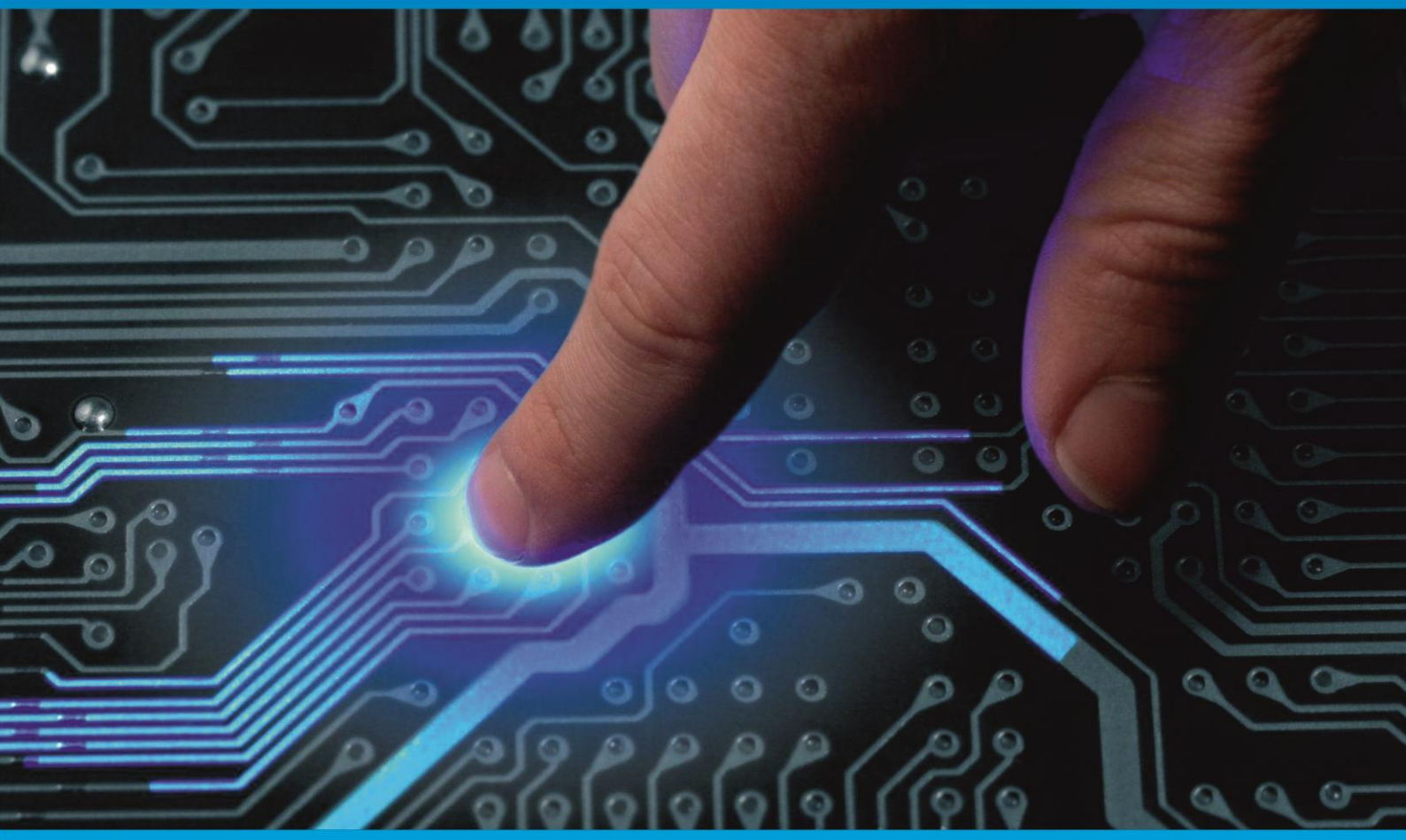
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 10, October 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

A Review of Prediction of Heart Diseases using Machine Learning Approaches

Kehkasha Khan[1], Prof. Susheel Kumar Gupta[2]

M.Tech Scholar, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India[1]

Assistant Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India[2]

ABSTRACT: Heart disease (HD) is one of the most common diseases nowadays, and an early diagnosis of such a disease is a crucial task for many health care providers to prevent their patients for such a disease and to save lives. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. Therefore classify the healthy people and people with heart disease, noninvasive-based methods such as machine learning are reliable and efficient. This paper studied a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Various results show that each technique has its unique strength to prediction parameters.

