iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51836

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 11 Issue V May 2023- Available at www.ijraset.com

Heart Disease Prediction Using Improved Squirrel Search Algorithm

S. Rajalakshmi[1], R. Ragulraja[2], K. Kalvi[3], S. Dhinesh Kumar[4], B. Murugan[5]
M. Tech, Assistant Professor / CSE, [2, 3, 4, 5]Final Year / CSE, Achariya College of Engineering Technology, Puducherry

Abstract: ***In recent years, the volume in globally recognized medical data sets are increasing both with attributes and number of records. Machine learning algorithms aiming to detect and diagnose ischemic heart diseases requires high efficacy and judgment. The state of art ischemic heart disease data sets present several issues, including feature selection, sample size, sample imbalance, and lack of magnitude for some characteristics etc. The proposed study is primarily concerned with improving feature selection and reducing the number of features yet giving better decisions.***
Index Terms: SSA, Optimization, RF, Ischemic Heart Disease, Feature Selection.

I. INTRODUCTION

Ischemic heart disease (IHD) is a leading cause of death in India, accounting for a significant proportion of all deaths each year. According to recent estimates, the total number of deaths due to IHD in India is around 1.6 million per year. Predicting the death rate of ischemic heart disease (IHD) is a challenging task that requires the analysis of various risk factors, including age, gender, lifestyle habits, medical history, and clinical parameters such as blood pressure, cholesterol levels, and electrocardiogram (ECG) readings [1]. Several statistical and machine learning techniques can be used to develop predictive models for IHD mortality, including logistic regression, decision trees, random forests, and support vector machines (SVMs) [2]. These models can be trained on large datasets of patients with IHD to identify the most important risk factors and their relationship to mortality. Heart disease prediction is a complex problem that involves the analysis of various factors, including medical history, lifestyle habits, and genetic predisposition. The Improved Squirrel Search Algorithm (ISSA) is a metaheuristic optimization technique that is inspired by the behaviour of squirrels [3][4]. To use ISSA for heart disease prediction, you could start by identifying the relevant features or risk factors that may contribute to the development of heart disease, such as age, gender, smoking status, blood pressure, cholesterol levels, and family history. You would then need to collect data on these factors for a large sample of individuals who have or have not been diagnosed with heart disease. Next, you could apply the ISSA algorithm to search for the optimal combination of risk factors that can predict heart disease with high accuracy [5][6]. ISSA could be used to identify the most important features and to optimize the weights or coefficients associated with each feature in a predictive model. To implement the ISSA algorithm for heart disease prediction, you would need to define the search space and the fitness function. The search space would include all possible combinations of risk factors and their associated weights or coefficients. The fitness function would evaluate the performance of each combination in terms of its ability to predict heart disease based on the available data. Once ISSA has identified the optimal combination of risk factors and weights, you could use this information to develop a predictive model that can be applied to new patients to predict their risk of developing heart disease. This model could be further refined and validated using additional data and statistical methods. In summary, ISSA could be a promising approach for heart disease prediction, but it would require careful implementation and validation to ensure that the resulting predictive model is accurate and reliable.

II. LITERATURE SURVEY

Jijesh et al. [7] created a supervised learning-based decision support system for multisensory healthcare data from wireless body sensor networks. The SSA is used to select features. Other methodologies such as Conventional Neural network, Deep Belief Network, and Artificial Neural Network are compared to the proposed M-DBN (modified deep belief network) technique. The proposed method outperforms all other techniques currently in use. Based on data mining techniques, Shan et al. [8] suggested a more accurate and practical risk prediction system. To reduce dimensionality, CFS Subset Evaluation was used with the Best-First-Search approach to picking key characteristics. The data is taken from the Cleveland Heart Disease Database and PKU People's Hospital's Cardiology inpatient dataset. SVM takes twice as long as the random forest classifier. To pick salient aspects of heart disease, an imperialist competitive algorithm with a meta-heuristic approach is proposed in this paper [9]. In comparison to other optimization algorithms, this approach can deliver an additional best solution for FS toward genetics. The categorization is also done using the K-nearest neighbour approach.

