iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.61179

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Improving Cardiovascular Stroke Risk Prediction with Machine Learning

Mrs. T.Ganga Bhavani[1] , S Chaitanya Sai Bhaskar[2], N N Venkat Prabhash[3], A Ajaykumar Simhadri[4], Beera Vamsidhar[5], Bula Nithin[6,]

[1]*Assistant Professor* , [2, 3, 4, 5, 6]*B.tech Students Department of Information Technology, Pragati Engineering College, Surampalem , Andhra Pradesh, India*

Abstract: ***Cardiovascular diseases have overtaken all other causes of death in industrialized, developing, and underdeveloped countries over the previous several decades. Receiving clinical care and detecting heart abnormalities early can help to reduce death rates. We created a model for forecasting heart illness and detecting imminent heart disease using machine learning techniques such as logistic regression, SVM, Multinomial Nave Bayes, Random Forest, and Decision Tree based on the patient's various cardiac data. The most common type of input is numerical data representing a variety of parameters. Real-time output results are then generated to determine whether or not the patient has a disease. Before deciding on the best supervised machine learning technique for the model, we will try a variety of approaches. Current systems use imprecise and inefficient traditional deep learning models. They take slightly longer to process and are less precise than the suggested model.***

Keywords: Machine learning, Black Box, Logistic regression, Heart ailment, SVM

I. INTRODUCTION

Machine learning is used in many different industries throughout the world. The healthcare industry is no different. Machine learning can be very useful in predicting the incidence of ailments such as cardiovascular disease, locomotor disorders, and others. Such evidence, if expected, can provide medical practitioners with valuable information that allows them to modify their diagnostic and treatment techniques. The heart is one of the body's major organs. It moves blood through the circulatory system's blood vessels. The circulatory system is essential to the body because it transports blood, oxygen, and other nutrients to all organs. The circulatory system's most important component is the heart. A faulty heart can cause major health problems, including death. The health care business has access to large amounts of medical data; thus, machine learning algorithms are crucial for accurate cardiac illness prediction. Recently, researchers have focused on combining these methodologies to create hybrid machine learning algorithms. In the study proposal, data pre-processing is used to reduce noisy data, complete missing data when necessary, offer default values when appropriate, and categorize features for various degrees of prediction and decision-making. The treatment plan's success is evaluated using techniques such as classification, accuracy, sensitivity, and specificity analysis. This study compares the accuracy of SVM Classifier, Decision Tree, Logical Regression, Random Forest, and Multinomial Naïve Bayes models in predicting cardiovascular disease outcomes.

II. LITERATURE SURVEY

[1] To determine the usefulness of machine learning methods for forecasting cardiac illness, such as k-nearest neighbor, decision tree, linear regression, and support vector machine, the researchers test and train using data from the UCI repository.

Using classification techniques such as logistic regression, K-nearest neighbor, support vector machines, decision trees, and random forests, [2] aimed to develop a method for predicting cardiovascular disease and compare the efficacy of the proposed scheme to the uniform standard.

[3] The writers of this article provide a quick overview of various classification methods. The majority of these classification algorithms use threshold values and benchmark methodologies to determine whether the dataset's properties are optimum.

[4] The Heart Disease Prediction System offers consumers with a prediction result indicating their risk of developing CAD. The system employs the MLP machine learning algorithm. Machine learning algorithms have expanded significantly as a result of recent technological breakthroughs, and the authors chose MultiLayered Perceptron (MLP) for the proposed system due to its accuracy and efficiency.

©IJRASET: All Rights are Reserved

[5] This study presents a novel strategy for forecasting cardiovascular illnesses using machine learning. Betthe Authorsen's various machine learning algorithms are compared to assess their effectiveness on the dataset; the proposed solution employs a 92.30 percent accurate K Nearest Neighbor.

As per [6,] machine learning has demonstrated efficacy in deriving inferences and projections from a substantial quantity of data accumulated over an extended period by the healthcare industry. Artificial neural networks (ANN), decision trees (DT), random forests (RF), support vector machines (SVM), naive Bayes (NB), and the k-nearest neighbor approach are some of the machine learning techniques employed in this prediction of heart disease.

[7] For instance, machine learning may identify the most important information in a patient's medical records and utilize data to predict the patient's prognosis. The right combination of relevant features must be chosen in order to increase the estimating approaches' efficacy.

