IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

Liver Cirrhosis Prediction Using Machine Learning

Akash M J[1], Abhishek I Hanchinamani[2] , Venkatesha H[3], Karthik S Badiger[4], Anu C.S[5]

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[1]

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[2]

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[3]

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[4]

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[5]

ABSTRACT: Liver cirrhosis is the most common type of chronic liver disease in the globe. The ability to forecast the onset of liver cirrhosis disease is critical for successful treatment and the prevention of catastrophic health implications. As a result, we are going to design a prediction model using machine learning techniques. The proposed model for the prediction of liver cirrhosis disease uses Ensemble learning models (Naive Bayes classifier, Classification and Regression Tree (CART), and Support Vector Machine (SVM) with 10-fold cross-validation). Accuracy, precision, recall, and F1 Score were used to evaluate the model's performance. Ensemble learning techniques may provide a more accurate prediction for liver cirrhosis disease. This approach can be used to help doctors make better clinical decisions.

KEYWORDS: Liver functions tests, Data preprocessing, Deep learning, Ensemble Model.

I. INTRODUCTION

Liver cirrhosis is an important type of liver damage. It usually occurs as a result of long term damage of liver caused by many forms of liver diseases and circumstances, such as hepatitis and chronic alcoholism or through genetics. Each time the liver is injured it tries to repair itself fibrous scar tissue can be deposited in place of the missing cells which forms the cirrhosis. As cirrhosis progresses, more and more scar tissue forms, hence making it difficult for the liver to function. Advanced cirrhosis is life threatening. The liver damage done by cirrhosis generally can't be undone. But if liver cirrhosis is diagnosed early and the cause is treated, further damage can be limited and, rarely, reversed. In addition to fibrosis, the complications of cirrhosis include portal hypertension, ascites, hepatorenal syndrome and hepatic encephalopathy.

A poor correlation exists between histologic findings of cirrhosis and the clinical picture. Some patients with cirrhosis are completely asymptomatic and have a reasonably normal life expectancy while some individuals have severe symptoms of end-stage liver disease and limited chance for survival. Common signs and symptoms may arise from decreased hepatic synthetic function (coagulopathy), decreased detoxification capabilities of the liver (hepatic encephalopathy) or portal hypertension (variceal bleeding) (Wolf & Katz, 2013). ICT has been globally credited for changing the course of history and adding value to human lives in various ways. Of all the technologies that add value and enhance human life, the introduction of telemedicine which perhaps go down in history as the most defining and has the potential to impact positively on humans, especially those living in the rural areas (Ezeorah, Ayatalumo & IbeEnwo, 2009).

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

1.1 Causes of Liver Cirrhosis:

- **Chronic Alcoholism**: Excessive and prolonged alcohol consumption is a common cause of liver cirrhosis. Alcohol-related liver disease occurs when the liver is unable to metabolize alcohol effectively.
- **Viral Hepatitis:** Chronic infections with hepatitis B or C viruses can lead to inflammation and damage to the liver over time, contributing to cirrhosis.
- **Non-Alcoholic Fatty Liver Disease (NAFLD):** Accumulation of fat in the liver, often associated with obesity and metabolic syndrome, can progress to non-alcoholic steatohepatitis (NASH) and eventually lead to cirrhosis.
- **Autoimmune Hepatitis:** The immune system mistakenly attacks healthy liver cells, causing inflammation and, in some cases, cirrhosis.
- **Genetic Disorders**: Inherited conditions such as hemochromatosis, Wilson's disease, and cystic fibrosis can lead to the accumulation of toxins in the liver, contributing to cirrhosis.
- **Biliary Atresia:** This is a rare condition where the bile ducts inside or outside the liver are absent or damaged, leading to bile accumulation and liver damage, especially in infants.
- **Primary Biliary Cirrhosis (PBC) and Primary Sclerosing Cholangitis (PSC):** These are autoimmune conditions affecting the bile ducts, leading to inflammation, scarring, and eventually cirrhosis.