©IJRASET: All Rights are Reserved

The efficiency of the FS method improved by applying the proposed algorithm, according to the evaluation results. The data for these tests came from the UCI ML repository's HD data and the Tehran Shahid Rajaei hospital. The method proposed by Fuad et al. [10] is feature optimization, which aims to find the better variable size n-gram attributes for the supervised ML, also known as ''discrete weights-based n-gram feature selection.'' Pre-processing, appropriate attribute selection, attribute FS, and the classification approach used to predict the extent of the disease in a specific patient's medical records are all things to think about. An Internet of Medical Things framework for the analysis of HD based on modified salp swarm optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS) is presented to increase prediction accuracy. The Levy flight method, the proposed MSSO-ANFIS improves search abilities. For all iterations, the suggested Levy-based crow search algorithm for FS acquired the maximum fitness values. When compared to existing approaches, the suggested MSSO-ANFIS strategy provides superior F1-score, precision, accuracy, and recall as well as the lowermost classification error [11].

III. EXISTING SYSTEM

Filtering metrics are used in hybrid techniques to minimize the computational complexity of wrapper algorithms, and they have been shown to provide better feature subsets. Although filtering metrics select features based on their importance, the majority of them are insecure and subjective toward the metric in question.

- *Disadvantages:* This existing system results in poor robustness and efficiency due to selection of features based on importance. Accuracy is not up to the level due to feature selection.

IV. PROPOSED SYSTEM

The proposed studies only focus on improving the selection process and reducing the number of features while enabling better decisi on making. In this study, a meta-
heuristic approach was used to refine the squirrel search optimization algorithm to select salient features of heart disease. A compara tive study of the proposed Ischemic Heart Disease Squirrel Search Optimization (IHDSSO) model with a random forest classifier fo r better selection.

- *Advantage:* The Proposed system results in better robustness and efficiency. Improving the accuracy by using the proposed model.

V. ARCHITECTURE

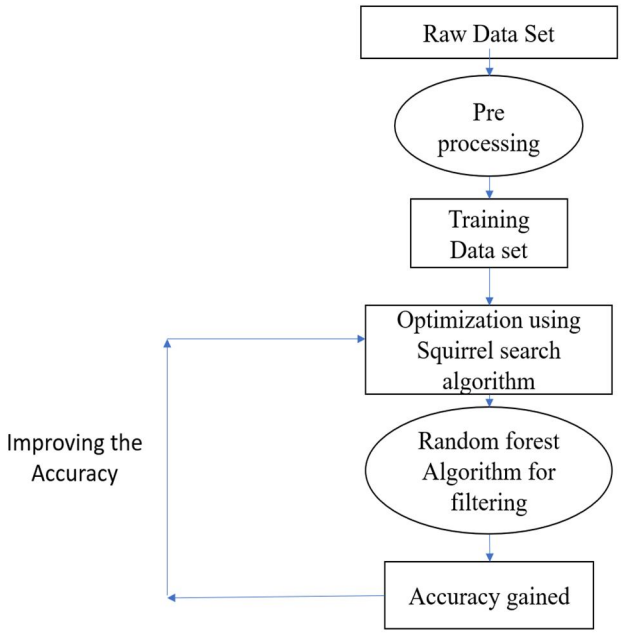


Fig.V.1 Architecture Diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

VI. MATERIALS AND METHODS

A. *Data Pre-Processing*

The dataset is a collection of connected data, with a report for each instance based on the data it represents, and an attribute for each attribute in the dataset. This dataset from the Kaggle for Ischemic Heart Disease Prediction.

Ischemic Heart Disease Attributes

Attribute	Description
cp	Chest pain type
trestbps	Resting blood sugar
Chol	Serum cholesterol
Oldpeak	ST depression induced by exercise relative to rest
Slope	Slope of peak exercise ST segment

Table IV.1 Attributes and Description

B. *Feature Selection And Reduction*

Among 5 attributes, the attributes which are used to identify the personal information are removed like age, sex and the remaining attributes are considered as they are important in finding the heart disease.

VII. OVERVIEW OF ALGORITHMS

In machine learning, random forest is the most powerful and widely used algorithm. It belongs to supervised machine learning [12]. It is used for both classification and regression problems in Machine learning. The process of random forest is: It collects the information. It builds decision trees on different samples. It takes the average of the decision trees It can handle the dataset containing categorical variables but compared to a single decision tree it is slower. Doesn't handle missing values.

