iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45985

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 10 Issue VII July 2022- Available at www.ijraset.com

Heart Attack Predictions Using Machine Learning

Manoj Y.R [1], Dr Nagarathna [2]
[1]*Student,* [2]*HOD and Professor, Department of computer science and engineering, PES Collage of Engineering Mandya*

Abstract: ***In recent years, predicting heart disease has grown to be one of the most challenging problems in medicine. About one person dies from heart disease every minute in the modern era. In the healthcare industry, machine learning is crucial for handling massive amounts of data. Since predicting cardiac disease is a complicated procedure, it is necessary to automate the process in order to minimize risks and forewarn patients. The heart disease dataset from the machine learning repository is used in this study. By implementing KNN AND SVM, RF the suggested study predicts the risk of heart disease and categorizes the patient's risk level. As a result, this essay presents a comparison research by examining the effectiveness. of several machine learning techniques. The test results show that, when compared to other ML algorithms used, the algorithm gets the maximum accuracy of 94.3 percent.***
Keyword: Machine learning, KNN, SVM, RF.

I. INTRODUCTION

The second most important part of the human body, after the brain, is the heart. The coronary heart's main function is to circulate bl ood throughout the entire body [1]. Heart disease is any conditions that may impair the heart's ability to pump blood.There are sever al different types of coronary heart disorders, including coronary artery.. Any disease which could lead to disturbing the functionality of the heart is known as heart disease. Several types of coronary heart disorder are there in the world; Disorder and coronary heart failure are the maximum common heart illnesses that are present. The major reason behind the coronary heart disorder is blockage or narrowing down of the coronary arteries Coronary arteries additionally responsible for imparting blood to the coronary heart. CAD is the main purpose of demise over 26 million human beings are stricken by coronary heart disorder round the sector, and it's far growing 2% yearly because of 17five million deaths globally. In the developing world, 2% of the populace round the Sector is stricken by CAD, and 10% of the human beings are older than sixty five years. Approximately 2% of the once a year healthcare price range spent best to deals with the CAD disorder [2].

It truly causes doctors headaches. Today, doctors are applying a wide range of scientific technologies, including numerous devices that are quite effective at predicting a wide range of ailments. When using a heart attack prediction system that leverages machine le arning to get accurate findings, doctors frequently may not make precise recommendations. the use of numerous traditional risk indi cators, including vital characteristics such as age, gender, blood pressure, etc., to forecast coronary heart problems.

A dataset from the kaggle [3] that has a number of physiological features as its attributes is chosen to continue with this task. The actual diagnosis is determined by looking at these characteristics .the dataset is cleaned in order to make it understandable the ML model. This phase's name is data preparations the data is checked for empty values, and any missing entries are the n filled. In label encoding is followed if necessary, by one encoding it convert string value into integers. The dataset is divided into train and test data following data preparation. After that, a model is built utilizing the new dataset and a range of categorizations methods. Accuracy is determined for each of these methods in order to determine which predictions model is the best-trained compared.

We develop the classifier, test its accuracy, and the develop an HTML website and flask applications the user enters the predictions parameters in the web applications. The web applications and trained model are connected through the flask applications. Study determines which algorithm is most suited for heart attack predictions giving it through evaluations.

II. RELATED WORKS

In the aforementioned research, we examined various classification algorithms**[1]** that may be applied to categorise databases of car diac illness, as well as various classification strategies and the technology that results from them. According to the instruments made to conduct the work, this survey provides information on various technologies utilised in various sorts of papers with various featur es and variable degrees of accuracy.

The research describes a machine learning prediction method for cardiac disease. **[2]**MLP gives the user a prediction result that reve als their state prior to CAD. Machine learning algorithms have significantly changed as a result of recent technical breakthroughs; as a result, we are adopting Multilayer Perceptions (MLP) in the suggested system due to its effectiveness and accuracy.