II. LITERATURE REVIEW

C. Geetha et al. has proposed a work on "**Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms"** in **2021.** In this Study, methods used are Support Vector Machine, Decision Tree. While its Accuracy is 70%. This work focused on algorithms for classification of healthy people from liver datasets. Centre on their success variables, this research also aims to compare the classification algorithms and to provide prediction accuracy **[1]**.
Jianxia Wen et al. demonstrated a work on **"Research Progress and Treatment Status of Liver Cirrhosis with Hypoproteinemia"** in **2022.** In this paper, Support Vector Machine is used. While its Accuracy is 55%. This study comprehensively analyzed the common complications, pathogenic mechanisms, and treatment status of cirrhosis caused by hypoproteinemia and proposed research prospects for dealing with this increasingly serious problem

[2]. Manjula Devarakonda Venkata Sumalatha Lingamgunta et al. proposed a work on **"Health Care Automation"** in **2022.** The ML-based system for the early prediction of liver disease based on the Indian dataset has been successfully developed using the RF algorithms. The performance analysis of the technique is presented in terms of metrics for evaluation. The performance of predicting the positive cases from the dataset is approximately 95% which is evident from the recall metrics. Further, the precision reported to be approximately 74% emphasizing that the performance of predicting the positive class which are actually positive

[3]. Md. Fazle Rabbi, et al. demonstrated a work on **"Prediction of Liver Disorder Using Machine Learning Algorithm"** in **2020.** In this research, Logistic regression, Decision Tree, Random Forest. ML algorithms such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Extra Trees (ET) for classifying Indian Liver Patient Dataset (ILPD). Pearson Correlation Coefficient based feature selection (PCC-FS) is applied to eliminate irrelevant features from the dataset. Also, a boosting algorithm (AdaBoost) is utilized to enhance the predictive performance of those algorithms. The comparative analysis is evaluated in terms of accuracy, ROC, F-1 score, precision, and recall. After comparing experimental results, we have found that boosting on ET provides the highest accuracy of 92.19%

[4]. **Hartatik et al.** demonstrated a work on **"Prediction for Diagnosing Liver Disease in Patients using KNN and Naive Bayes Algorithms"** in **2020.** Based on the results of testing the Naive Bayes and KNN algorithms to solve predictive problems for patients with liver disease or not using the python application. Data were taken from the UCI Machine Learning Repository, namely the Indian Liver Patient Dataset (ILPD). The results show that of the two algorithms, the Naive Bayes algorithm provides a better value than the KNN by using six variables in the prediction model, which gives an increase in accuracy compared to the results of previous studies

[5].Sateesh Ambesange et al. presented a work on **"Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques"** in **2020.** In this work, ML models are built using various preprocessing techniques to balance the unbalanced data and predicted using RF algorithm. If Data set is imbalanced, only performing pre-processing like – Replacing missing value, outlier treatment and transforming the data set does not improve the results. Even hyper parameter tuning, Feature selection and PCA methods improve performance up to a certain value,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

as mentioned in the result table. To further fine tune a model, the balancing the data set is essential which is done using many oversampling and under sampling techniques in this work. It has observed that prediction accuracy from oversampled dataset is less compared to under sampled dataset, which indicates that over sampling dilutes the correlation and relationship of features with target label and increases variance in result. It also indicates, more data not always gives better results, it is quality of data, cleaned data essential for building efficient models. In future same techniques we can use in another set of data and check for the prediction accuracy

[6]. **Maria Alex Kuzhippallil et al.** proposed work on **"Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients"** in **2020.** In this work, liver disease prediction has been studied and analyzed. The data is cleaned by performing various techniques such as imputation of missing values with median, label encoding to convert categorical into numerical data for easy analysis, duplicate value elimination and outliers are eliminated using Isolation forest in order to improve the performance. Genetic algorithm combined with XGBoost is used to fetch the best attributes required for prediction of liver disease. Different classification algorithms are used to predict the presence or absence of liver disease. Performance metrics such as accuracy, precision, recall, f-measure and time complexity is effectively utilized to analyze the performance of various classification algorithms

[7]. **A.Sivasangari et al.** proposed a work on **"Diagnosis of Liver Disease using Machine Learning Models"** in **2020.** In this paper, the different machine learning algorithms is evaluated for the prediction of liver disease. Due to the subtle nature of its symptoms, liver disease is particularly difficult to diagnose. Liver disease prediction followed the step of preparing data in that data was collected from the public database preprocessing of data for -1 value replacement. Data division into the entire array of data split into training and research. Eventually, quantitative measurement metrics such as precision, accuracy and recall are measured over various machine learning models

