iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36372

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Heart Disease Prediction using Machine Learning

Prof. Dr. R. Sandhiya[1], A. Shivani[2], J. Shalini[3], K. Tejasvi[4]
[1, 2, 3, 4]Information Technology, Coimbatore Institute of Technology, Coimbatore, Tamilnadu

***Abstract:** In recent times, the diagnosis of heart disease has become a very critical task in the medical field. In the modern age, one person dies every minute due to heart disease. Data science has an important role in processing big amounts of data in the field of health sciences. Since the diagnosis of heart disease is a complex task, the assessment process should be automated to avoid the risks associated with it and alert the patient in advance. This paper uses the heart disease dataset available in the UCI Machine Learning Repository. The proposed work assesses the risk of heart disease in a patient by applying various data mining methods such as Naive Bayes, Decision Tree, KNN, Linear SVM, RBF SVM, Gaussian Process, Neural Network, Adabost, QDA and Random Forest. This paper provides a comparative study by analyzing the performance of various machine learning algorithms. Test results confirm that the KNN algorithm achieved the highest 97% accuracy compared to other implemented ML algorithms.*
***Keywords:** Naive Bayes, Decision Tree, KNN, Linear SVM, RBF SVM, Gaussian process, Neral network, AdaBoost, QDA, Random Forest and Heart Disease Prediction.*

I. INTRODUCTION

The proposed work in this paper centers basically around different information mining rehearses that are utilized in coronary illness expectation. Human heart is the main piece of the human body. Essentially, it manages blood stream all through our body. Any inconsistency to heart can cause trouble in different pieces of body. Such an aggravation to ordinary working of the heart can be named a Heart illness. In the present contemporary world, coronary illness is one of the essential purposes behind event of most passings. Coronary illness may happen because of unfortunate way of life, smoking, liquor and high admission of fat may cause hypertension . As per the World Health Organization in excess of 10 million bite the dust because of Heart illnesses each and every year all throughout the planet. A sound way of life and soonest location are just approaches to forestall the heart related illnesses. The principle challenge in the present medical care is arrangement of best quality administrations and viable precise analysis . Regardless of whether heart illnesses are found as the great wellspring of death on the planet as of late, they are the ones that can be controlled successfully. The entire precision in administration of a sickness lies on the legitimate season of identification of that infection. The proposed method makes an endeavor to identify these heart infections at beginning phase to stay away from deplorable outcomes. Records of enormous arrangement of clinical information made by clinical specialists are accessible for examining and removing significant information from it. Data mining procedures are the methods for removing important and concealed data from the huge measure of information accessible. For the most part the clinical data set comprises of discrete data. Thus, dynamic utilizing discrete information becomes unpredictable and extreme undertaking. AI (ML) which is subfield of information mining handles enormous scope very much arranged dataset effectively. In the clinical field, AI can be utilized for analysis, discovery and forecast of different illnesses. The primary objective of this paper is to give a device to specialists to recognize coronary illness as beginning phase. This thus will assist with giving viable treatment to patients and keep away from extreme results. ML assumes a vital part to distinguish the secret discrete examples and subsequently dissect the given information. After examination of information ML methods help in coronary illness forecast and early conclusion. This paper presents execution examination of different ML strategies, for example, Naive Bayes, Decision Tree, Logistic Regression and Random Forest for foreseeing coronary illness at a beginning phase.

II. AIM AND GOAL OF THE PROJECT

A. The aim of the project is to predict the patients having heart disease by collecting the reports from the hospitals (vector value).
B. Using Machine Learning algorithms , we do trial and testing of the datasets to get maximum accuracy.
C. The output is shown to the patients using graph with different analytics.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

III. LITERATURE REVIEW

Numerous studies have been conducted in medical centres employing various data mining approaches and machine learning algorithms to develop disease prediction systems.