KEYWORDS: Prediction, Data mining, Heart Disease, Machine Learning, Decision Trees, Naïve Bayes.

I. INTRODUCTION

The heart disease (HD) has been considered as one of the complex and life deadliest human infections on the planet. In this infection, ordinarily the heart can't push the necessary measure of blood to different pieces of the body to satisfy the typical functionalities of the body, and because of this, eventually the cardiovascular breakdown happens [1]. A significant test confronting medical care associations (clinics, clinical focuses) is the arrangement of value administrations at moderate expenses. Quality help suggests diagnosing patients accurately and directing medicines that are compelling. Poor clinical choices can prompt deplorable results which are thusly unsuitable. Clinical choices are frequently made dependent on specialists' instinct and experience instead of on the information rich information covered up in the information base. This training prompts undesirable inclinations, mistakes and exorbitant clinical costs which influence the nature of administration gave to patients. In the vast majority of the scientists study, coordination of clinical choice support with PC based patient records could lessen clinical mistakes, upgrade persistent wellbeing, decline undesirable practice variety, and improve understanding result [2]. Most medical clinics today utilize a type of medical clinic data frameworks to deal with their medical services or patient information [3]. Sadly, this information are infrequently used to support clinical dynamic.

The pace of coronary illness in the US is high [2]. The manifestations of coronary illness incorporate windedness, shortcoming of actual body, swollen feet, and weariness with related signs, for instance, raised jugular venous weight and fringe edema brought about by practical heart or noncardiac irregularities [3]. The examination methods in beginning phases used to distinguish coronary illness were convoluted, and its subsequent unpredictability is one of the significant reasons that influence the norm of life [4]. The coronary illness finding and treatment are exceptionally unpredictable, particularly in the agricultural nations; because of the uncommon accessibility of symptomatic contraption and lack of doctors and others assets which influence appropriate forecast and treatment of heart patients [5]. The precise and legitimate conclusion of the coronary illness danger in patients is fundamental for lessening their related dangers of extreme heart issues and improving security of heart [6]. The European Culture of Cardiology (ESC) announced that 26 million grown-ups overall were determined to have coronary illness and 3.6 million were analyzed each year. Roughly half of coronary illness individuals experiencing HD kick the bucket inside beginning 1-2 years, and concerned expenses of coronary illness the executives are around 3% of medical services budgetary spending plan [7].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

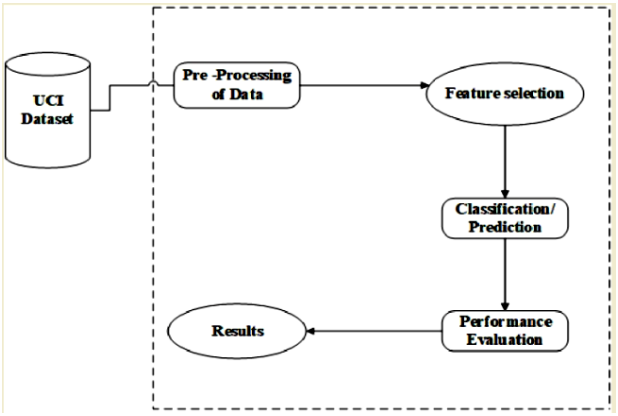


Figure 1: Basic process of Prediction

Figure 1 shows the basic operation of heart disease prediction using classification or data mining approach.

II. LITERATURE REVIEW

S. Mohan, et al.,[2019] Coronary illness is one of the most imperative explanations behind mortality on the planet today. Desire for cardiovascular infection is a fundamental test in the region of clinical data examination. AI (ML) has been exhibited to be suitable in assisting with making decisions and desires from the tremendous measure of data conveyed by the therapeutic administrations industry. it is have similarly watched ML techniques being used in late enhancements in different regions of the Web of Things (IoT). Various examinations give simply an investigate anticipating coronary illness with ML techniques. In this work, it is propose a novel procedure that objectives finding basic features by applying AI systems achieving improving the precision in the figure of cardiovascular infection. The figure model is given different mixes of features and a couple of known characterization techniques. it is produce an overhauled display level with an exactness level of 88.7% through the desire model for coronary illness with the cream irregular forest area with a straight model (HRFLM). [1]