[8]. **Sateesh Ambesange et al.** demonstrated a work on **"Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques**" in **2020**. In this work they developed the K-Nearest Neighbor model to diagnose and predict liver disease. The data is transformed and further dimensionality reduction is performed to reduce the features to improve the model performance. The performance of classification and prediction techniques are evaluated on different performance measures some of them are precision , accuracy, recall and score of F-1. Grid Search is used for tuning the model's hyper parameters like solver, max iterations, random-state etc. The model not only gives best accuracy, it also gives a perfect score in terms of AUC-ROC curve, precision, recall and other matrices of the model. The K NN model performs better with an accuracy of 91%. In future this model can be utilized for larger and real time datasets with more attributes, so that the model can perform even more accurately

2.1 Literature Review Summary

The literature review aimed to explore existing research on the early prediction of liver cirrhosis, focusing on machine learning techniques. The review encompassed studies published in last few years, investigating various algorithms and methodologies employed in predicting liver cirrhosis. While the literature presented promising advancements, it also acknowledged certain challenges. The main challenge is that the Existing Systems are unable to predict the liver cirrhosis in its early stages.

Gaps Identified:

There are some other challenges like low accuracy, unable to predict the stages, data quality and model generalization, the review points towards collaborative efforts for large, diverse datasets. The future direction emphasizes the need for early detection along with the stage in which the patient belongs.

Existing System :

The existing system for liver cirrhosis assessment often relies on traditional diagnostic methods, including clinical evaluation, liver function tests, imaging studies, and invasive procedures like liver biopsy. While these approaches are fundamental, they may have limitations in terms of early detection and widespread applicability. Machine learning has emerged as a complementary tool to enhance the existing system. By leveraging vast datasets and advanced algorithms, ML models can analyze diverse patient data to identify patterns and subtle indicators associated with liver cirrhosis. This offers the potential for earlier and more accurate predictions, enabling timely interventions and personalized healthcare strategies.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

III. PROBLEM STATEMENT

Liver cirrhosis has become a common disease around the world. The death rate due to the disease is becoming alarming. Early detection of the disease may reduce the complication of the disease misfortune on patients. The ease of use of inventive technologies such as the one anticipated in this research may help in alleviating the troubles of holdup in the uncovering and treatment of liver cirrhosis. The Machine learning tools are used to predict whether patient is positive or negative for the Disease. One more significant drive behind this is Predicting Stage also in which stage patient is there.

1. Proposed System:

The proposed system for liver cirrhosis prediction through machine learning is designed to revolutionize current diagnostic approaches. This system integrates a diverse array of patient data to develop a robust predictive model using advanced machine learning algorithms. The primary objective is to predict the liver cirrhosis in the early stages. Overall, this proposed system represents a significant step toward more accurate and user-friendly liver cirrhosis prediction system.

2. Objectives:

1. To design a model to analyse various patient data and predict the presence of liver cirrhosis.
2. To design a model to predict the stage of liver cirrhosis using Ensemble Classification algorithm for model creation.
3. To design front End application using flask for user usage.

3. Methodology

This section provides a summary of the datasets, the suggested method, the structural design of the system, and the algorithms utilized for categorizing liver disease.

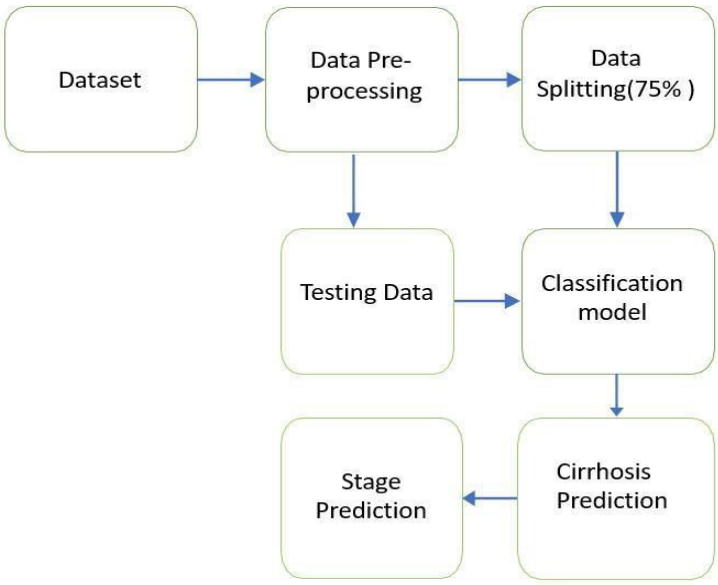

Figure 8.1: Design Frame work

3.1 Dataset Description

Liver disease categorization is performed using the dataset pertaining to Indian Liver Patients (ILPD) sourced from the UCI Machine Learning Repository.[2] It comprises 13 columns. Table 1 presents a summary of the feature characteristics for the patients

.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

3.2 Dataset Pre-Processing

The target feature in the dataset represents the categorical health condition of patients' livers. The control group, also known as the negative class, consists of blood donors. On the other hand, the positive classes include patients diagnosed with Cirrhosis.