Prediction of Heart Disease Using Multiple Regression Model was proposed by K. Polarajuet al, and it proved that Multiple Linear Regression is appropriate for predicting heart disease risk. The work is done with the help of training. The job is carried out with the use of a training data set that contains 3000 cases with the 13 different features described earlier. The data set is separated into two sections, with 70% of the data used for training and 30% for testing. Based on the findings, it is obvious that the Regression algorithm's classification accuracy is superior to that of other algorithms.

KStar, j48, SMO, and Bayes Net, as well as Multilayer perception using WEKA software, were used to produce heart disease prediction by Marjia et al. Using k-fold cross validation, SMO and Bayes Net outperform KStar, Multilayer Perception, and J48 approaches based on performance from multiple factors. The accuracy levels attained by those algorithms are still insufficient.

KStar, j48, SMO, and Bayes Net, as well as Multilayer perception using WEKA software, were used to produce heart disease prediction by Marjia et al.accuracy's performance is improved more to give better decision to diagnosis disease.

S. Seema et al. investigate strategies for predicting chronic disease by mining data from past health records utilising Nave Bayes, Decision Trees, Support Vector Machines (SVM), and Artificial Neural Networks (ANN). A comparison analysis of classifiers is conducted to determine which performs better on an accuracy rate.. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

Heart Disease Prediction System Using Data Mining Techniques was proposed by MeghaShahi et al. WEKA software is used in healthcare centres to provide automatic disease diagnosis and service quality ratings. SVM, Nave Bayes, Association rule, KNN, ANN, and Decision Tree were among the algorithms utilised in the study. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

Prediction and Analysis of the Occurrence of Heart Disease Using Data Mining Techniques was advocated by ChalaBeyene et al. The major goal is to predict the emergence of cardiac disease in order to make an early automatic diagnosis of the condition in a short amount of time. The proposed strategy is also important in healthcare organisations with experts who lack knowledge and expertise. It analyses a variety of physiological characteristics, such as blood sugar and heart rate, as well as age and sex, to determine whether or not a person has heart disease. WEKA software is used to compute dataset analyses.

R. Sharmila et al. proposed using a non-linear classification system to predict cardiac disease. For heart disease prediction using an optimal attribute set, it is proposed to employ bigdata techniques such as Hadoop Distributed File System (HDFS), Mapreduce, and SVM. This study looked into the usage of various data mining approaches for predicting cardiac disorders. It recommends using HDFS for storing big amounts of data across several nodes and running the SVM prediction algorithm on multiple nodes at the same time. SVM was applied in parallel, which resulted in a faster computation time than sequential SVM.

Jayami Patel et al. proposed employing data mining and machine learning algorithms to detect heart disease. The purpose of this research is to use data mining techniques to uncover hidden patterns. In comparison to LMT, the best algorithm J48 based on UCI data has the highest accuracy rate. Purushottam et al. presented a data-mining-based approach for predicting heart disease. This system aids medical practitioners in making informed decisions based on a set of parameters. A International Journal of Computer Applications(0975–8887) Volume 181– No.18, September 2018 It provides 86.3 percent accuracy in the testing phase and 87.3 percent in the training phase for 22 specific parameters.. K.Gomathi et al. proposed employing data mining approaches to predict many diseases. In today's world, data mining is critical for forecasting a variety of diseases. The number of tests can be lowered by employing data mining techniques. This research focuses on predicting heart disease, diabetes, and breast cancer, among other things. Heart disease prediction using ANN algorithm in data mining was proposed by P.Sai Chandrasekhar Reddy et al. There was a need to design a new method that could anticipate heart illness due to the rising costs of heart disease diagnosis. After evaluating numerous data such as heart rate, blood pressure, and cholesterol, a prediction model is utilised to anticipate the patient's health. The system's accuracy is demonstrated in Java. Data mining approaches and machine learning were recommended by Jaymin Patel et al. to predict cardiac disease. Predicting the cardiac system has two objectives. Assume any prior knowledge of the patient's records. To run against a huge number of records, the system must be scalar. WEKA software is used to implement the system. WEKA's classification tools and explorer mode are used for testing.

