



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.524**

**Volume 14, Issue 1, January 2025**

# **Optimization Accuracy of Smart Healthcare for Heart Disease Diagnosis using Gradient Boosting Technique**

**Krishna Kaushal<sup>1</sup>, Dr. Ram Milan Chadhar<sup>2</sup>**

M. Tech. Scholar, Department of Electronics and Communication Engineering, Madhyanchal Professional University,  
Bhopal, India<sup>1</sup>

Head of Department, Department of Electronics and Communication Engineering, Madhyanchal Professional  
University, Bhopal, India<sup>2</sup>

**ABSTRACT:** Heart disease is a major contributor to global mortality, necessitating advanced diagnostic tools for timely detection and treatment. This research focuses on optimizing the accuracy of heart disease diagnosis within smart healthcare systems using the Gradient Boosting technique, a powerful machine learning algorithm known for its efficiency and predictive capabilities. By leveraging real-time health data from IoT devices and patient records, the study integrates feature engineering, hyperparameter tuning, and data balancing strategies to enhance model performance. The proposed Gradient Boosting model demonstrates superior accuracy, precision, and recall compared to traditional methods, making it a robust solution for heart disease diagnosis. In the current digital era, remote monitoring devices collect vast amounts of patient data, contributing to 30% of global data generated by the healthcare industry annually, with an expected increase of 36% by 2025. A regularized gradient boosting model for heart disease diagnosis is also developed that has provided 99.20% accuracy and 98.46% accuracy for RF on the 10016 heart dataset.

**KEYWORDS:** Heart Disease, Machine Learning, Accuracy

## **I. INTRODUCTION**

Cardiovascular diseases (CVDs) are associated with arrhythmias, vascular disorders, heart failure, heart attacks, strokes, and other heart diseases, and are one of the leading causes of death worldwide [1]. According to reports from several organizations [2], more than 17 million people died from these diseases in 2019, accounting for more than 30% of the total number of deaths worldwide in that year. Therefore, timely and accurate detection of diseases and identification of key risk factors are of great importance in healthcare, especially in CVD. Several risk prediction algorithms have been recommended, mainly using regression methods that integrate data on known risk factors [3]. However, these methods ignore the complexity and nonlinear characteristics of risk factors that have little or no interaction [4], which has a significant negative impact on the prediction of CVD. CVD includes factors related to blood pressure, cholesterol levels, blood sugar levels, lifestyle habits, and smoking, and can be controlled by taking medication and certain preventive measures. However, factors such as age, ethnicity, and family history of cardiovascular disease do not change with drug therapy. Therefore, to accurately predict cardiovascular disease, many factors must be considered, taking into account their complexity and nonlinearity. This justifies the use of artificial intelligence (AI) tools that help in predicting and classifying cardiovascular diseases. AI, machine learning (ML), and fuzzy logic play an important role in healthcare to effectively diagnose various diseases in patients. ML is an advanced tool that allows a system to automatically learn and improve based on the experience of the target system. Supervised and unsupervised ML algorithms are the most common algorithms. Supervised ML algorithms use the system's previous knowledge on new data. However, unsupervised ML algorithms use raw data that is neither classified nor labeled. ML techniques are more reliable and efficient in predicting cardiovascular disease than simple ML and regression methods [5, 6]. In the past decade, several ML algorithms have been proposed for predicting cardiovascular disease using different parameters, datasets, and approaches. For example, an ML approach has been proposed for predicting cardiovascular disease using body mass index (BMI) [7]. Various ML models such as decision trees, support vector machines (SVMs), artificial neural networks (ANNs), naive Bayes (NB), and random forests have been used for diagnosing cardiovascular disease [8]. Among these models, ANN showed the highest accuracy at 84.25%. Mental load

and other influences were statistically evaluated using analysis of variance (ANOVA) tests using electroencephalography (EEG) signals from 30 volunteers [9]. The predictive ability of ML algorithms using CVD data was evaluated in [11]. These authors found that SVMs may be better than others when the goal is to maximize a certain mathematical function with respect to a given dataset. Similarly, a comparison of the capabilities of ML methods found that SVMs can identify hidden patterns in complex medical data. Several researchers have used ML methods to validate their prediction frameworks. For example, ML methods were used to gain new insights into a dataset of over 40,000 Swedish patients with heart failure [10]. Using cluster analysis, the authors discovered four new disease phenotypes in this group.