W. Chang, et al.,[2019] The changes for an incredible duration rhythm and improvement in material levels that happened starting late extended the amount of people encountering hypertension on the planet. Thusly, as a cardiovascular intricacy of hypertension, the regularity of hypertensive coronary illness has extended each year, it has genuinely jeopardized the prosperity of human life, and the convincing desire for hypertensive coronary illness has become a general issue. This work uses the as of late proposed XGBSVM hybrid model to envision whether hypertensive patients will make hypertensive coronary illness inside three years. The last preliminary exhibits that through this model, hypertensive patients can pick up capability with their threat of hypertensive coronary illness inside 3 years and afterward experience zeroed in on preventive treatment, thusly decreasing the psychological, physiological and financial weight. This work certifies that the AI can be viably applied in the biomedical field, with strong genuine massiveness and exploration regard. [2]

T. S. Brisimi, et al.,[2018] Metropolitan living in present day enormous metropolitan networks has gigantic disagreeable ramifications for prosperity, growing the peril of a couple of unremitting sicknesses. it is revolve around the two driving lots of consistent sicknesses, coronary illness and diabetes, and make data driven techniques to envision hospitalizations as a result of these conditions. it is base these desires on the patients' therapeutic history, later and dynamically out of reach, as portrayed in their Electronic Wellbeing Records (EHRs). it is figure the desire issue as a twofold characterization issue and consider a collection of AI strategies, including kernelized and small Support Vector Machines (SVMs), sparse determined backslide, and irregular boondocks. To locate some sort of congruity among accuracy and interpretability of the desire, which is critical in a restorative setting, it is propose two novel methods: K - LRT, a likelihood extent test-based methodology, and a Joint Grouping and Arrangement (JCC) technique which perceives covered calm gatherings and changes classifiers to each bundle. it is make theoretical out-of-test guarantees for the last strategy. it is favor our figurings on colossal enlightening assortments from the Boston Clinical Center, the greatest prosperity net crisis facility system in New England.[3]

A. Mdhaffar, et al.,[2017] This work displays a novel prosperity examination approach for heart frustration conjecture. It relies upon the use of complex event taking care of (CEP) development, gotten together with genuine techniques. A CEP engine systems moving toward prosperity data by executing edge based examination rules. Instead of having to truly set up limits, our novel genuine count thusly registers and updates edges as demonstrated by recorded chronicled data. Test outcomes show the advantages of our strategy to the extent speed, precision, and recall.[4]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

G. S. Karanasiou et al.,[2016] Heart dissatisfaction (HF) is an endless infection depicted by bad quality of life, discontinuous hospitalization and high mortality. Adherence of patient to treatment proposed by the authorities has been shown a basic snag of the recently referenced authentic outcomes. In any case, the non-adherence rates are through and through high; a reality that includes the noteworthiness of predicting the adherence of the patient and enabling authorities to change in like way lenient watching and the chiefs. The purpose of this work is to foresee the adherence of patients with HF, using AI systems. Specifically, it plans to portray a patient as remedy devotee or not, yet furthermore as pupil or not to the extent medication, sustenance and actual activity (overall follower). Two characterization issues are watched out for: (I) if the patient is overall pupil or not and (ii) if the patient is solution supporter or not. Around 11 characterization figurings are used and gotten together with feature decision and resampling systems. The classifiers are evaluated on a dataset of 90 patients. The patients are portrayed as remedy and overall supporter, taking into account clinician assessment. The most important revelation precision is 82 and 91% for the first and the subsequent order issue, respectively.[5]

D. Tay, et al.,[2015] Myocardial dead tissue (MI) is one of the principle wellsprings of death in many made countries. In this manner, early ID of MI events is essential for incredible insurance medicines, potentially reducing avoidable mortality. One procedure for early sickness conjecture is the use of danger desire models made using AI systems. One critical portion of these models is to outfit clinicians with the versatility to change (e.g., the figure range) and use the danger desire model that they regarded commonly profitable for their patients. Thusly, in this work, it is make MI desire models and exploration the effect of test age and conjecture objectives on the introduction of MI risk desire models. The cardiovascular prosperity study dataset was used in this examination. Results exhibit that the figure model made using SVM estimation is good for achieving high affectability, expressness, and changed precision of 95.3%, 84.8%, and 90.1%, independently, over a period length of 6 years. Both model age and conjecture objectives were found not to altogether influence the show of MI risk desire models made using subjects developed 65 or more. This recommends chance desire models made using particular model age and estimate objectives is an achievable system. These models can be facilitated into a PC supported screening instrument which clinicians can use to decipher and envision the MI chance status of the individual patients in the wake of playing out the significant clinical examinations (e.g., scholarly limit, actual limit, electrocardiography, general changes to prosperity/lifestyle, and remedies) needed by the models. This could offer a strategies for clinicians to screen the patients at risk for having MI in the near future and underwrite early therapeutic intercession to reduce the risk.[6]