Boshra Brahmi et al, developed different data mining techniques to evaluate the prediction and diagnosis of heart disease. The main goal is to compare and contrast several classification algorithms such as J48, Decision Tree, KNN, SMO, and Nave Bayes.Following that, some performance in terms of accuracy, precision, sensitivity, and specificity is assessed and compared. The greatest strategy for predicting heart disease is J48 and decision tree.

Noura Ajam suggested using an artificial neural network to diagnose cardiac problems. Feed forward Back propogation learning techniques were employed to test the model based on their ability. Classification accuracy increased to 88 percent when appropriate function was taken into account, using 20 neurons in the buried layer. For the prediction of cardiac disease, ANN provides a substantial outcome.

S.Prabhavathi et al. introduced the Decision Tree Based Neural Fuzzy System (DNFS) technique for analysing and predicting various heart diseases. The studies on heart disease diagnostics is reviewed in this study. Decision tree based Neural Fuzzy System is abbreviated as DNFS. The goal of this study is to develop an intelligent and cost-effective system, as well as to improve the performance of the current system. Data mining techniques are specifically applied in this work to improve heart disease prediction. The findings of this study reveal that SVM and neural networks may accurately predict heart disease in a highly positive manner. Still, data mining approaches aren't promising when it comes to predicting heart disease..

To forecast cardiac disease, SairabiH. Mujawar et colleagues employed k-means and nave bayes. The purpose of this study is to develop a system that uses a historical cardiac database to provide diagnostics. The system was built with 13 attributes in mind. Data mining techniques such as clustering and classification methods can be used to extract knowledge from databases. From the Cleveland Heart database, 13 attributes with a total of 300 records were used.. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

Sharan Monica.L et colleagues proposed a cardiovascular disease analysis. This study presented data mining techniques for illness prediction. Its goal is to provide a survey of current strategies for extracting data from datasets that will be valuable to healthcare professionals. The time it takes to generate the decision tree for the system can be used to estimate performance.

To predict cardiac illness, Sharma Purushottam et al suggested c45 rules using a partial tree approach. This paper can uncover a set of criteria for predicting patient risk levels based on a specific health parameter. The outcomes can be measured in terms of accuracy classification, error classification, rules developed, and overall performance. Then, using C4.5 and a partial tree, a comparison was made. The The outcome demonstrates that there is potential for prediction and that it is more efficient. Table 2 depicts the accuracy of heart disease diagnosis using various methodologies, as indicated below.

IV. CONCEPTAND EVALUATION

A. *Existing System*

Detecting heart disease is a complex and dangerous process. Since it is directly related to people's wellbeing, precision is crucial. If not calculated correctly, this can be fatal. As a result, the emphasis of this study is on comparing various data mining methods for generating estimates. It is the product of a comparison of various approaches. To compare approaches, cross-validation error is used.

B. *Proposed System*

Using data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression, and Random Forest, the suggested work evaluates the probability of Heart Disease and classifies the patient's risk level. As a result, the performance of several machine learning algorithms is compared and contrasted in this work.