3.3 Utilizing Improved Pre-processing for Liver Disease Prediction via Machine Learning Algorithms.

This study assesses how ensemble-driven machine learning techniques perform on the Dataset and conducts a comparative analysis of their outcomes. The Ensemble methodology involves a distinct strategy where we merge several machine learning models, whether similar or dissimilar, to execute prediction tasks, such as logistic regression (LR), KNN, support vector machines (SVM), and so forth [19]. The ensemble models employ foundational estimators or base learners. There exist numerous rationales for favoring ensemble models over conventional ones.

Our results indicate that the utilization of ensemble classification approaches leads to higher accuracy when compared to individual classifiers [20]. The amalgamation of these algorithms demonstrated superior performance in contrast to using a single algorithm. The discovery was made that selecting classifiers with independence and divergent perspectives leads to enhanced outcomes

IV. RESULTS

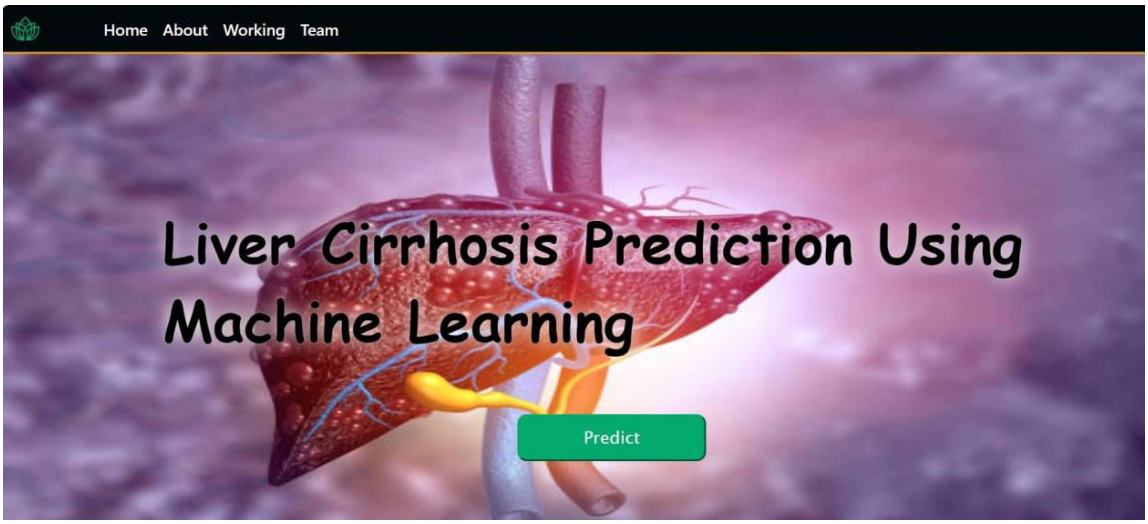

Fig. 9.1: User Interface Design of Home Page

Fig. 9.2: User Interface Design of Form for Data Entry

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

This is the form which accepts Patient information from various medical reports.