## II. SYMPTOMS OF HD

Dyspnea is often caused by strenuous exercise in healthy, well-trained people, and by continued exercise in normal people who are not accustomed to exercise, so it should be considered abnormal if this symptom occurs during breaks in physical work and is suddenly brought on. It is associated with all Huntington's diseases. It occurs only when resting, and almost always not when working hard. The effective onset of dyspnea is when there is increased awareness of breathing, caused by increased pain in the cardiovascular peak area, or a delay of more than 2 hours, a dull chest pain, and difficulty in taking enough air into the lungs [9].

### Chest pain or discomfort

Chest pain is one of the most common symptoms of Huntington's disease, but it is difficult to determine whether it is caused by the heart or some other factor other than the heart. Chest pain is also one of the symptoms. People assume that chest pain is heart failure. If chest pain is caused by other disorders, heart disease is the most common and dangerous, but heart pain itself can be treated. It is used to describe irritation, pressure, tightness, numbness, and other discomfort in the chest, neck, and upper trunk, along with pain in the jaw, head, arms, etc. It can last for a few minutes, a few days, or even a week, depending on your expectations. The meaning of chest pain is very unclear, and there are some medical cases of HD that cause symptoms [10].

Syncope is described as "a loss of consciousness that causes a decrease in blood flow to the brain". Since around 2017, syncope has been defined as "a sudden, temporary loss of consciousness" and is a common symptom that many people experience only once in their lifetime and is not a serious illness. However, syncope can represent a dangerous, life-threatening condition. If syncope occurs, it must be diagnosed and treated. The causes of syncope are divided into two levels: neurological, metabolic, vasomotor, and cardiac. Syncope is a disease that causes sudden death. It can be caused by a variety of conditions that change the heartbeat. Palpitations, also known as "missing beats" or a fast, irregular heartbeat, are a common symptom. Some people with palpitations have arrhythmias, which are considered abnormal heart rhythms. Regardless of the cause, there are many different types of arrhythmias that can cause palpitations [11]. These side effects are normal for any illness. Weakness may be expected as an inability to function normally. Somnolence indicates lack of sleep or insomnia. Patients often fall asleep suddenly during the day. Indicative of HD, slurred displays indicate dysfunction of various organs of the body. Similarly, dizziness, lack of energy, fatigue, and lethargy usually require treatment to identify the specific cause. Disorders that cause nighttime sleep, such as restless legs syndrome and insomnia, are called somnolence [12].

## III. PROPOSED METHODOLOGY

Gradient Boosting is a powerful machine learning technique that builds predictive models by combining the strengths of multiple weak learners, typically decision trees, to create a strong learner. The algorithm works iteratively, where each new tree is trained to correct the errors or residuals of the previous trees. By minimizing a specific loss function, Gradient Boosting ensures that the model focuses on the most challenging aspects of the data at each iteration. The "gradient" in Gradient Boosting refers to the gradient descent optimization technique, which guides the model to minimize the error by making small, incremental improvements. This approach makes Gradient Boosting highly effective in handling both regression and classification problems. With its ability to handle non-linear relationships and complex datasets, Gradient Boosting has become a preferred choice for applications requiring high accuracy, such as heart disease diagnosis. Additionally, advanced implementations like XGBoost, LightGBM, and CatBoost have introduced efficiency and scalability, making Gradient Boosting suitable for large-scale real-world problems.