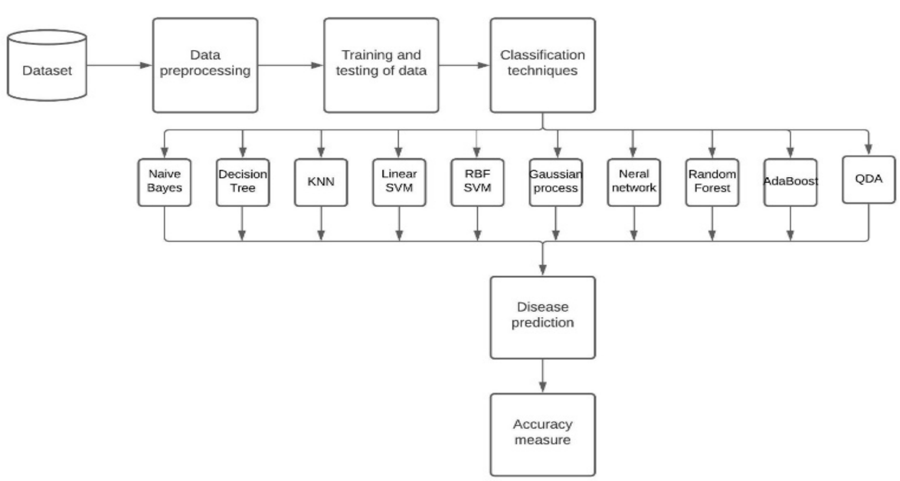


Fig-1:Basic Block diagram of the system.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

V. IMPLEMENTATION METHOD

A. Data Collection Module
The data was obtained via kaggle. Age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal as input columns and goal as output columns make up the dataset.

B. Data Preprocessing Module
Following the acquisition of various records, data on heart illness is pre-processed. A total of 5K patient records are included in the dataset, with some missing variables. These records are deleted from the dataset and utilised to pre-process the patient records that remain. Multiclass variables and binary classification are introduced for the properties of a given dataset. The presence or absence of cardiac disease is investigated using multi-class variables. The value is set to 1 in the case where the patient has heart illness; otherwise, the value is set to 0 to indicate that the patient does not have heart disease. Data is pre-processed by transforming medical records into diagnostic values. According to the results of data pre-processing for patient records, a value of 1 indicates the presence of heart disease, while a value of 0 indicates the absence of heart illness.

C. Model Building
Building machine learning models that are capable of generalizing on any future data requires a lot of prior consideration and accurate assumptions about the feature set and available training algorithms.
Steps involved in building a model:
1) Simplicity of the Model
2) Evaluate existing models
3) Design model structure and metrics
4) Relaunching and Reiterating

D. Training And Testing
The following are the data files that were used:
The training set is train.csv, whereas the test set is test.csv.

E. Validation
The classification model is built using the training data.Then the proposed model is tested using the test data here 20% of the data is used for testing.

F. Prediction
The predicted values are found using above classification models .Then based on the predicted values either 0 or 1 is predicted.The model is saved and various performance metrics is calculated.

G. Advantages
1) It's simple and straightforward to put into practise.
2) Approach to continuous optimization.
3) Without utilising any Bayesian procedures, the Naive Bayes model can be employed. Naive Bayes classifiers perform effectively in a variety of complicated real-world situations.

VI. CONCLUSION AND FUTURE SCOPE
This is the most accurate model for predicting heart disease patients. This project will provide answers to complex questions, each with its own strengths in terms of model interpretation, comprehensive knowledge access, and accuracy. This project has the potential to be improved and extended. Data mining techniques are used to present this project. KNN, Naive Bayes, Decision Tree, Random Forest, and other algorithms are used to design the method based on logistic regression. Random Forest produces better outcomes and supports domain experts and even medical professionals in planning for a better and earlier diagnosis for patients. This device functions brilliantly even without retraining. Determine each algorithm's prediction performance and apply the proposed system to the area that requires it. To increase the accuracy of algorithms, employ more relevant feature selection approaches. If a patient has been diagnosed with a specific type of cardiac disease, there are numerous therapy options available to them.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

In conclusion, according to the literature survey, only a marginal success is achieved in the creation of predictive models for heart disease patients, and thus there is a need for combinational and more complex models to increase the accuracy of predicting the onset of heart disease at an early stage. The technology will become increasingly sophisticated as more data is fed into the database. There are numerous enhancements that might be investigated in order to improve the scalability and accuracy of this prediction system. Due to a lack of time, the following research / work will have to be completed in the future. For future investigations, I'd like to explore other discretization strategies, multiple classifier voting techniques, and different decision tree kinds.