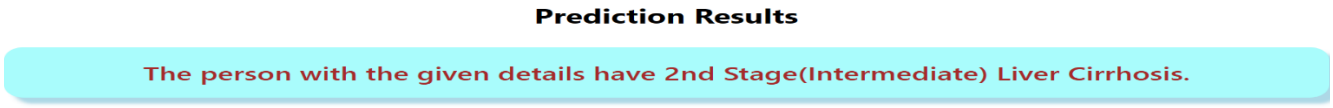

Fig. 9.3: Prediction Result Page

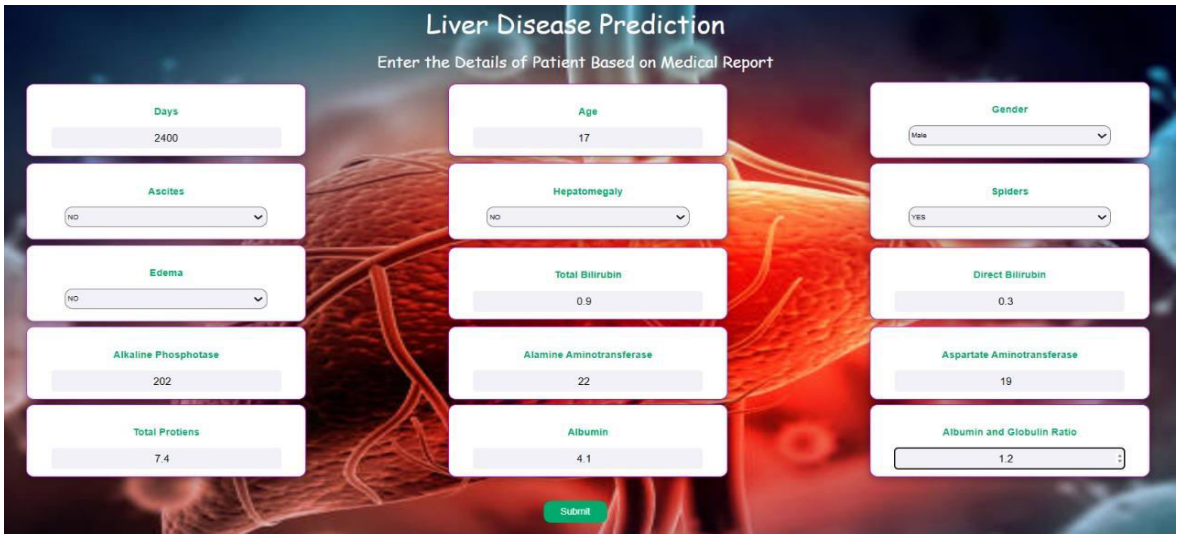

Fig. 9.4: User Interface Design of Form for Data Entry

Fig. 9.5: Prediction Result Page

Fig. 9.6: User Interface Design of Form for Data Entry

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204182 |

Fig. 9.7: Prediction Result Page

REFERENCES

1. A. Al-Aiad, S. Abualrub, Y. Alnsour, and M. Alsharo and was titled "DataMining Algorithms Predicting Different Types of Cancer: Integrative Literature Review." It debuted during the AMCIS 2020 TREOs. You may find the document at https://aisel.aisnet.org/treos_amcis2020/59.
2. R. D. Canlas Jr. finished an unpublished master's thesis titled "DATA MINING IN HEALTHCARE: CURRENT APPLICATIONS AND ISSUES" in August 2009. The ten-page thesis focuses on the use of data mining in healthcare.
3. Ibrahim and A. Abdulazeez wrote a paper in the Journal of Applied Science and Technology Trends titled "The Role of Machine Learning Algorithms in Disease Diagnosis." The essay appears on pages 10through 19 of volume 2, issue 1. It was published in 2021 and has the following DOI: 10.38094/jastt20179.
4. "Hepatitis C Virus Vaccine: Challenges and Prospects," co-authored by J. D. Duncan, R. A. Urbanowicz, A. W. Tarr, and J. K. Ball, was published in the journal "Vaccines." The paper goes from page 1 through page 23 of volume 8, issue 1. It was published in 2020 and has the DOI: 10.3390/vaccines8010090.
5. L. Syafa'ah, Z. Zulfatman, I. Pakaya, and M. Lestandy did study titled "Comparison of Machine Learning Classification Methods in Hepatitis C Virus," which was published in 2021 in the Journal ofOnline Information, volume 6, issue 1, page 73. The corresponding DOI is 10.15575/join.v6i1.719.
6. Günaydin, M. Günay, and engel co-authored a paper titled "Comparison of Lung Cancer Detection Algorithms," which was presented at the 2019 Scientific Meeting on Electrical, Biomedical Engineering, and Computer Science. The DOI for the publication is 10.1109/EBBT.2019.8741826, andit is tied to the EBBT 2019 event.
7. G. S. Rao, G. V. Kumari, and B. P. Rao contributed to "Network for Biomedical Applications," which was published by Springer Singapore in Volume 2, Issue 1 in January 2019. The DOI for this article is10.1007/978-981-13-1595-4.
8. "Enhanced Prognosis of Hepatocellular Carcinoma Fatality Using Ensemble Learning Approach," by M. Sharma and N. Kumar, was published in the Journal of Ambient Intelligence and Humanised Computing. The paper was published in Volume 13, Issue 12, pages 5763-5777 in 2022. The associated DOI is 10.1007/s12652-021-03256-z.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

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