©IJRASET: All Rights are Reserved

Additionally, the algorithm uses input from the user to provide a closeby trustworthy output.People who use the system more frequ ntly will be aware of their existing heart status and the death rate from heart disease will eventually go down as a results.
This research aims to offer a method that can be incorporated into automated heart disease prediction
[3] systems. The many methods described in this article as related work have become more widely used in recent years to quickly an d accurately diagnose heartdisease.The data and the model's purpose should be taken into consideration while choosing a model. Re searchers are utilising predictive analytic models to identify the disease, which is motivated by the rising annual global mortality rat e of heart disease patients and thaccessibility of enormous amountsof data. heart. The use of prediction models by medical experts ai ds in the diagnosis of cardiac disease. The predictive analytics model used to choose the most suitable course of action for patients withless attention has been paid to people with heart problems.

The paper's broad perspective is the primary cause of mortality is a heart attack, which must be anticipated in its early stages to prevent fatalities. [4] To predict the outcomes and patterns with a respectable degree of accuracy and dependability, the prescient research involves the separation of data from current datasets. This study analyses a few variables and forecasts heart attacks. Thus, based on Naive Bayes arrangement and Framework structure, it suggests a heart attack forecasting framework. Naive Bayes classifiers are a category of fundamental probabilistic classifiers used in machine learning with the intention of adopting the Bayes hypothesis.

This paper is essential to foretell the earlier actions that heart attack[5] are a major cause of death and might affect a person's daily li fe.The anticipated surveys involve isolating the data from alreadyexisting data sets in order to forecast outcomes, accuracy values, a nd valuable reliability.In this white paper, various criteria are explained and a heart attack is predicted.The Naive Bayes, Hadoop, an d Mahout Stored Heart Structure Strength Assault Framework is advised. The Naive Bayes classifier is a collection of probabilistic f undamental classifiers that take the hypothesis into consideration in machine learning.

Ten machine learning (ML) classifiers from various categories, including Bayes, functions, lazy, meta, rules, and trees, [6] were trained in this study to predict the risk of heart disease accurately using both the Cleveland heart dataset's full set of attributes and the best attribute sets determined by three attribute evaluators. A 10-fold cross-validation testing option was used to evaluate the algorithms' performance. Finally, we tuned the instance-based classifier's hyper parameter k, which stands for the number of nearest neighbours. The greatest accuracy value obtained by the sequential minimal optimization (SMO) when employing the entire set of attributes was 86.468 percent. chi-squared attribute evaluator's best attribute set, or set of attributes. Meanwhile, employing both the complete and ideal attribute sets acquired from the Relief attribute evaluator, the Meta classifier bagging with logistic regression (LR) provided the greatest ROC area of 0.91. Overall, the SMO classifier was the most accurate approach for making predictions when compared to other methods, and improved accuracy by 8.25 percent by setting the hyper parameter "k" to 9 using the chi-squared attribute set

One of the most prevalent diseases is heart disease. Since this condition is now fairly widespread, we used various characteristics that were closely related to heart diseases to find the best way to predict it[7]. We also used algorithms to make predictions. On the basis of risk variables, the Naive Bayes method is examined on the dataset. Based on the aforementioned characteristics, we also employed decision trees and a combination of algorithms to predict heart disease. According to the findings, decision trees produce accurate results when the dataset is large and the naive Bayes method produces accurate results when the dataset is small.

In this study, numerous heart disease-related characteristics are presented, along with a model built using supervised learning techniques such Naive Bayes, decision trees, K-nearest Neighbor, and random forest algorithm.[8] It makes use of the current dataset from the UCI heart disease patient repository's Cleveland database. There are 303 instances and 76 attributes in the collection. Only 14 of these 76 attributes—which are crucial to proving the effectiveness of various algorithms—are taken into consideration for testing. The purpose of this study work is to predict the likelihood that patients would develop heart disease. The findings show that K-nearest Neighbor achieves the highest Accuracy score.

In order to compare the findings and analyses of the UCI Machine Learning Heart Disease dataset, various machine learning techniques including deep learning are used in this study. [9] The dataset has 14 key attributes that will be used in the research. The accuracy and confusion matrix is used to achieve and validate a number of promising results. Isolation Forest is used to manage some unimportant aspects in the dataset, and data is standardized for better outcomes. Also highlighted is how this study may be coupled with other multimedia technology, such as mobile devices. The accuracy was 94.2 percent using the deep learning method.

The primary focus of the research paper is on which patients, given certain medical characteristics, are more likely to suffer heart disease. Using the patient's medical history, we developed a system to determine if a heart disease diagnosis is likely or not for the patient. [10] To forecast and categorize the patient with heart disease, we employed a variety of machine learning methods, including KNN and logistic regression.