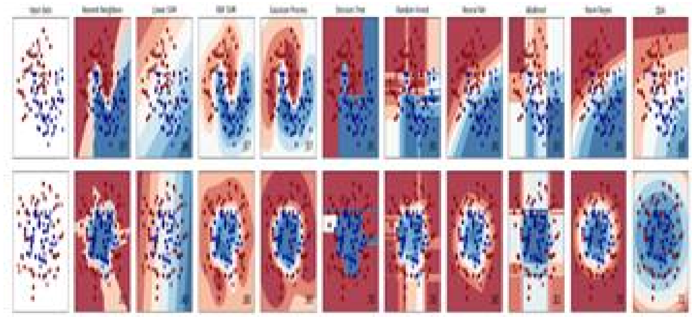
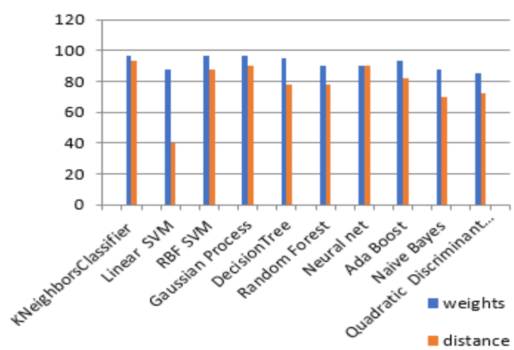


Fig-2:1O different classifier-percentage values.

To visualize the comparison model graphically, a bar graph has been plotted.



CLASSIFICATION ALGORITHM	ACCURACY (%)	
	WEIGHTS	DISTANCE
K Neighbors Classifier	97	93
Linear SVM	88	40
RBF SVM	97	88
Gaussian Process	97	90
DecisionTree	95	78
Random Forest	90	78
Neural network	90	90
Ada Boost	93	82
Naive Bayes	88	70
Quadratic Discriminant Analysis	85	72

Fig-3:Tabulated and graphical comparison of algorithms.

VII. ACKNOWLEDGEMENT

First and foremost, we thank Coimbatore Institute of Technology's Information Technology Department and Dr. R. Sandhiya, Associate Professor Coimbatore Institute of Technology's Information Technology Department. Dr. NK. Karthikeyan, Head of Department, Information Technology Department, Coimbatore Institute of Technology, deserves special thanks for his ongoing advice and assistance for our project effort.

REFERENCES

[1] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2018, pp.2137-2159.
[2] V.Krishnaiah, G.Narsimha, N.Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.
[3] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2017.
[4] T.Mythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, vol. 68, 16 April 2018.
[5] Nimai Chand Das Adhikari, Arpana Alka, and rajat International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 18, September 2018.Garg, "HPPS: Heart Problem Prediction System using Machine Learning".
[6] K.Polaraju, D.Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.
[7] Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold cross-validation", The Institute of Electrical and Electronics Engineers, March 2017.
[8] Ashok kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer, 17 September 2018.
[9] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466.
[10] Mr. ChalaBeyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.
[11] R.Sharmila, S.Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.
[12] Jayami Patel, Prof. TejalUpadhay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, pp.962-969."Efficient Heart Disease Prediction System", 2016,
[13] K.Gomathi, Dr.D.ShanmugaPriyaa, "Multi Disease Prediction using Data Mining Techniques", International Journal of System and Software Engineering, December 2016, pp.12-14.
[14] Mr.P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, "Heart Disease Prediction using ANN Algorithm in Data Mining", International Journal of Computer Science and Mobile Computing, April 2017, pp.168-172.
[15] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277-281.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