[9] The recommended diagnostic system is optimized using the grid search method. The Cleveland dataset, an online database on heart failure, was used in the studies. The proposed approach is more economical and less complex than the typical random forest model, with a 3.3 percent higher accuracy using only 7 features.

[10] The authors of this study describe a novel method for using machine learning techniques to uncover essential traits, hence boosting the accuracy of cardiovascular disease prediction. The forecast model is introduced using a variety of popular classification methodologies and key criteria.

III. SYSTEM ANALYSIS

A. *Existing System*

The present model for predicting cardiovascular stroke is primarily based on classic deep learning algorithms. When compared to the suggested model, it is clear that these models are less accurate and efficient. They typically have worse accuracy and take longer to process. In the current approach, numerical parameters related to the patient's cardiac characteristics are typically used as input data. The result, which is generated in real time, anticipates the patient's likelihood of having heart disease. The system uses standard machine learning methods like logistic regression, SVM, Random Forest, Multinomial Naive Bayes, and Decision Tree to provide these forecasts. However, because to their inefficiencies and reduced accuracy, these models are less helpful for clinical care and early diagnosis of heart issues, which are critical in lowering the death rate associated with cardiovascular diseases.

DISADVANTAGES OF THE EXISTING SYSTEM

1) *Inefficiency:* The computational inefficiency of traditional deep learning models frequently results in longer processing times.
2) Reduced Precision: More advanced machine learning techniques may be more precise and accurate than the current process when it comes to predicting heart disease.
3) *Lack of Scalability:* Conventional models may struggle to scale well when faced with large datasets or when adding additional features and data sources.
4) *Limited Feature Engineering:* There's a chance that the existing system will limit your capacity to perform feature engineering, which is crucial for finding complex connections in the data.
5) *Reduced Adaptability:* These conventional models are less suitable for early detection and real-time cardiac status prediction because they could not be as adaptable to changing data patterns and evolving trends.

B. *Proposed System*

The proposed system has a machine learning model that has been trained to identify an individual's risk of cardiovascular disease. It anticipates the threat and alerts them to it. The prediction approach applies techniques commonly used in categorization problems, such as logistic regression, SVM, machine learning, and Naïve Bayes. Several structured data sets should be created from our data based on the patient's cardiac features.

A logistic regression model is developed with the available data to predict the patient's illness. The initial step would be to collect and prepare the data. Our dataset, which was obtained from Kaggle, is quite effective at recognizing and forecasting. High accuracy requires a huge dataset.

The data are preprocessed after collection to remove any null values from the dataset. The ML model's training process will be significantly influenced by the dataset's Null values. The dataset needs to have a lot of its data standardized before it can be utilized for training. Data transformation is the process of converting data into a format better suited for data mining.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