The Boost algorithm, together with the weak ML scheme from a boosted learning technique, is suggested in this chapter to sustain optimum classification accuracy and robustness. Adaptive boosting can be utilized with a supervised ML approach for maximum efficiency after monitoring the output and undertaking a performance evaluation. With essentially no tuning limits, Boost is speedy, straightforward, and easy to program. It is adaptable since it may be used in conjunction with the SVM, neural networks, random forest and other algorithms.

It is one of the most used algorithms for boosting for adapting to problem-solving activities. Gradient boost combines numerous "poor classifiers" into a single "strong classifier." It places a higher priority on difficult-to-characterize events and less on those that are already well-managed. Gradient Boost is a very efficient algorithm that can compete with the best in the industry, if not better than.

This Classifier, on the other hand, by painstakingly fitting several noisy sets of data until every data point in the training set matched without error, it defies simple statistical norms. Perhaps more amazing, at least to statisticians, is that it will continue to improve a strategy that already removes generalization mistake. Gradient Boost takes a statistical approach to boosting, it's a step-by-step optimization of an exponential failure that requires tree regularization and step number management.

#### Algorithm for Proposed Methodology:-

Step 1: Start

Step 2: Import Library: define all loading dataset, visualization, data preprocessing, data splitting, confusion matrix, machine learning and accuracy library

Step 3: Upload Dataset (files .upload command)

Step 4: Stored in a Pandas Data frame (pd.read\_csv command)

Step 5: Pre-processing

Step 6: Variable (input variable i.e. age, sex, cp, chol, bp)

(6.1) if

(6.2) (Input variable >= 2)

(6.3) end if;

Step 7: Classes; 0 for not disease and 1 for disease

Step 8: Applied Machine learning Technique

Step 9: Create Confusion Matrix

Step 10: Threshold (T), maximum value of T is 100

(10.1) if

(10.2) (Accuracy >= T)

(10.3) end if;

Step 11: Calculate Parameter

Step 12: End

## IV. SIMULATION RESULTS

### Simulation Parameter

The accuracy can be measured according to Eq. 1

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (1)$$

For a heart classification problem, its measures include Precision-Recall and accuracy. The formula to derive these measures is given in Eq. 2 and Eq. 3.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (4)$$