D. R. Patil et al.,[2014] In therapeutic field the finding of coronary illness is most inconvenient endeavor. It depends upon the wary assessment of different clinical and masochist data of the patient by remedial experts, which is trapped technique. As a result of movement in AI and information development, the researchers and therapeutic experts in colossal degree are enthusiastic about the progression of robotized structure for the desire for coronary illness that is significantly definite, convincing and obliging in early assurance. In this work it is accessible a desire structure for coronary illness using multilayer perceptron neural framework. The neural framework in this system recognizes 13 clinical features as data and it is readied using back-spread computation to foresee that there is closeness or nonappearance of coronary illness in the patient with most raised precision of 98% like various structures. The precision thusly got with this structure shows that it is ideal and profitable over various systems. [7]

R. Wijaya, et al.,[2013] In this work discussed the improvement of coronary illness figure using AI (for this circumstance the Fake Neural Organization or ANN). There are 13 factors that can choose coronary illness as demonstrated by Miss Chaitrali paper. Gauge of a person's coronary illness one year ahead is performed by examining the model pulse data. Data is taken by using gadget, for instance, splendid mirror, smart mouse, progressed cell phones and keen seat. Pulse data were assembled through the Web and accumulated in a worker. Learning in this system is performed for a period of one year to get enough data to make estimates. Farsighted of future coronary illness in one year can assemble a person's knowledge of coronary illness itself. The system is similarly expected to diminish the amount of patients and the amount of passings from heart disease.[8]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

Table 1: Summary of literature survey

Sr No	Author Name	Year of Publication	Proposed Work	Outcome
1	S. Mohan	IEEE, 2019	A novel method aims at finding significant features by applying ML	An accuracy level of 88.7%
2	W. Chang,	IEEE, 2019	The newly proposed XGBSVM hybrid model	An accuracy level of 82%.
3	T. S. Brisimi	IEEE, 2018	A JCC method which identifies hidden patient clusters	Validate algorithms on large data sets from the Boston Medical Center.
4	A. Mdhaffar	IEEE, 2017	A novel health analysis approach for heart failure prediction.	The merits of our approach in terms of speed, precision, and recall.
5	G. S. Karanasiou	IEEE, 2016	The adherence of patients with HF.	The highest detection accuracy is 82 and 91%
6	D. Tay,	IEEE, 2015	It is develop MI prediction models.	Accuracy of 92.3%, 84.8%, and 90.1%,
7	D. R. Patil	IEEE, 2014	A prediction system for heart disease using multilayer perceptron neural network	Accuracy of 88% comparative to other systems.
8	R. Wijaya	IEEE, 2013	The development of heart disease prediction machine learning in this case the ANN.	Expected to reduce the number of patients and the number of deaths from heart disease.

III. VARIOUS METHODS

The data mining has four main techniques namely

1. Classification,
2. Clustering,
3. Regression, and
4. Association rule.

Data mining techniques have the ability to rapidly mine vast amount of data. Data mining is mainly needed in many fields to extract useful information from a large amount of data. The fields like the medical field, business field, and educational field have a vast amount of data, thus these fields data can be mined through those techniques more useful information. Data mining techniques can be implemented through a machine learning algorithm. Each technique can be extended using certain machine learning models.

There are various types of classification algorithms and machine learning, such as decision trees, naive bayes, linear discriminate analysis, k-nearest neighbor, logistic regression, neural networks, and support vector machines.

1. **K-Nearest Neighbors Algorithm (KNN)**

   The K-Nearest Neighbors is an algorithm for supervised learning and is a classification algorithm that takes a bunch of labeled points and uses them to learn how to label other points. This algorithm classifies cases based on their similarity to other cases. In K-Nearest Neighbors, data points that are near each other are said to be neighbors. K-Nearest Neighbors is based on this paradigm. Thus, the distance between two cases is a measure of their dissimilarity.