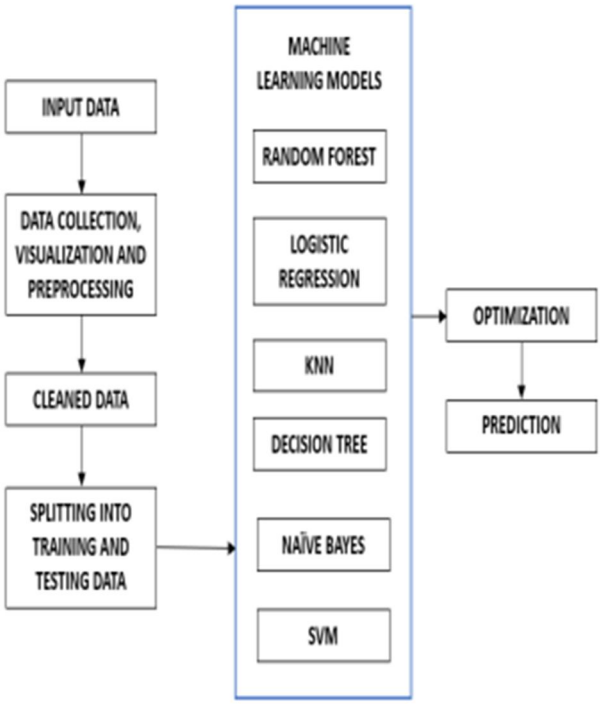


Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

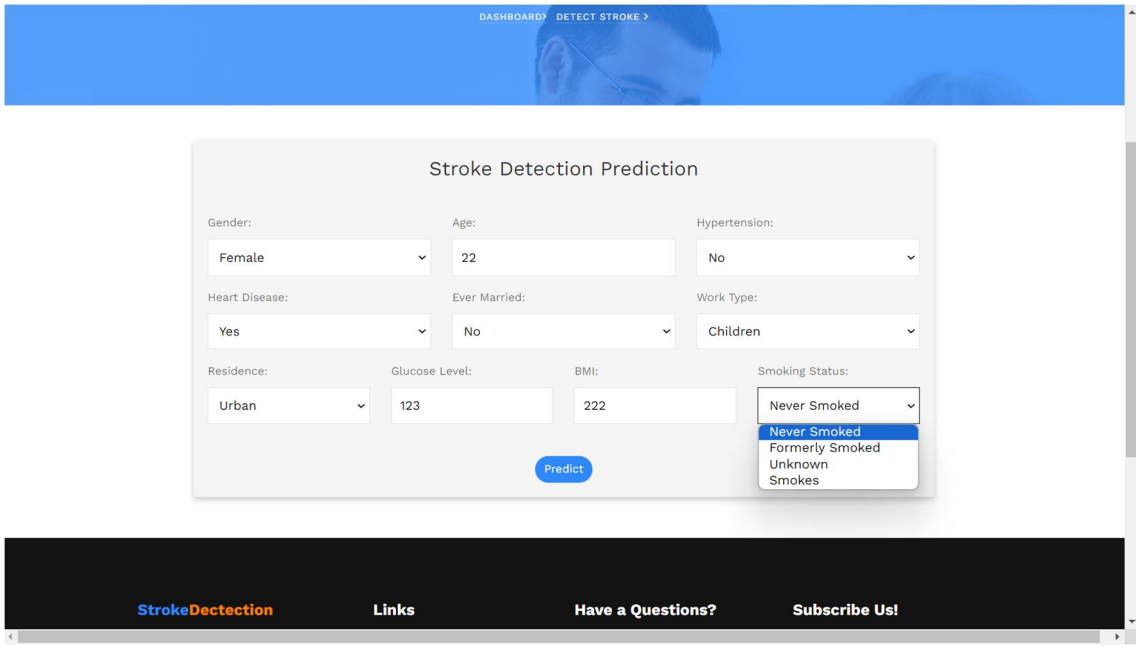
1) *Data Collection and Preprocessing:* This department is in responsible of compiling the relevant medical history, lab findings, and vital signs for each patient. Preprocessing involves cleaning, normalizing, and preparing the data for machine learning algorithms.
2) *Selecting and extracting features:* involves using patient data to identify important characteristics that can be used for prediction. Features that are most relevant to the model are selected, which reduces dimensionality and enhances model performance.
3) *Training and optimizing machine learning models:* such as logistic regression, SVM, Multinomial Naive Bayes, Random Forest, and Decision Tree, utilizing historical patient data. Finding relationships and patterns that can be used to forecast the future is the aim.
4) *Real-time Prediction:* Following model training, this module handles real-time predictions. By estimating the likelihood of cardiovascular disease based on recently gathered patient data and the trained model, it gives medical professionals rapid feedback.
5) *Model Evaluation and Continuous Improvement:* This module assesses the performance of the machine learning models using metrics including accuracy, precision, recall, and F1-score. It also facilitates the process of retraining models with new data to improve forecast accuracy and adapt to changing health patterns.