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig. 1: Heart Dataset

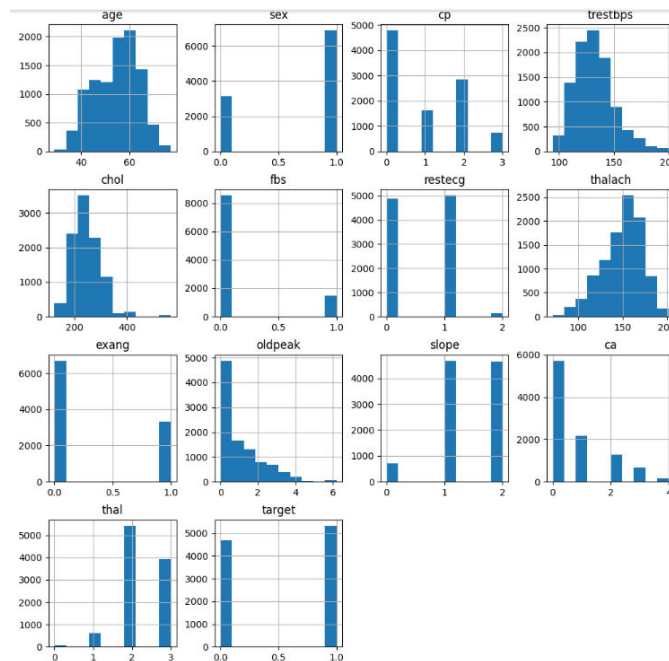


Fig. 2: Histograms of HD Dataset

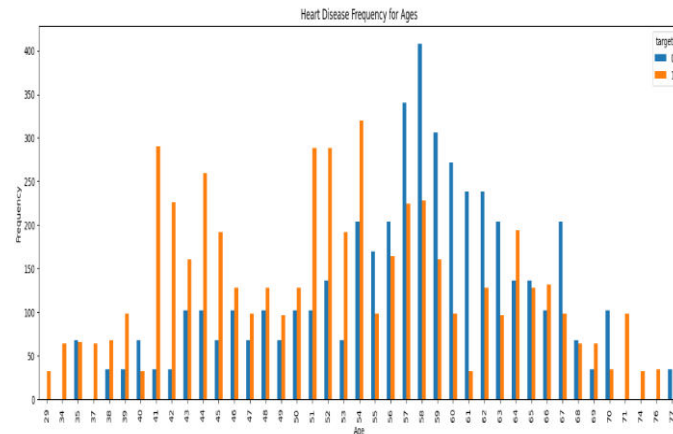


Fig. 3: Frequency of HD Dataset

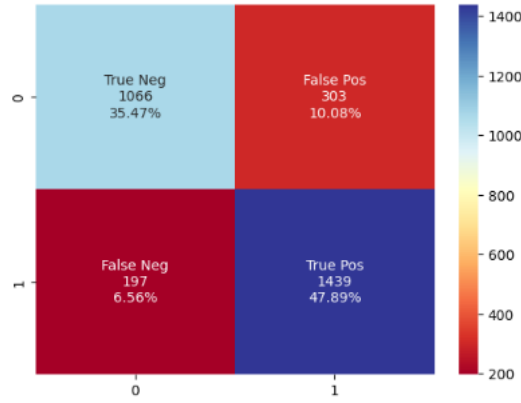


Fig. 4: Confusion Matrix for Navies Bayes

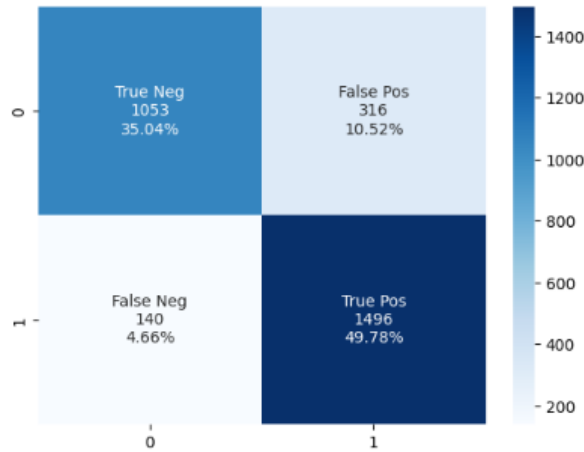


Fig. 5: Confusion Matrix for Logistic Regression

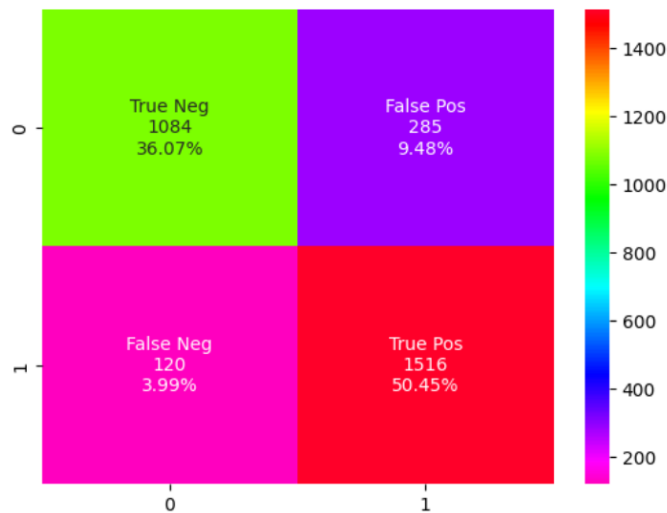


Fig. 6: Confusion Matrix for K-NN

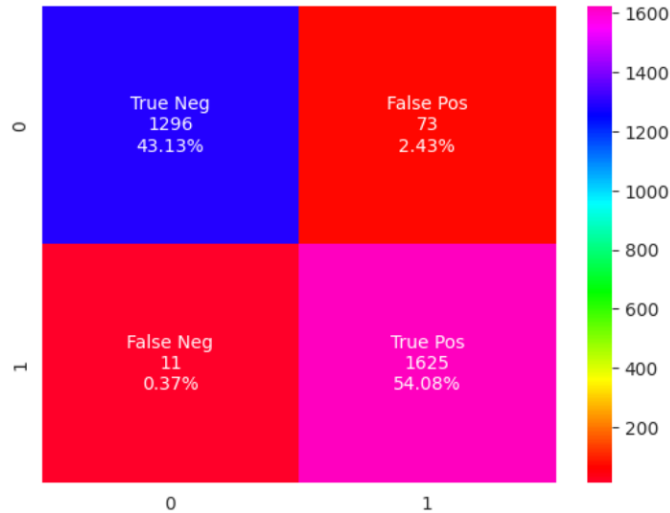


Fig. 7: Confusion Matrix for Gradient Boosting

Model	Accuracy
Logistic Regression	84.82%
Navies Bayes	83.36%
Random Forest	98.46%