The regulation of how the model can be utilized to increase the precision of heart attack prediction in any individual was done in a very helpful way .When compared to the previously employed classifiers, such as naive bayes, etc., the proposed model's accuracy in predicting signs of having a heart disease in a specific individual was quite satisfactory. It did this by using KNN and Logistic Regression. Thus, utilizing the provided model to determine the likelihood that the classifier can correctly and accurately identify heart illness has relieved quite a bit of pressure.

III. EXISTING SYSYTEM

The patient provides the input information for this system. Then, using ML approaches, heart disease is assessed from the user inputs .The outcomes are now contrasted with the outcomes of the same models' current iterations domain and discovered to be enhanced. Data from KNN, Support Vector Machines SVM,andRF are used to discover patterns in data from heart disease patients that was obtained from the Kaggle platform. Different approaches are compared for output performance and accuracy. When compa red to the other strategies that are currently accessible, the proposed hybrid strategy produces outcomes at a 94,3 % rate.

IV. PROPOSED SYSYTEM

We performed heart disease research using python and pandas operations after assessing the findings from the existing approaches. for the data acquired from the Kaggle platform, classification repository. We evaluated the results from the previous methodologies and conducted heart disease study utilising pandas operations and Python. Categorization repository for the data acquired from the Kaggle platform. It has an accessible aspect. Representation of the dataset, the environment in which the prediction models were cre ated, and preprocessing the data comes first in the ML process, then feature extraction.selection based on data purification, classific ation, and performance modelling evaluation. Precision is improved by employing a forest technique to replace randomness.

V. . METHODOLOGIES

Various Kaggle datasets were investigated to continue the development. Out of all the publicly accessible datasets, a suitable dataset for model training was chosen. After the dataset has been obtained, the next step is to clean and organize it so that the computer can understand it more easily. Data preparation is what it was. This stage includes handling uneven data, addressing null values, and performing attribute encoding for these particular dataset. In terms of accuracy measurements, the model comparison provides the testing procedure's appropriate outcome. To make it easier for users to enter input data and receive results after a model asbeen appli ed, an HTMLformatted website is developed. A flask application, which is essentially a Python framework that connects the model and the web application, provides the

Model with Userentered data.The flask application receives predictions about the output from the model based on input data. The re sults will now be shown for users to review on the website page using this application.

That flowing architecture show the flowing from of the heart attack predictions using machine learning that all detailed description of the algorithms showed in flowing in details

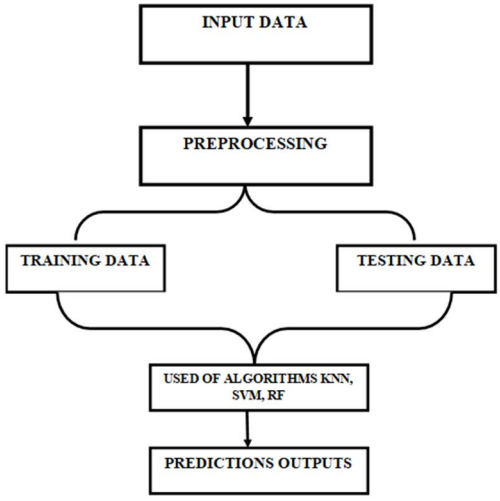


FIG: 1 depicts the flowchart of the suggested system's technique

A. Data Collections

Many details regarding patients' disorders and diseases have been collected in clinical databases in the website kaggle [3]. Heart disease contains a variety of conditions, contains heart-related conditions a cardiovascular condition is the primary factor in deaths worldwide. The phrase "cardiovascular illness" includes a wide variety of diseases affecting the heart, blood vessels, and the way that the heart pumps and circulates blood body. Records containing medical information were Taken from the database for heart disease in Cleveland. Using as a guide the dataset, key trends related to heart attacks extraction of a diagnosis the data were distributed equally. Training dataset and testing dataset. We were able to gather 10268 records in total with 14 medical attributes. Every attribute has a numerical value. There are only 14 attributes that we are focusing attributes.