A. *The Improved Squirrel Search Algorithm*

This section presents an improved squirrel search optimization algorithm by introducing four strategies to enhance the searching capability of the algorithm. In the following, the four strategies will be presented in detail.

1) *An Adaptive Strategy of Predator Presence Probability:* When flying squirrels generate new locations, their natural behaviours are affected by the presence of predators and this character is controlled by predator presence probability [17].
2) *Flying Squirrels' Random Position Generation Based on Cloud Generator:* The foraging behaviour of flying squirrels has the characteristics of randomness and fuzziness. These characteristics can be synthetically described and integrated by a normal cloud model [13]. In the model, a normal cloud model generator instead of uniformly distributed random functions is used to reproduce new location for each flying squirrel.
3) *A Selection Strategy between Successive Positions:* When new positions of flying squirrels are generated, it is possible that the new position is worse than the old one. This suggests that the fitness value of each individual needs to be checked after the generation of new positions by comparing with the old one in each iteration. If the fitness value of the new position is better than the old one, the position of the corresponding flying squirrel is updated by the new position [14].
4) *Enhance the Intensive Dimensional Search:* The main drawback of this procedure is that different dimensions are dependent and the change of one dimension may have negative effects on others, preventing them from finding the optimal variables in their own dimensions [19]. To further enhance the intensive search of each dimension, the following steps are taken for each iteration: (i) find the best flying squirrel location; (ii) generate one more solution based on the best flying squirrel location by changing the value of one dimension while maintaining the rest dimensions; (iii) compare fitness values of the new-generated solution with the original one, and reserve the better one; (iv) repeat steps (ii) and (iii) in other dimensions individually

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

B. Random Forest

In machine learning, random forest is the most powerful and widely used algorithm [15][16]. It belongs to supervised machine learning. It is used for both classification and regression problems in Machine learning.

The process of random forest is:

1) It collects the information
2) It builds decision trees on different samples
3) It takes the average of the decision trees

It can handle the dataset containing categorical variables but compared to a single decision tree it is slower. Doesn't handle missing values.

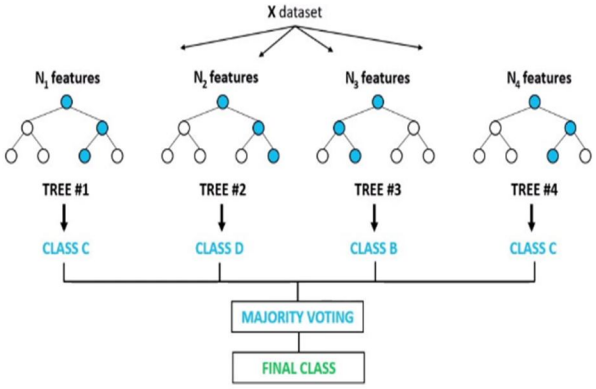


Fig. VII.1 RF algorithms decision tree

VIII. METHODOLOGY

''Age, Sex, Cp, Trestbps, Chol, Fbs, Restecg, Thalach, Exang, Oldpeak, Slope, Ca, and Thal'' are selected from the preprocessed data set of HD. This feature information has some not null values and unnecessary details. By using data preprocessing techniques, we further make the features with more specific data. Along these features, we have to find, which of them, give more accurate prediction of ischemic heart disease. [20] For feature selection process, we used Squirrel search algorithm. It sets the lower bound and upper bound limits. It is very useful for optimization of the features. According to these algorithm concept, it generates random location for the n number of squirrels, by that way, it generates random features to be selected within the limit concession. It evaluates the more accurate fitness value for every selected feature. It sorts the feature of heart disease in ascending order depending upon the fitness value [18]. It will provide a more accurate classification result of the classification of ischemic heart disease. The performance of the Squirrel Search algorithm evaluated by the metrics of accuracy, sensitivity, specificity. The model evaluation will be done by using Random Forest Algorithm. It aggregates a greater number of Decision Trees to make more accuracy of the model**.**

