It also includes visualization of tools. The models used and their classification rate was Naïve base classifier 83.4983, Support Vector Machine 84.1584, Decision Tree 77.5578, K-Nearest Neighbor 76.2376. The naïve base classifier is the best as compared to Support Vector Machine, Decision Tree, and k-Nearest Neighbor.

Heart disease prediction using machine learning and data mining technique [2] this paper compares different algorithms of Decision Tree classification gives better performance in heart disease diagnosis using WEKA. The models which are tested are the J48 algorithm, Logistic model tree algorithm, and Random Forest algorithm. This chosen algorithm is implemented using the WEKA tool. This paper has specifically on decision tree classifiers for heartbeat prediction within WEKA. J48 algorithm achieved higher sensitivity and accuracy while LMT achieved higher specificity than J48 and Random Forest algorithm. It is concluded that J48 (with Reduced Error Pruning) has got the best overall performance. J48 tree technique turned out to be the best classifier for heart disease prediction because it contains more accuracy. The best calculation J48 dependent on UCI information has the most noteworthy precision i.e., 56.76% and the complete time to fabricate display is 0.04 seconds while LMT calculation has the least exactness i.e., 55.77% and the absolute time to manufacture show is 0.39seconds. Taking everything into account, as distinguished through the writing audit, just a negligible achievement is accomplished in the making of a prescient model for coronary illness patients and subsequently there is a requirement for combinational and progressively complex models to build the exactness of foreseeing the early beginning of coronary illness.

Prediction of heart disease using neural network [3] in this paper, a heart disease prediction system which uses an artificial neural network back-propagation algorithm is proposed. 13 medical features were used as input for the neural network and then the neural network was trained with back-propagation algorithm to predict absence or presence of heart disease with prediction accuracy of 95%. Prediction of heart disease is called supervised learning problem. Because of having output variables are in category type, the prediction heart disease is "classification type of supervised learning". The proposed heart disease prediction system which uses a multi-layer perceptron neural network was developed in MATLAB. The proposed system gives a 95% accuracy rate which means a very good rate according to related studies on this field. As a further study, the proposed methodology can be enhanced as a hybrid model with other classification algorithms in order to obtain a more accurate diagnosis of heart disease.

Human Heart Disease Prediction System using Data Mining Techniques [4] this paper gives the review about various arrangement strategies utilized for foreseeing the hazard dimension of every individual dependent on age, sexual

orientation, Blood weight, cholesterol, beat rate. The patient hazard level is grouped utilizing datamining order strategies, for example, Naïve Bayes, KNN, Decision Tree Algorithm, and Neural Network. And so on. Accuracy of the risk level is high once exploitation a great deal of scope of properties. The coronary illness exactness dimension of the expectation was expanded up to 80.6%. According to the examination mode, it is seen that numerous creators utilize different innovations and diverse number of qualities for their investigation. Henceforth, extraordinary advancements give distinctive exactness relying upon various characteristics considered. Utilizing KNN and ID3 calculation the hazard rate of coronary illness was identified and precision level additionally accommodated diverse number of properties. In future, the quantities of properties could be decreased and exactness would be expanded utilizing some different calculations.

Performance evaluation of different machine learning techniques for prediction of heart disease [5] it is essential to have a framework that can effectually recognize the prevalence of heart dis-ease in thousands of samples instantaneously. At this crossroads, the capability of six AI systems was assessed for expectation of coronary illness. The most noteworthy classification precision of 85 % was accounted for utilizing calculated relapse with affability and specificity of 89 and 81 %, separately. The least accuracy is for decision trees of 77%.

Heart Disease Prediction Using Artificial Neural Networks [6] this paper shows a framework for desire for coronary disease using huge risk components. This technique incorporates two best data mining instruments, neural frameworks and genetic estimations. The crossbreed structure executed usages the overall improvement purpose of inclination of innate computation for in statement of neural framework loads. The learning is snappy, all the more relentless and definite when stood out from back inciting. The system was completed in Mat lab and predicts the threat of coronary disease with an accuracy of 89%.

Early prediction of heart disease using decision tree algorithm [7] the wellbeing area today contains shrouded data that can be vital in deciding. Information mining calculations, for example, choice tree and gullible Bayes are connected in this examination for expectation heart assaults. Datamining empower the wellbeing area to anticipate designs in the informational collection. The datamining apparatus MATLAB is utilized for the test. A disarray lattice is gotten to figure the precision of order. The most astounding exactness is for choice tree of 98% and least precision for credulous Bayes of 89%.

Comparative Analysis of Machine Learning in Heart Disease Prediction by R Language [8] This study compares different machine learning algorithms seeking better performances in heart disease prediction using R tool. The models used are the Logistic Regression Model, Random Forest Tree Model, and Neural Network Model. The efficiency of these models is compared through sensitivity, specificity, and accuracy. The existing data set of Heart Disease patients from Cleveland Database of UCI repository is utilized. Accuracy for logistic regression is 83.22%, Neural Network is 86.58, and Random Forest is 85.91%. We can see the highest accuracy belongs to the Neural Network algorithm followed by the Random Forest algorithm and Logistic Regression. It is also observed that Neural Networks has taken maximum time to construct. Logistic Regression is a fast build model of 10.42millisec. The effective and more perplexing models are used to increase the accuracy of predicting the early onset of heart disease.

III. DATASET DESCRIPTION

UCI heart disease is a CAD data set. The Cleveland dataset from UCI machine learning repository is used hear and this dataset has 76 attributes and 303 records. However, only 13 attributes are used in this study and testing. Since all the published experiments for CAD diagnosis, One decision label is the attribute named as num which indicates whether the risk of heart disease exists or not, in the patient dataset for each patient.

The 13 attributes (categorical, integer, real) are listed as follows:

• Age: The age in years. Hear the age of the patient is taken.

Heart disease is mostly seen in more aged people with age greater than 40, taken as the numerical value.

• Sex: Gender of the patient, 0 indicates female and 1 is for the male.

It is found that male have more heart disease conditions than female

• CP: chest pain type, takes values equal to 1, 2, 3 or 4 indicating typical angina, atypical angina, non- anginal pain and asymptomatic, respectively.

• Trestbps: resting blood pressure.

• Chol: serum cholesterol mg/dl.

If cholesterol is above then there is a chance of heart disease

• FBS: (fasting blood sugar > 120 mg/dl), takes 1 or 0 as yes or no, respectively.

• Resting: resting electrocardiographic results, takes values equal to 0, 1 or 2 indicating: normal, ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) and probable or definite left ventricular hypertrophy by Estes' criteria, respectively. Hear if, resting blood pressure is above 180, then there is a chance of heart disease.

• Thach: maximum heart rate achieved. If maximum heart rate achieved is above 100 then there is a chance of heart attack.

• Exchange: exercise-induced angina takes 1 or 0 as yes or no, respectively.
• Old peak: ST depression induced by exercise relative to rest (based on the ST segment on the electrocardiogram).
• Slope: the slope of the peak exercise ST segment, takes values equal to 1, 2 or 3 indicating upsloping, flat and down-sloping, respectively.
• Ca: value 0-3 indicating the number of major vessels colored by fluoroscopy.
• Thal: takes values equal to 3, 6 or 7 indicating normal, fixed defect and reversible defect, respectively.
• Num: diagnosis label which is an integer constant that can take any value from 0 to 4. 0 indicates non- CAD and values 1, 2, 3, and 4 are for CAD risk. The 1, 2 and 3 show the number of clogged vessels. It is the response variable.

IV. METHODOLOGY

The machine learning application which is trained by a UCI dataset, we input its specific medical details to get the prediction of heart disease. The algorithm will calculate the probability of the presence of heart disease. The result will be the accuracy of each model. Thus minimizes the cost and time required to predict the disease. The necessity of the robust machine learning algorithm is very important that can decrease the noise content which is mostly present in the medical data set, it may also consist of many redundancies of data so this process will avoid such issues and predict the result in high processing speed. The analysis of the data for prediction is done well using supervised learning algorithms, which gives efficient results. An analytical comparison has been done for finding out the best available model for the medical data set. This type of system for prediction of heart disease helps doctors to predict heart disease easily with high processing speed. The comparison of taken three algorithms leads to find the best accuracy from the best model. The following system architecture shows the process of the data that leads to find the accuracy and choose the best fit model for further implementation.

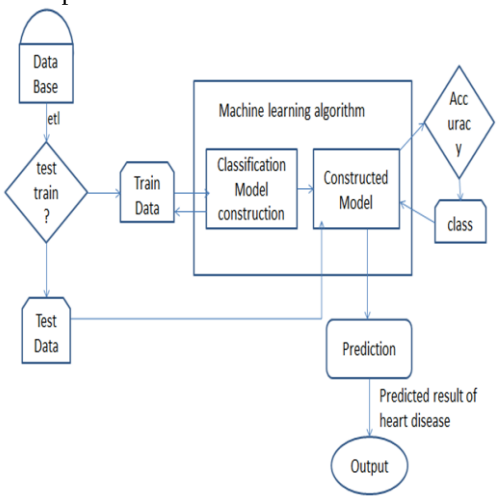

Figure 1. System Architecture

The data as input that plays a crucial role in the design shown, while entering the data to the R tool check its proper format which is taken in csv file format and if it is not as per need then error dialog box will be displayed. We first store the data in the required database, by creating a separate folder in our database to it, save the data set file in csv format such that we can access the data directly from R. The R programming language is an open source scripting language for prescient investigation and information visualization .To perform complex information measurable examination and show the outcomes in any of huge number of visual graphics. The R programming language incorporates capacities that help direct displaying, non-straight demonstrating, established insights, orders, grouping and more. Users can peruse information and burden it to the work space, determine directions and get results. Directions can be anything from straightforward numerical administrators, including +, - , * and/, to progressively confused capacities that perform straight relapses and other propelled calculations. Users can likewise compose their very own capacities. Upload the data in R tool and proceed for further implementation. After loading the data in R the data is read and it is subdivided into two parts as seventy and thirty, the 70% of the entries in the data set have been used for training and the remaining 30% have been used for testing the accuracy of the algorithm. First, we undergo with the train data set which is subdivided from the input data set taken, the predictive model is always built with train data set. The way to identify the train data set is that always contains a responsive variable which is one of the attributes present in the data set. After the process then the data is trained with the taken classification model and find the accuracy of the model by the trained data. The test data that undergoes with the trained model that is done previously, find the accuracy of the already trained model by using the test data set. Hear the accuracy is compared for both the test and trained data set, mostly the test data set has the highest accuracy which leads to further implementations.

The accuracy is found by applying the confusion matrix to the model. A confusion matrix is a precis of prediction results on class trouble. The range of accurate and wrong predictions are summarized with matter values and broken down through each class. This is the key to the confusion matrix. The confusion matrix suggests the methods wherein your type version is burdened when it makes predictions. In the field of machine learning, specifically the matter of applied math classification, a confusion matrix, additionally referred to as a slip matrix. It's a table that's typically accustomed describe the performance of a classification model (classifier) on a group of check knowledge that truth values are legendary. It permits the visual image of the performance of associate formula. The confusion matrix example that related to heart disease is as shown below.

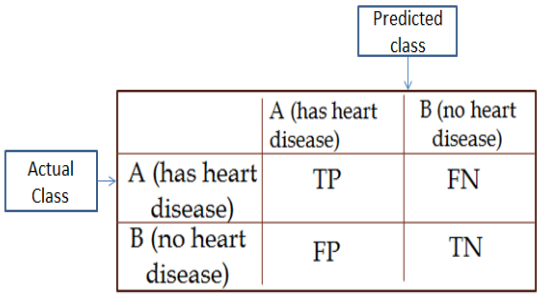

Figure 2. Confusion matrix

The following three algorithms are implemented:

- Neural Networks
- Decision Tree
- Logistic Regression

The operating of those algorithms has been explained within the sections ahead. The thought about models is trained exploitation the info set obtained from UCI repository. Few ways are taken for optimizing the algorithms for up the accuracy. These steps embody cleansing the information set and data pre-processing. The machine learning models were thought about supported their accuracy and it had been determined that the decision tree was the foremost correct out of the three algorithms with 88% of high accuracy. Hence it had been select as the best and will develop exploitation of different applications.

Inputs: Data set.

Outputs: Accuracy.

V. ALGORITHMS

A. Decision trees

A type of supervised Learning methodology, that works for each categorical and continuous input and output variables. Decision trees are graphics module of decision Matrix Analysis and have three varieties of nodes: decision nodes, fortuity nodes and terminating nodes. They split the population or sample into 2 or additional uniform sets (or sub-populations) supported most significant splitter or person in input variables and if there is a high non-linearity and complicated relationship between dependent and freelance variables, a tree model can outstrip a classical regression methodology. A choice tree model will significantly higher than a linear model and are less complicated to interpret than rectilinear regression. The decision tree mechanism is clear and that follow a tree structure simply to determine however the choice is created a decision tree could be tree structure consisting of internal and external nodes connected by branches. An enclosed node could be a higher cognitive process unit that evaluates a decision, operates to see that child node to go to next. The external node, on the opposite hand, has no child nodes and is related to a label or price that. The decision trees undergo the following methods for more accuracy.

1. Entropy:

Entropy is that the degree or quantity of uncertainty within the randomness of components or in alternative words it's a life of impurity. Intuitively, it tells us regarding the certainty of a specific event. Entropy calculates the homogeneity of a sample. If the sample is totally consistent the entropy is zero associated if the sample is an equally divided its entropy of 1.

2. Information gain:

It measures the relative modification in entropy with relevant to the freelance attribute. It tries to estimate the knowledge contained by every attribute. Constructing a choice tree is all regarding finding the attribute that returns the best info gain.

3. Iterative Dichotomies :

Id3 is a decision tree algorithm that is used for the classification of objects using iterative inductive approach. It's basically used to construct decision trees that consist of the top nodes, known as the root nodes and the other adjoining nodes, known as the leaf nodes.

4. Pruning:

In the machine, learning to prune is a technique associated with decision trees. Pruning reduces the dimensions of tree by removing components of the tree that don't offer power to classify instances. Decision trees are the foremost vulnerable out of all the machine learning algorithms to overfitting and effective pruning will scale back this chance.

5. Cross-validation:

Cross-validation is typically performed additionally to assembling a tree from the full information set. The cross-validation is completed simply to estimate the prognostic accuracy of the hypothesis or for parameter tuning.

B. Neural Network

An Artificial Neural Network (ANN) is a data processing paradigm that's galvanized by the method of biological nervous systems, like the brain, process data. It's composed of a vast sized range of extremely interconnected process parts (neurons) operating in unison to resolve specific issues. Adaptive learning: a capability to be told the way to do tasks supported the information given for coaching or initial expertise. Self-Organization: an ANN will produce its own organization or illustration of the data it receives throughout learning time. Real-Time Operation: ANN computations are also applied in parallel, and special hardware devices are being designed and made that cash in of this capability. The proposed heart disease prediction system was designed as a multilayer perceptron neural network. The designed ANN

has three layers: namely an input layer, a hidden layer, and an output layer. ANN additionally supports Real Time Operation wherever computations are often allotted in parallel, together with special hardware devices being designed and made taking advantage of this capability

1. Normalization

Normalization is needed in order that all the inputs are at a comparable vary and is important as a result of an input layer the increased worth of weight and input variable ought to activate to terribly little but three thus it's necessary to get the higher result it should be normalized. The neural network will simply counter your normalization since it simply scales the weights and changes the bias. When numeric x-data values are normalized, neural network training is usually economical, that results in an improved predictor. Basically, if numeric information isn't normalized, the magnitudes of 2 predictors are so much a part, and then a modification within the value of a neural network weight has the much more relative influence on the x-value with larger magnitudes.

C. Logistic Regression

The method or machine learning model performed on the data sets was Logistic Regression. This model is a Statistical method that analyses a data set having one or more independent variables to determine an outcome. The outcome is measured with a dichotomous variable, which means there could be two possible outcomes (True or False). Logistic Regression is used when finding the best fit model to describe the relationship between the dependent variable and set of freelance variables, The Linear model Equation (1) is:

$$y = b0 + b1x$$

Where b0 is the constant and b1 is the slope which defines the steepness of the curve. The Logistic model Equation (2) is in the odd's ratio as:

$$p = 1/ (1 + e - (b0 + b1x))$$

Logistic Regression is employed once the response variable y could be a binary variable, and autonomous factors or predictors are numeric. It models the association amongst x and y by fitting a calculated supplying bends, that seems like formed or Sigmoid Curve. It fits a curve (y=f(x)) . Once y is in the type of chances, or binary ("0" or "1"," failure" or "success"). It finds application in such examples to predict, who can win the following IPL match, next tournament based on their quality and strength, whether an individual have heart disease or not (Yes or no).

VI. Comparitive Study

For the classification model construction, the R studio is used to hear. R is freely available by the general public license, which is an open source. The steps involved in R script are:

1) Start
2) Load data to R.
3) Partition data -train 70%, test 30%.
4) Implementing the model.
5) Confusion matrix –test/train.
6) Miss classification error.
7) Accuracy.
8) Classification.

END

The comparison of the three models is done by following the above steps. The accuracy of the model can also be increased by changing the test and train data percentage. The models are compared with a single data set such that we can find which model is more accurate for finding the heart disease. A further implementation is required for the model accuracy implementation; it depends on the algorithm chosen.

VII. Results and Discussion

Finally, we show a summary of the best test results of different datasets on the 3 models. And can be shown in the table. The misclassification rate and the accuracy of the train and test dataset of three algorithms have been given in the table. By modifying the training process of execution we can implement the test accuracy rate. By changing the test and train percentage taken we can implement the classification model, such that it leads to high accuracy. The more accurate algorithm can predict the best results for heart disease prediction. Hear we got the highest accuracy of test data is for logistic regression. So it is the best model in predicting heart disease for the given data set.

	Train data		Test data	
	Misclassification	accuracy	Misclassification	accuracy
Decision trees	13.08	86.91	17.97	82.02
Neural networks	47.66	52.33	41.57	58.42
Logistic regression	16.82	83.17	11.23	88.76

Figure 3. Results

VIII. Conclusion and Future Scope

The most efficient algorithm was to be selected based on various criteria. We found out that the logistic regression algorithm was the most efficient out of the three with an accuracy of 88%. Decision tree, neural network had an accuracy of 82% and 58% respectively. Thus the logistic regression algorithm was further implemented using different applications. The highest training accuracy is 86% for

decision trees. For this, R studio was used. Since heart diseases are major killer in India and throughout the world, the application of a promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. There are numerous conceivable enhancements that could be investigated to improve the adaptability and exactness of this predicted system. By training the model with different dataset may lead to best fit model because this heart disease data set may vary with years. It could be more benefited by changing the data set and by implementing different algorithms for the prediction of heart disease may increase the efficiency of the prediction.

REFERENCES

[1] Sanjay Kumar Sen, *"Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms",* International Journal of Engineering and Computer Science, ISSN: 2319-7242, Volume.6, Issue.6, June 2017, Page No. 21623-21631.

[2] Jaymin Patel, Prof. TejalUpadhyay, Dr. Samir Patel, *"Heart disease prediction on using machine learning and data mining technique."*, IJCSC, Volume.7, Number.1, pp.129-137, 2016.

[3] Tülay Karayilan, izkan Kiliç, *"Prediction of Heart Disease Using Neural Network"*, IEEE, Vol.978, Issue.1, pp.5386-0930, 2017.

[4] Theresa Princy. R, J. Thomas, *"Human Heart Disease Prediction System using Data Mining Techniques"* 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT] , Vol.978, Issue.1, pp.5090-1277, 2016 IEEE.

[5] Ashok Kumar Dwivedi, *"Performance evaluation of different machine learning techniques for prediction of heart disease"*, Springer, DOI 10.1007/s00521-016-2604-1. 2016.

[6] Akash Mukherjee, Raj Manjrekar, Ashish Marde, Prof. Rajesh Gaikwad, *"Heart Disease Prediction Using Artificial Neural Networks"*, IJSRD, National Conference on Technological Advancement and Automatization in Engineering, January 2016, ISSN: 2321-0613.

[7] A. Sankari karthiga, M. Safish Mary, M. Yogasini *"Early Prediction of Heart Disease Using Decision Tree Algorithm"* , International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST), Vol.3, Issue.3, March 2017.

[8] Avni Sharma, Deeksha Tyagi, Dr. Tarun Kumar Gupta, *"Comparative Analysis of Machine Learning Techniques in Heart Disease Prediction by R Language"*, IJSRD - International Journal for Scientific Research & Development|, Vol. 5, Issue 02, 2017.

Authors Profile

Nukala V V Pravallika pursed Bachelor of Technology from University of K L, Guntur in 2017 and She is currently pursuing PG in Department of Computer Science and Technology from University of Adikavi Nannaya, Rajahmundry since 2018. She is a member of CSI Computer Society of India since 2018. She has published a research papers in reputed international journal and it's also available online. Her main research work focuses on Machine Learning Algorithms, Big Data Analytics, Data Mining.

Prof P Suresh Varma M.Tech, Ph.D. He is currently working as Professor in Department of Computer Science, University of Adikavi Nannay, India. He has published more than 150 research papers in reputed international and national journals. His main research work focuses on Network Security, Cloud Computing, Computer Networking, IoT, wireless Computing,web services. He has 25 years of teaching and Research experience.