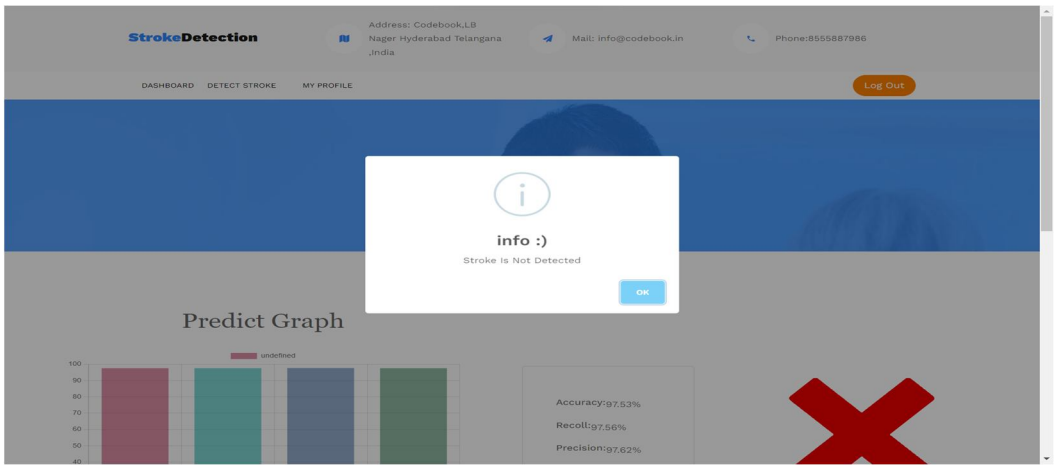
VI. RESULTS AND DISCUSSION

To improve performance, 10-fold cross validation was first applied to the training set. Second, results were obtained by applying the model; third, features were modified; and fourth, the framework was confirmed. The current status of mining studies on cardiovascular disorders is reviewed in this publication. With machine algorithms, it is feasible to efficiently "mine" pertinent data from the enormous amounts of data produced by the medical industry.

©IJRASET: All Rights are Reserved

The outcomes are significantly better when many mining methods are applied. Cardiovascular disease monitoring is quickly and efficiently implemented thanks to a precise combination of mining algorithms and their exact application on the supplied data.