B. Data Preprocessing

There are no null values in the dataset. However, there were a lot more outliers which needed to be handled carefully, and the datase t's distribution is also off. Two approaches were used. The outcomes were underwhelming without the use of outliers, a feature selection procedure, or applying the data directly to the algorithms. The findings obtained are extremely encouraging after employing the normal distribution of the dataset to overcome the overfitting issue and then applying machine learning methods for the outlier's detection. All of these preprocessing methods are crucial when transferring data for categorization or prediction.

1) *Preprocessing:* NaN values can be found in the database. The values of NaN programmed cannot process, hence these variables are required into numerical quantities to convert. In this method, the phrase NaN values are substituted for by the values once the column is calculated mean.
2) *Dividing:*The database is divided into training and testing portions. Database. The bulk of the data is exploited in training. The remaining 40% of the data are used for testing. Testing is done with the remaining 60% of the data.

C. Algorithm

The training data is trained using few major alternative ways of machine learning algorithms include SVM, KNN, and random forest algorithms Is fully described. The dataset that was used to predict the attack on the heart is seriously imbalanced. The entire dataset consists of records, 10256 of which indicate the likelihood of a brain stroke and 4861 of which indicate there will be no attack. Even while building a machine learning model with such data may result in accuracy, other accuracy measures like recall and precision are limited. If such skewed data is not handled properly, the forecast will be inaccurate and the results will be erroneous. Therefore, in order to develop a suitable system, it is first necessary to address the unequal data.

Heart attacks are the most frequent disease detected by medical professionals, and their frequency is rising yearly. Using the publicl y accessible, heart attack analysis prediction dataset, the study investigated ten widely used ML techniques for detecting brain stroke recurrence, which are as follows:

1) K Nearest Neighbor
2) Support Vector Machine
3) Random forest algorithm

As a consequence, came up with testing and training scores each method, and visually in the table below.

According to our algorithm analysis for the heart prediction model, we applied all of the aforementioned algorithms to compare which one was better. We choose the KNN model for prediction since testing score of trained models is more and because it is best method for model prediction according to metrics of TrainingScore of Trained Model of Trained Model of SVM, Random Forest, KNN classifier is more.

TABLE-1: Training and Testing Score of Trained Model

AGE	63	37	17	56
CP	1	1	0	1
TRESTBPS	3	2	0	0
CHOL	233	250	170	200
FBS	1	0	1	1
RESTECG	0	1	0	1
THALCH	150	187	77	79
EXANG	0	0	1	1
OLDPEAK	2.3	3.5	0	1
SLOPE	0	0	2	2
THAL	1	2	3	3
TARGET	1	1	1	1
HEART DISEASE	YES	YES	NO	NO