K Nearest Neighbors	86.65%
Decision Tree	96.07%
Gradient Boosting	99.20%

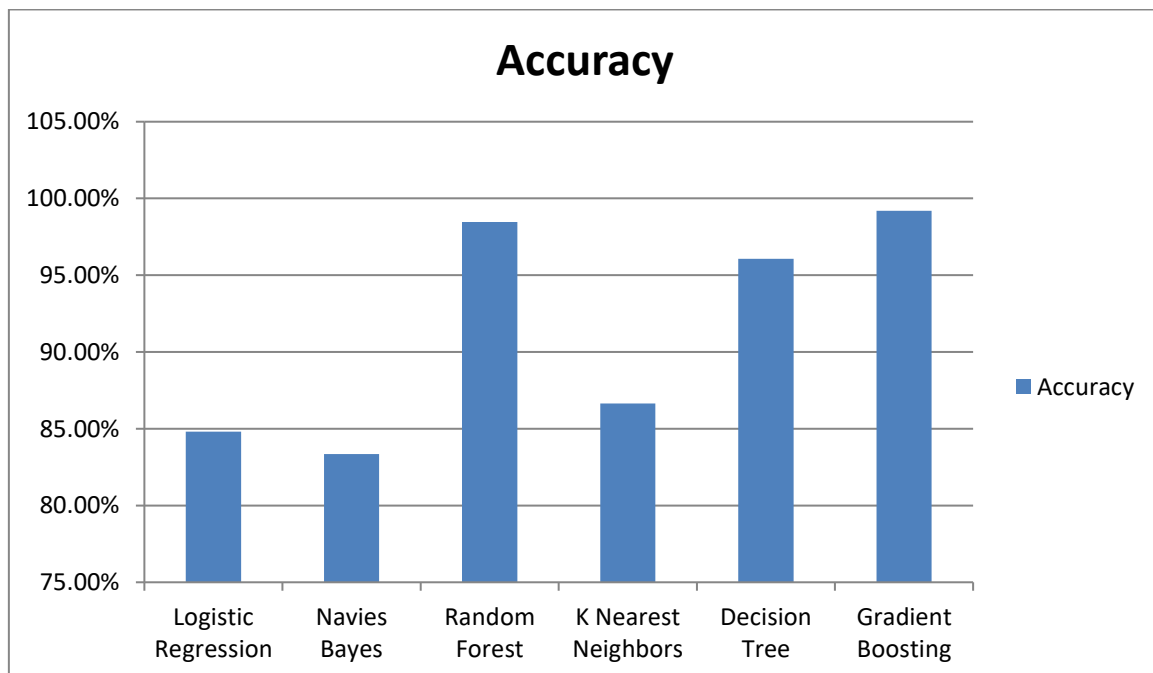


Fig. 8: Graphical Accuracy

## V. CONCLUSION

The Gradient Boosting technique has emerged as a highly effective method for heart disease diagnosis within smart healthcare systems. Its ability to iteratively refine predictions by addressing residual errors makes it a robust choice for handling complex, multidimensional datasets commonly encountered in healthcare. By leveraging real-time health data, this approach not only enhances diagnostic accuracy but also supports early detection and personalized treatment plans, contributing to improved patient outcomes.