2. **Decision Tree**

   Decision trees are built using recursive partitioning to classify the data, i.e., by splitting the training set into distinct nodes, where one node contains all of or most of one category of the data. A decision tree can be constructed by considering the attributes one by one.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

3. **Naïve Bayes**

Naïve Bayes classifier is a supervised algorithm which classifies the dataset on the basis of Bayes theorem. The Bayes theorem is a rule or the mathematical concept that is used to get the probability is called Bayes theorem. Bayes theorem requires some independent assumption and it requires independent variables which is the fundamental assumption of Bayes theorem.

4. Logistic Regression vs. Linear Regression

While Linear Regression is suited for estimating continuous values (e.g. estimating house price), it is not the best tool for predicting the class of an observed data point. In order to estimate the class of a data point, we need some sort of guidance on what would be the most probable class for that data point. For this, we use Logistic Regression. The difference between linear and multiple linear regression is that the linear regression contains only one independent variable while multiple regression contains more than one independent variables. The best fit line in linear regression is obtained through least square method. Linear regression finds a function that relates a continuous dependent variable, y, to some predictors.

5. Support Vector Machine

Support Vector Machine (SVM) is a supervised algorithm that can classify cases by dividing a data set into two or more classes using a separator.

6. Random Forest

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it's simplicity and the fact that it can be used for both classification and regression tasks. In this post, you are going to learn, how the random forest algorithm works and several other important things about it.

IV. CONCLUSION

From these studies, it is clear that there are various methods of data analysis of any application. Heart decease dataset is available from UCI Machine Learning Repository. It has been further preprocessed and cleaned out to prepare it for classification process. Decision trees, naive bayes, linear discriminate analysis, k-nearest neighbor, logistic regression, neural networks, and support vector machines are studied in this paper. The fact is that computers cannot replace humans and by comparing the computer-aided detection results with the pathologic findings, doctors can learn more about the best way to evaluate areas that computer aided detection highlights.

REFERENCES

1. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019.
2. W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou and W. Cao, "A New Hybrid XGBSVM Model: Application for Hypertensive Heart Disease," in *IEEE Access*, vol. 7, pp. 175248-175258, 2019.
3. T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," in *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690-707, April 2018.
4. A. Mdhaffar, I. Bouassida Rodriguez, K. Charfi, L. Abid and B. Freisleben, "CEP4HFP: Complex Event Processing for Heart Failure Prediction," in *IEEE Transactions on NanoBioscience*, vol. 16, no. 8, pp. 708-717, Dec. 2017.
5. G. S. Karanasiou *et al.*, "Predicting adherence of patients with HF through machine learning techniques," in *Healthcare Technology Letters*, vol. 3, no. 3, pp. 165-170, 9 2016.
6. D. Tay, C. L. Poh, E. Van Reeth and R. I. Kitney, "The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1178-1185, May 2015.
7. D. R. Patil and J. S. Sonawane, "Prediction of heart disease using multilayer perceptron neural network," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-6.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

8. R. Wijaya, A. S. Prihatmanto and Kuspriyanto, "Preliminary design of estimation heart disease by using machine learning ANN within one year," *2013 Joint International Conference on Rural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T)*, Bandung, 2013, pp. 1-4.
9. N. G. B. Amma, "Cardiovascular disease prediction system using genetic algorithm and neural network," *2012 International Conference on Computing, Communication and Applications*, Dindigul, Tamilnadu, 2012, pp. 1-5.
10. K. R. Taylor *et al.*, "AudioGene: Computer-based prediction of genetic factors involved in non-syndromic hearing impairment," *2011 9th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, Sharm El-Sheikh, 2011, pp. 75-79.
11. B. A. Thakkar, M. I. Hasan and M. A. Desai, "Health Care Decision Support System for Swine Flu Prediction Using Naïve Bayes Classifier," *2010 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, 2010, pp. 101-105.
12. H. Y. Wang, H. Zheng and F. Azuaje, "Evaluation of computational classification methods for discriminating human heart failure etiology based on gene expression data," *2006 Computers in Cardiology*, Valencia, 2006, pp. 277-280.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details