IX. CONCLUSION

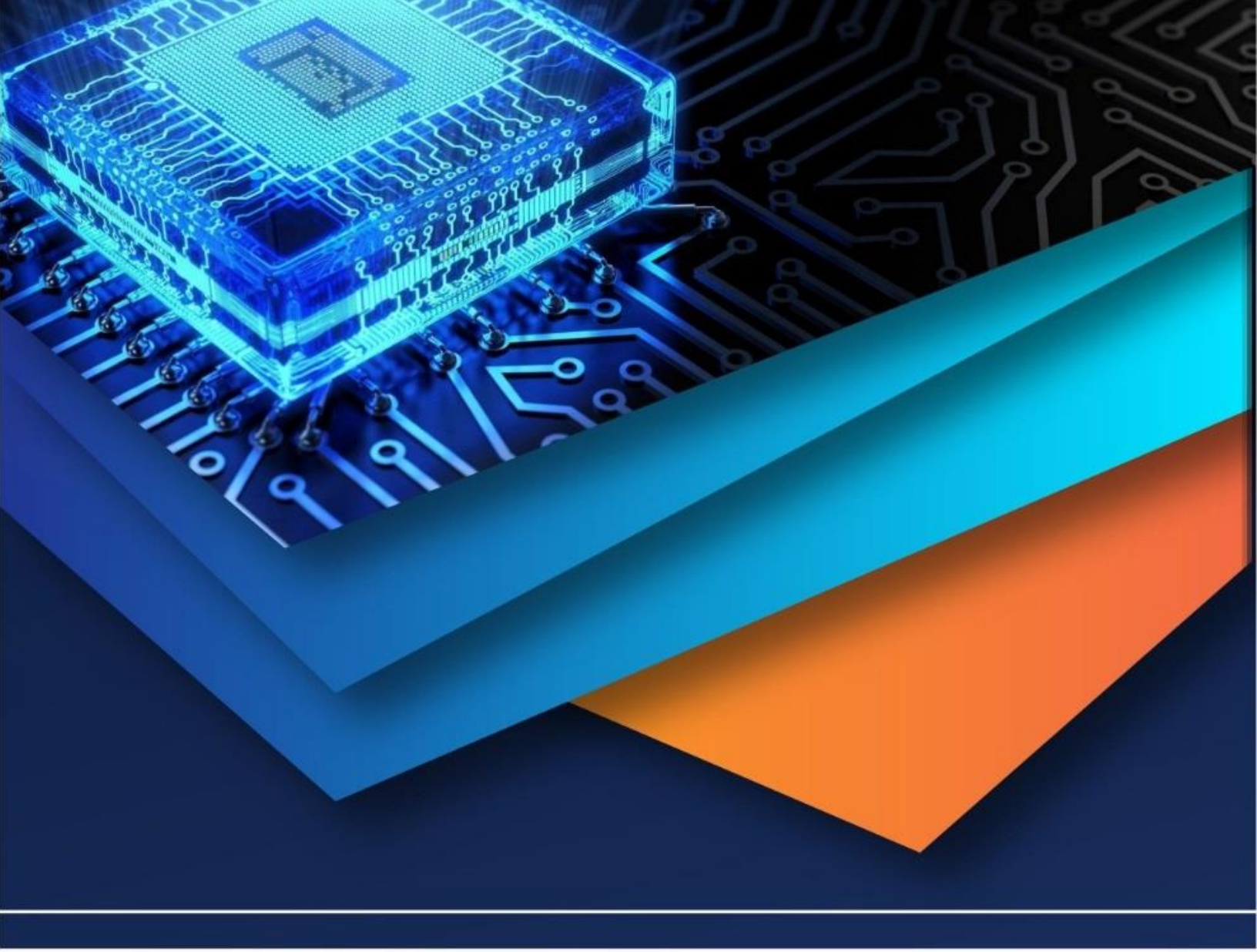
Ischemic Heart Disease Squirrel Search Optimization feature selection algorithm is verified considering the UCI heart disease dataset. The proposed IHDSSO model classifies the most essential features which could be used as a strong predictor for heart disease classification. In the study, we verified that the features like ''Cp'', "trestbps'', ''oldpeak'', ''Ca'', and ''thal" are more essential for the prediction of heart disease. The proposed IHDSSO model in conjunction with a random forest classifier ensured accuracy of over 98.38% and optimal heart disease prediction. The proposed IHDSSO model will be incredibly helpful in supporting healthcare solutions for diagnosing ischemic heart disease.