©IJRASET: All Rights are Reserved

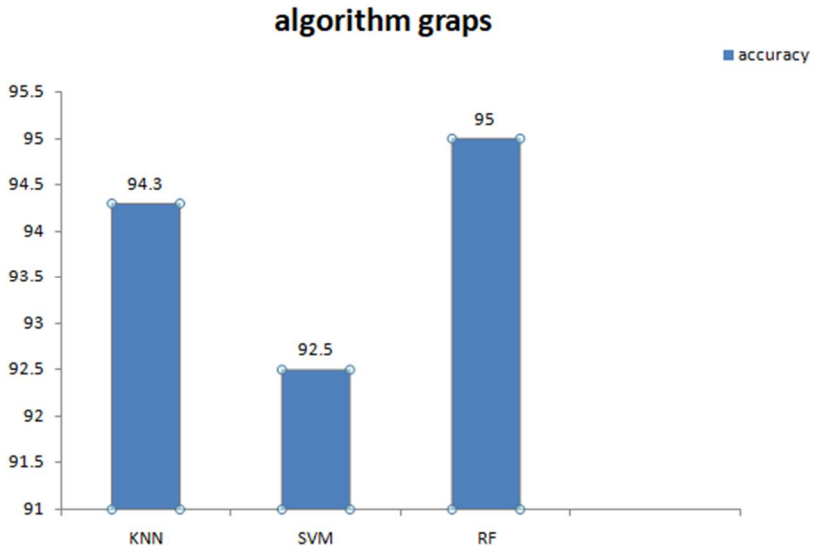


FIG-2: Graphical representation of Training and Testing Score of Trained Model I

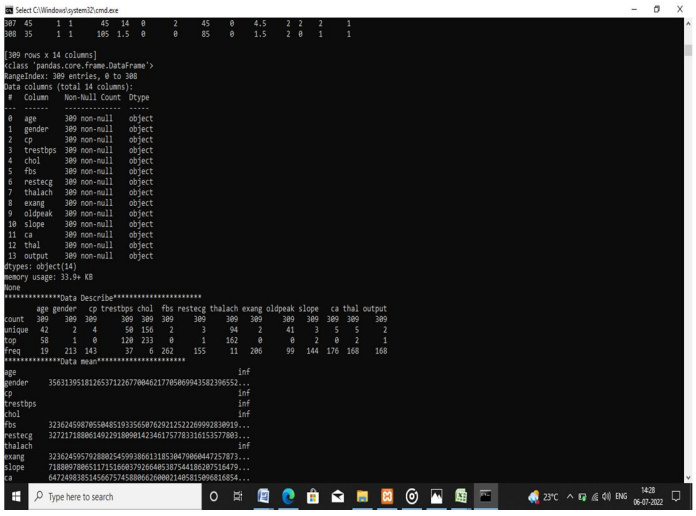


FIG: 3 Representation of Classification Report for KNN SVM AND RF Model

VI. WORKING OF THE SYSTEM

1) *Step 1:* Upload a data by entering the values using the selection and textbox given.
2) *Step 2:* Uploaded data will be loaded to the machine learning prediction model. Checks whether the metrics is having heart attack or not.
3) *Step 3:* If the uploaded metrics is having heart attack message will be displayed to user.

VII. EXPECTED RESULT

To make sure the algorithm used in the system is reliable, we tested it both with and without a heart attack, the programmed correctly anticipated the user inputted no-attack and attack metrics. The KNN, SVM, RF algorithm.

VIII. CONCLUSIONS

The main goal is to outline various ML methods that can be used to forecast heart disease accurately. Our objective is efficient and accurate prediction using fewer features and tests. We used K-nearest neighbour, SVM, random forest, and ML classification approaches to classify only the 14 most important attributes. The model employed pre-processed data that had been previously altered. The algorithms producing the greatest results in this model include K-nearest Neighbor, SVM, and random forest.