This research highlights the potential of Gradient Boosting to optimize accuracy in heart disease prediction, demonstrating its superiority over traditional methods. However, challenges such as computational complexity, data privacy concerns, and model interpretability must be addressed to ensure practical deployment. Future work should focus on integrating explainable AI, real-time data processing, and scalable architectures to further enhance the role of Gradient Boosting in smart healthcare systems. These advancements can pave the way for more reliable and efficient healthcare solutions, ultimately revolutionizing the early diagnosis and management of heart disease.

## REFERENCES

- [1]O. Taylan, A. Alkabaa, H. Alqabbaa, E. Pamukçu and V. Leiva, "Early prediction in classification of cardiovascular diseases with machine learning neuro-fuzzy and statistical methods", *Biology*, vol. 12, no. 1, pp. 117, 2023.
- [2]R. Bhavani, V. Ramkumar, V. Ravindran, R. Sindhuja and K. Swaminathan, "An efficient SAR image detection based on deep dense-mobile net method", *7th International Conference on Computing in Engineering & Technology (ICCET 2022)*, vol. 2022, pp. 92-95, 2022, February.
- [3]I. Sutudja, "Descriptive and predictive analysis on heart disease with machine learning and deep learning", *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1-6, 2021, October.
- [4]Karna Vishnu Vardhana Reddy, Irraivan Elamvazuthi, Azrina Abd Aziz, Sivajothi Paramasivam and Hui Na Chua, "Heart Disease Risk Prediction using Machine Learning with Principal Component Analysis", *International Conference on Intelligent and Advanced Systems (ICIAS)*, pp. 01-05, IEEE 2021.
- [5]P. Rani, R. Kumar, N. Ahmed and A. Jain, "A decision support system for heart disease prediction based upon machine learning", *Journal of Reliable Intelligent Environments*, vol. 7, pp. 263-275, 2021.
- [6]M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, P. Singh and N. Kumar, "Latest trends on heart disease prediction using machine learning and image fusion", *Materials Today: Proceedings*, vol. 37, pp. 3213-3218, 2021.
- [7]M. Ganesan and Dr. N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models", *International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 01-05, IEEE 2019.
- [8]Priyan Malarvizhi Kumar, Usha Devi Gandhi, "A novel Internet of Things architecture with machine learning algorithm for early detection of heart diseases", *Computers and Electrical Engineering*, Vol.65, pp. 222–235, 2018.
- [9]Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A., "A comprehensive survey on support vector machine classification: Applications, challenges and trends", *Neurocomputing*, Vol. 408, pp. 189–215, 2020.
- [10]A. Javeed, S. Rizvi, S. Zhou, R. Riaz, S. Khan and S. Kwon, "Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification", *Mobile Information Systems*, pp. 1-11, 2020.
- [11]C. Latah and S. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques", *Informatics in Medicine Unlocked* 16, pp. 100203, 2019.
- [12]C. S. M. Wu, M. Badshah and V. Bhagwat, "Heart disease prediction using data mining techniques", *Proceedings of the 2019 2nd international conference on data science and information technology*, pp. 7-11, 2019, July.
- [13]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, March.
- [14]Amin Khatami and Abbas Khosravi, "Medical image analysis using wavelet transform and deep belief networks", *Journal of Expert Systems with Applications*, Vol. 3, Issue 4, pp. 190–198, 2017.
- [15]T. Shu, B. Zhang and Y. Tang, "Effective Heart Disease Detection Based on Quantitative Computerized Traditional Chinese Medicine Using Representation Based Classifiers", *Evidence-Based Complementary and Alternative Medicine* 2017, pp. 1-10, 2017.





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details