REFERENCES

[1] Khan. M. A (2020). An IoT framework for heart disease prediction based on MDCNN classifier, in IEEE Access, vol. 8, (pp. 34717–34727).
[2] Khan.M.A, M. T. Quasim, N. S. Alghamdi, and M. Y. Khan. (2020). A secure framework for authentication and encryption using improved ECC for IoTbased medical sensor data, in IEEE Access, vol. 8, (pp. 52018–52027).

[3] Mohan.S, Thirumalai.C, and Srivastava.G. (Jan. 2019). Effective heart disease prediction using hybrid machine learning techniques, in IEEE Access, vol. 7, (pp. 81542–81554).
[4] Soundharya.R, Cenitta.D, and Arjunan.R.V. (2018). Information concealment and redemption through data anonymization technique, in J. Adv. Res. Dyn. Control Syst., vol. 10, no. 7, (pp. 22–26).
[5] World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseas%es-(cvds)
[6] Yogaamrutha.S.C, Cenitta.D, and Arjunan.R.V. (2019). Forecast of coronary heart disease using data mining classification technique, in J. Adv. Res. Dyn. Control Syst., vol. 11, no. 4, (pp. 25–36).
[7] Jijesh. J. J. (Feb. 2021). A supervised learning based decision support system for multisensor healthcare data from wireless body sensor networks, in Wireless Pers. Commun., vol. 116, no. 3, (pp. 1795–1813).
[8] Xu. S, Zhang. Z, Wang. D, Hu. J, Duan. X, and Zhu.T. (Feb. 2021). Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework, in Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA), (pp. 228–232).
[9] Nourmohammadi-Khiarak. J, Feizi-Derakhshi M.-R, Behrouzi. K, Mazaheri. S, Zamani-Harghalani. Y, and Tayebi R.M. (Feb. 2021). New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection, in Health Technol., vol. 10, no. 3, (pp. 667–678).
[10] Al-Yarimi F.A.M, Munassar N. M. A, Bamashmos M. H. M, and Ali. M. Y. S. (Feb. 2021). Feature optimization by discrete weights for heart disease prediction using supervised learning, in Soft Comput., vol. 25, no. 3, (pp. 1821–1831).
[11] Khan M. A and Algarni. F. (2020). A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSOANFIS, in IEEE Access, vol. 8, (pp. 122259–122269).
[12] Cenitta D, Arjunan R. V, and Prema K. (2020). Missing data imputation using machine learning algorithm for supervised learning, in Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), (pp. 1–5).
[13] Ayyarao T. S. L. V, Ramakrishna N. S. S, Elavarasan R. M, Polumahanthi N, Rambabu M, Saini G, Khan. B, and Alatas. B. (2020). War strategy optimization algorithm: A new effective Metaheuristic algorithm for global optimization, in IEEE Access, vol. 10, (pp. 25073–25105).
[14] Gao. L and Ding. Y. (2020). Disease prediction via Bayesian hyperparameter optimization and ensemble learning, in BMC Res. Notes, vol. 13, no. 1, (pp. 1–6).
[15] Nagarajan. G and Babu. L. D. D. (2020). A hybrid feature selection model based on improved squirrel search algorithm and rank aggregation using fuzzy techniques for biomedical data classification, in Netw. Model. Anal. Health Informat. Bioinf., vol. 10, no. 1, (pp. 1–29).
[16] Jain. M, Singh. V, and Rani. A. (2020). A novel nature-inspired algorithm for optimization: Squirrel search algorithm, in Swarm Evol. Comput., vol. 44, (pp. 148–175).
[17] Wang. Y. and Du. T. (2020). An improved squirrel search algorithm for global function optimization, in Algorithms, vol. 12, no. 4, (p. 80).
[18] Alatas B. and Bingol. H. (2020). A physics based novel approach for travelling tournament problem: Optics inspired optimization, inInf. Technol. Control, vol. 48, (pp. 373–388).
[19] Alatas B. and Bingol. H. (Dec. 2020). Comparative assessment of light-based intelligent search and optimization algorithms, in Light Eng., vol. 28, no. 6, (pp. 51–59).
[20] Cenitta D, Arjunan R. V, and Prema K. (Sep. 2021). Cataloguing of coronary heart malady using machine learning algorithms, in Proc. 4th Int. Conf. Electr., Comput. Commun. Technol. (ICECCT), (pp. 1–6).

©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)