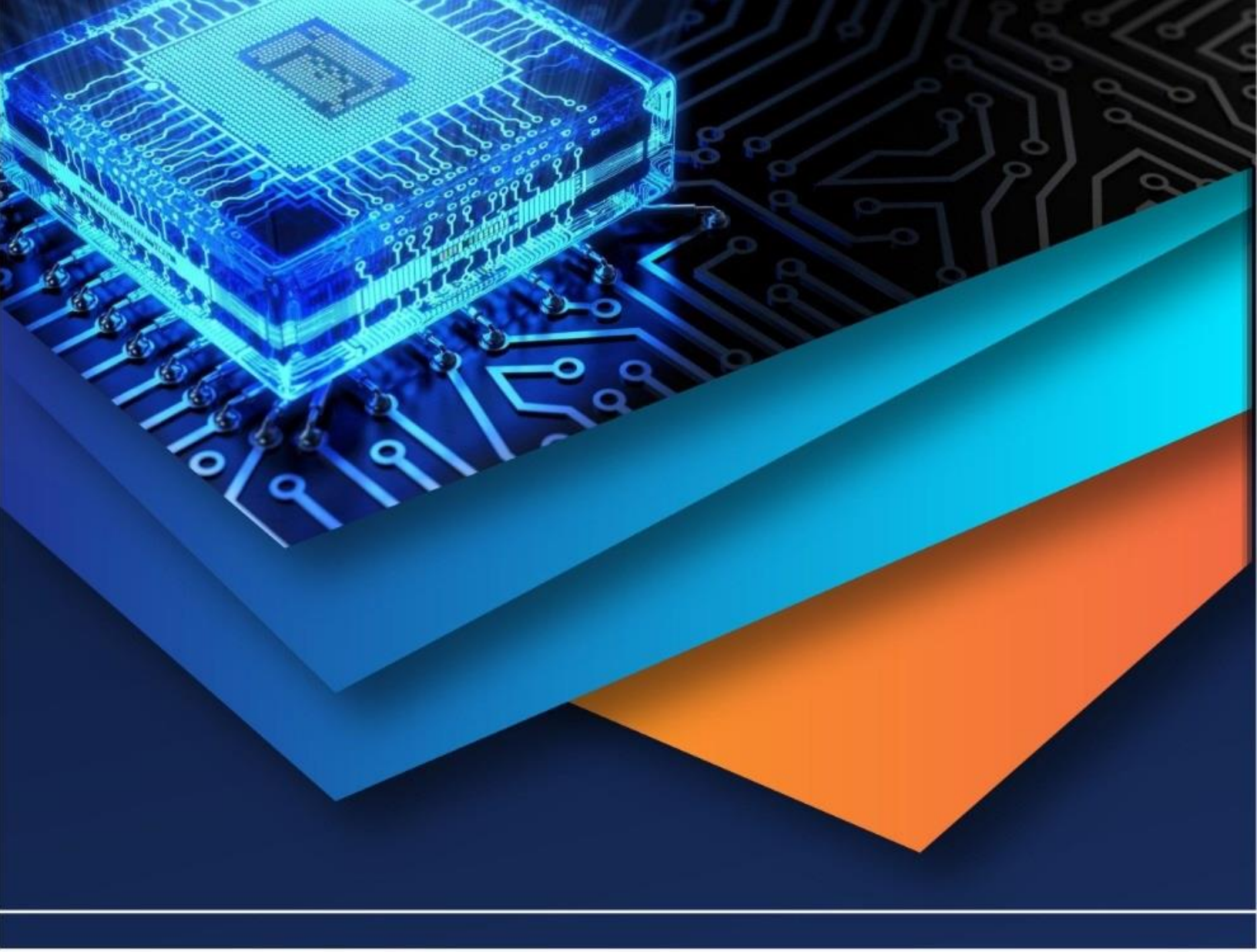
VII. CONCLUSION AD FUTURE WORK

Early detection can aid in the development as well as the prevention of illness. Machine learning technologies provide promise in aiding timely diagnosis and identifying important contributing elements. A deep learning model that can predict heart attacks and cardiovascular disorders is created with the use of the recommended technique. For this assignment, the SVM algorithm is the best choice. The model suggests that new machine learning techniques usually lead to better prediction accuracy. Early detection can aid in the development as well as the prevention of illness. Machine learning technologies provide promise in aiding timely diagnosis and identifying important contributing elements. A deep learning model that can predict heart attacks and cardiovascular disorders is created with the use of the recommended technique. For this assignment, the SVM algorithm is the best choice. The model suggests that new machine learning techniques usually lead to better prediction accuracy. In an attempt to determine precise and efficient methods, the existing methodologies are evaluated and compared. Machine learning algorithms enable people to obtain preventive medication and be detected early in the course of their ailment by considerably enhancing the accuracy of cardiovascular risk prediction. One could argue that machine learning techniques have a lot of potential for predicting heart- or circulatory-related diseases. The previously mentioned techniques have all performed remarkably effectively in every situation. By employing the multimodal technique, we were able to reduce processing time and increase accuracy.

REFERENCES

[1] Singh et al., "Heart Disease Prediction Using Machine Learning Algorithms", 2020 International Conference on Electrical and Electronics Engineering (ICE3), pp. 452-457, February 2020.
[2] K. Hashi and M.S.U. Zaman, "Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction",Journal of Applied Science & Process Engineering, vol. 7, no. 2, pp. 631-647, 2020.
[3] A. M.A and P. A. Thomas, "Comparative Review of Feature Selection and Classification modeling", 2019 International Conference on Advances in Computing Communication and Control (ICAC3), pp. 1-9, 2019
[4] A. Singh et al., "Heart Disease Prediction Using Machine Learning Algorithms", 2020 International Conference on Electrical and Electronics Engineering (ICE3), pp. 452-457, February 2020.
[5] K. Hashi and M.S.U. Zaman, "Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction",Journal of Applied Science & Process Engineering, vol. 7, no. 2, pp. 631-647, 2020.
[6] A. M.A and P. A. Thomas, "Comparative Review of Feature Selection and Classification modeling", 2019 International Conference on Advances in Computing Communication and Control (ICAC3), pp. 1-9, 2019
[7] A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease using Machine Learning", 2018 Second International Conference on Electronics Communication and Aerospace Technology (ICECA), pp. 1275-1278,2018.
[8] F. Mendonca, R. Manihar, A. Pal and S. U. Prabhu, "Intelligent Cardiovascular Disease RiskEstimation Prediction System", 2019 International Conference on Advances in Computing Communication and Control (ICAC3), pp. 1-6, 2019. 92 Authorized licensed use limited to: Universidad de Almeria. Downloaded on September 30,2023 at 14:01:46 UTC from IEEE Xplore. Restrictions apply. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)
[9] Rahul Katarya and Polipireddy Srinivas, "Predicting heart disease at early stages using machine learning: A survey", 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020.
[10] Davide Chicco and Giuseppe Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone", BMC medical informatics and decision making, vol. 20.1, pp. 1- 16, 2020.
[11] Mohammad Shafenoor Amin, Yin Kia Chiamand Kasturi Dewi Varathan, "Identification of significant features and data mining techniques inpredicting heart disease", Telematics andInformatics, vol. 36, pp. 82-93, 2019
[12] Ashir Javeed et al., "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection", IEEE Access, vol. 7, pp.180235-180243, 2019
[13] S. Mohan, C. Thirumalai and G. Srivastava,"Effective heart disease prediction using hybrid machine learning techniques", IEEE access, vol.7, pp. 81542-81554, 2019

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)