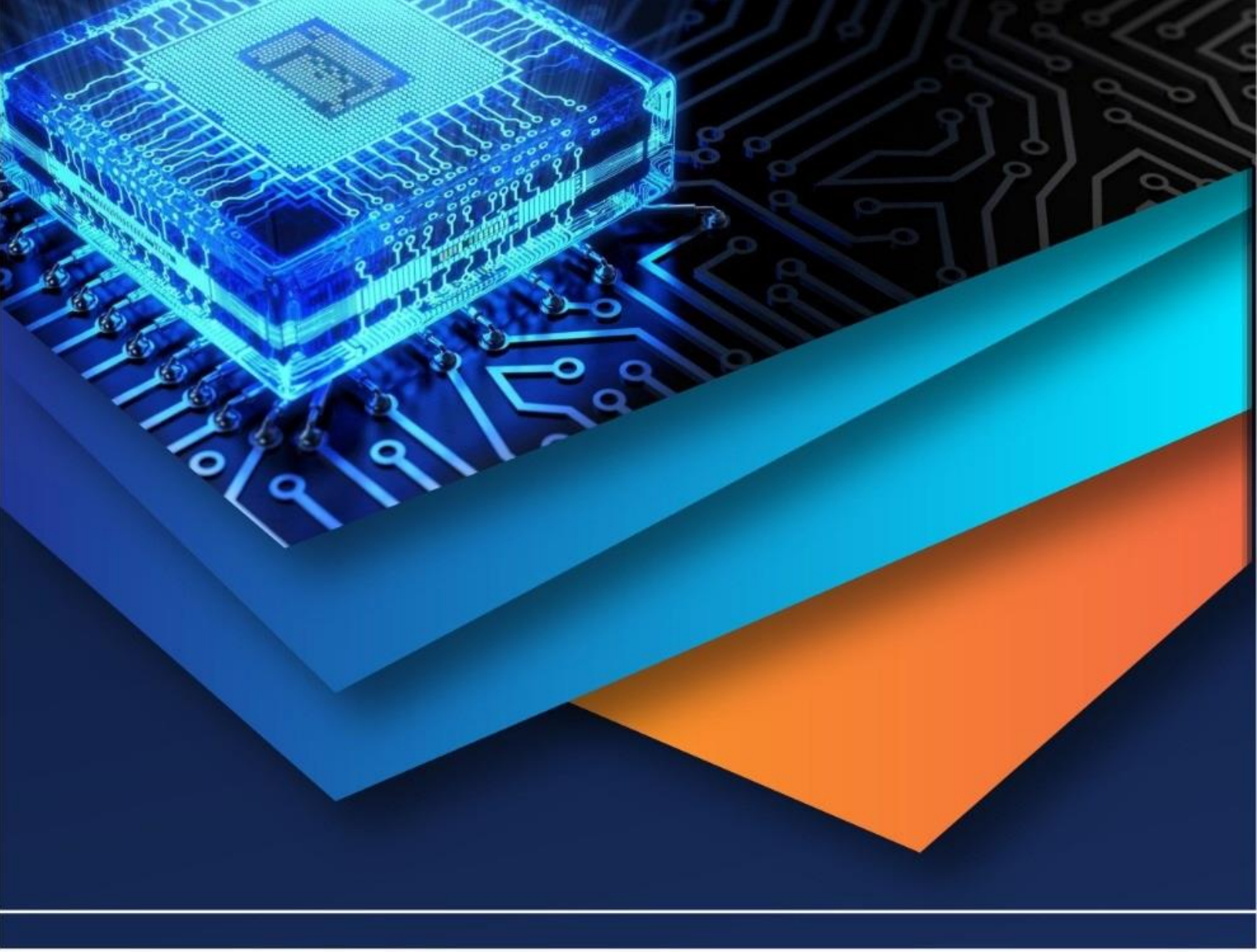
IX. FEATURE WORK

By combining more ML methods, such as time series, grouping and association rules, and evolutionary algorithms, we can further develop this work. Given the limits of this study, it is necessary to use more complicated models in conjunction to improve early prediction accuracy. Heart dieses

REFERENCES

[1] Concept of heart attack by Health line.
[2] Statistics of heart attack by Centers for Disease Control and Prevention.
[3] Dataset named 'heart attack using machine learning Dataset' from Kaggle. https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset
[4] Sonam Nikhar, A.M. Karandikar "Prediction of Heart Disease Using Machine Learning Algorithms" in International Journal of Advanced Engineering, Management and Science (IJAEMS) June2016 vol-2
[5] Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal
[6] Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data" in IEEE Journal of Biomedical and Health Informatics
[7] Murugesan, M., Thilagamani, S. ," Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network", Journal of Microprocessors and Microsystems, Volume 79, Issue November 2020, https://doi.org/10.1016/j.micpro.2020.103303
[8] DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Medical Data Classification" in Annual Conference on New Trends in Information & Communications Technology Applications - march2020
[9] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2019
[10] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dhavachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2019) Advances in Intelligent Systems and Computing, 381, pp. 341-351.
[11] Thilagamani, S., Nandhakumar, C. ." Implementing green revolution for organic plant forming using KNN-classification technique", International Journal of Advanced Science and Technology, Volume 29 , Isuue 7S, pp.
[12] V. Krishnaiah, G. Narasimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review" IJCA 2021.
[13] K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data Mining", IJARCSSE 2019.
[14] NagannaChetty, Kunwar Singh Vaisla, NagammaPatil, "An Improved Method for Disease Prediction using Fuzzy Approach", ACCE 2015.
[15] VikasChaurasia, Saurabh Pal, "Early Prediction of Heart disease using Data mining Techniques", Caribbean journal of Science and Technology,2018
[16] ShusakuTsumoto," Problems with Mining Medical Data", 0-7695- 0792-1 I00@ 2000 IEEE.
[17] Y. Alp Aslandoganet. al.," Evidence Combination in Medical Data Mining", Proceedings of the international conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.
[18] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," .
[19] Franck Le Duff, CristianMunteanu, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, page no. 1256-1259, 2004.
[20] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, KarstenSternickel,Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease Diagnosis", Proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16,page no. 305-310, 2006.
[21] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer 2006,Vol:345, page no. 721- 727.

©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)