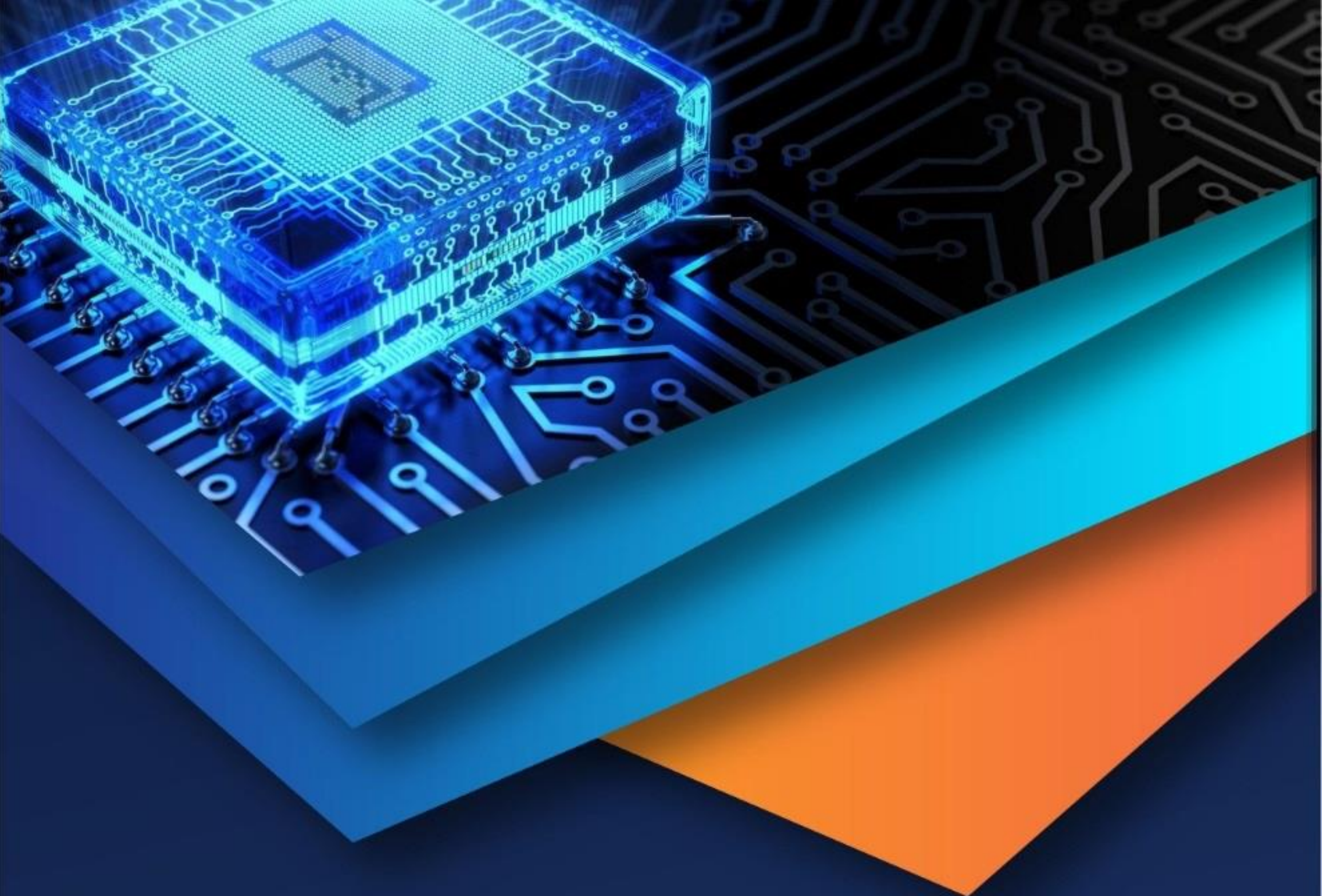
Techniques", Journals of Multidisciplinary Engineering

Dr.R.Sandhiya
Assistant Professor,
Coimbatore Institute of Technology
Coimbatore

Ms.Shivani A
B.Tech Information Technology
Coimbatore Institute of Technology
Coimbatore

Ms.Tejasvi K
B.Tech Information Technology
Coimbatore Institute of Technology
Coimbatore

Ms. Shalini J
B.Tech Information Technology
Coimbatore Institute of Technology
Coimbatore



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY