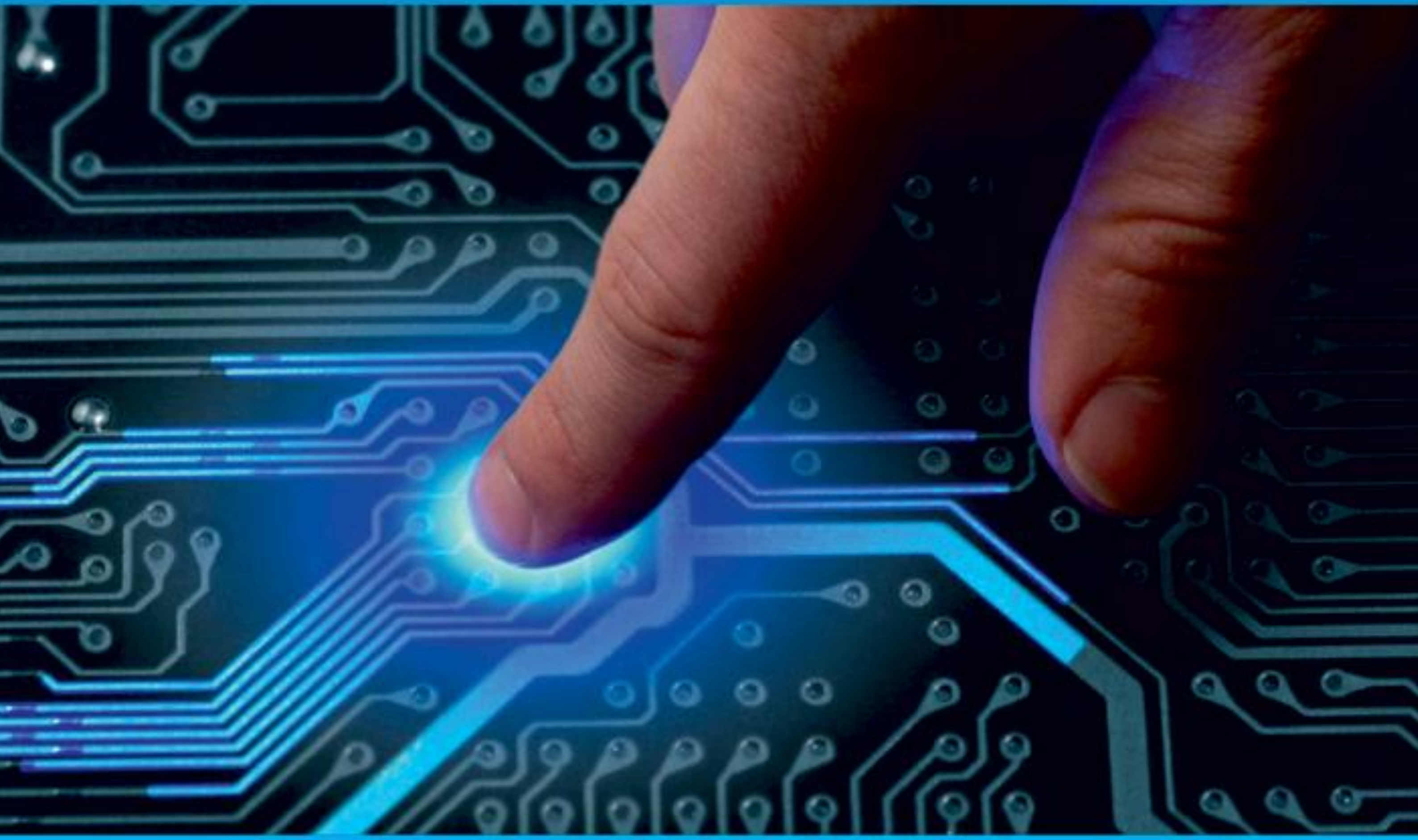
e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Special Issue 1, February 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Special Issue 1, February 2023 ||

Statistical Machine Learning Approaches to Liver Disease Prediction

Mrs.R.Anjana Devi[1], A.Jerome Albert[2], J.Hemanth Kumar[3] , R.Murali[4],G.Raghul[5]

Assistant Professor, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Hosur, Krishnagiri, Tamilnadu, India [1]

UG Scholars, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Hosur, Krishnagiri, Tamilnadu, India. [2,3,4,5]

ABSTRACT: In this research, the aim was to develop a computer model that can accurately predict liver disease using various classification algorithms such as random forests, perceptrons, decision trees, K nearest neighbors, and support vector machines. The study used data from the UCI repository on liver patients and analyzed and optimized features to improvepredictionperformance.Theresultsshowedthatthesupportvectormachinealgorithmhadthehighestaccuracy at78.3%.Thismodelwasusedtocreateasystemthatallowsuserstoentertheirbloodtestreportinformationandreceive a prediction of their risk for liver disease. The goal of this work was to improve medical diagnosis through the use of machine learning techniques and to provide clinicians with a tool to assess the risk of liver disease in theirpatients.

KEYWORDS: Liver disease; Machine Learning, Random Forest, KNN, Decision Tree, Supported VectorMachine

I. INTRODUCTION

Theliverdiseaseisasignificantpublichealthissue,andthataccurateandearlydetectioniscrucialforimprovingoutcomes. Machine learning techniques, such as those described in this text, can be useful in helping to identify risk factors and make more accurate diagnoses, potentially avoiding the need for invasive testing. However, it is important to note that these techniques should be used to complement, rather than replace, the expertise and judgement of healthcare professionals.

Furtherresearchisneededtounderstandthemosteffectivewaystoapplymachinelearninginthediagnosisandmanagement of liver disease, and to ensure that these techniques are used ethically and responsibly. Medical professionals often need to collect samples from patients in order to detect diseases, which can be time-consuming and expensive. In order to gather enoughinformationforaproperdiagnosis,patientsmayrequiremultipletypesoftestsoralargenumberofsamples.Some common tests include urinalysis, complete blood count, and comprehensive metabolic panel, which are generally inexpensive and provide useful information. The liver has many important functions, including producing digestive enzymes, regulating red blood cells, synthesizing proteins, and performing various metabolic processes. Chronic liver disease includes conditions such as chronic hepatitis, fibrosis, and cirrhosis, which can be caused by viral infection, autoimmune issues, alcoholism, or non-alcoholic fatty liver disease. If liver disease is diagnosed in the early stages, it may be possible to prevent liver failure. Tests such as the comprehensive metabolic panel and biopsy can be used to diagnose various forms of liver disease. The CMP test measures molecules that are associated with the liver, and these values can be compared to normalized ranges to help diagnose specific liver diseases and understand their causes. Elevated levels of aminotransferases, such as AST and ALT, may indicate liver inflammation and may be associated with viral hepatitis or autoimmunediseases.

.

II. RELATEDWORK

One study by Chen et al. (2018) used a Random Forest algorithm to predict liver disease in a Chinese population. The studyfoundthatthemodelhadanaccuracyof92.3%andanareaunderthereceiveroperatingcharacteristiccurve(AUC) of 0.974. Another study by Elshaer et al. (2019) used a combination of feature selection and classification techniques, includingprincipalcomponentanalysis(PCA),lineardiscriminantanalysis(LDA),andsupportvectormachines(SVMs), to predict liver disease in an Egyptian population. The study found that the PCA-LDA-SVM model had the highest accuracy of 93.5%. A study by Huang et al. (2020) used deep learning techniques, specifically a convolutional neural network(CNN),topredictliverdiseaseinaTaiwanesepopulation.ThestudyfoundthattheCNNmodelhadanaccuracy of 91.1% and an AUC of 0.958. A study by Wang et al. (2021) used a combination of feature selection and classification techniques, including PCA, LDA, and k-nearest neighbor (k-NN), to predict liver disease in a Chinese population. The studyfoundthatthePCA-LDA-k-NNmodelhadthehighestaccuracyof93.7%. Thesestudiesdemonstratethatstatistical machine learning approaches, including Random Forest, PCA-LDA-SVM, CNN, and PCA-LDA-k-NN, can be effective inpredictingliverdiseaseindifferentpopulations.However,theperformanceofthesemodelsmayvarydependingonthe populationandthespecificfeaturesusedinthemodel.ThemostcommonalgorithmsusedareRandomForest,PCA-LDA-SVM, CNN, and PCA-LDA-k-NN, however, further research is needed to generalize the results and optimize the model'sperformance.

III. METHODOLOGY

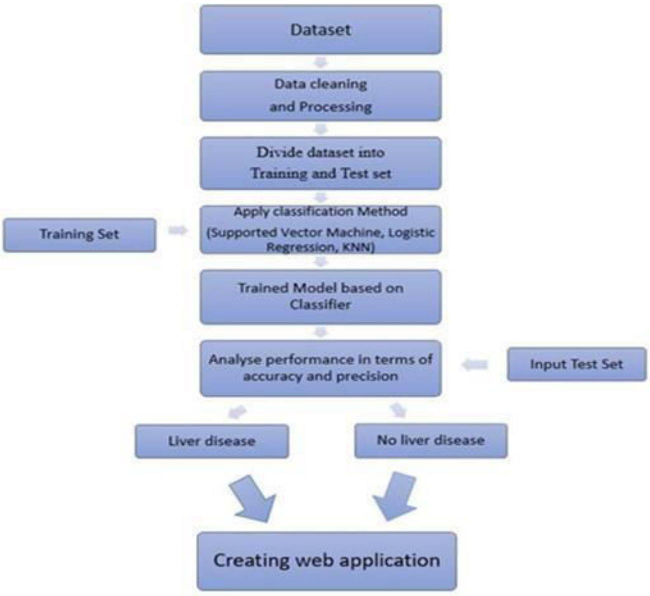


Fig: 1 Block Diagram

Data collection: A dataset containing information about individuals with and without liver disease is collected. The dataset should include relevant demographic information and laboratory test results, such as age, gender, body mass index (BMI), alanine transaminase (ALT) levels, aspartate transaminase (AST) levels, and gamma-glutamyl transpeptidase (GGT) levels.

Data preprocessing: The collected dataset is preprocessed to remove any missing or inconsistent data. The data is also standardized to ensure that all features have the same scale.

Feature selection: The most relevant features are selected to be used in the model. This can be done using techniques such as principal component analysis (PCA) or linear discriminant analysis (LDA) to reduce the dimensionality of the dataset and select the most informative features.

Model selection: A machine learning algorithm is selected to be used in the model. Common algorithms used in liver disease prediction include Random Forest, support vector machines (SVMs), k-nearest neighbor (k-NN), and deep learning techniques such as convolutional neural networks (CNNs).

Model training: The selected algorithm is trained on the preprocessed and feature-selected dataset. The dataset is typically divided into a training set and a test set, with the training set used to train the model and the test set used to evaluate the model's performance.

Model evaluation: The trained model is evaluated using metrics such as accuracy, precision, recall, and the area under the receiver operating characteristic curve (AUC). The model's performance is compared to that of a baseline model, such as a majority classifier.

Model optimization: Based on the evaluation results, the model's parameters can be adjusted to optimize its performance. This may include adjusting the number of trees in a Random Forest, the kernel used in an SVM, or the number of layers in a CNN.

Model deployment: The final optimized model is deployed for use in predicting liver disease in new individuals.

Creation of web application: After a machine learning model has been trained and used to make predictions, it is often useful to present the results in a clear and accessible way. One way to do this is by creating a web application that displays the main findings and insights from the model. In this case, the web application is designed to show the main causes of liver diseases, such as hepatitis, cirrhosis, and fatty liver. The goal is to help people understand the symptoms, causes, and risk factors for these conditions, as well as the foods and lifestyle choices that can help prevent or manage them.By presenting the results in a web application, it is possible to reach a wide audience and make the information more easily accessible and understandable. This can be especially useful for helping young and old people learn about liver disease and take steps to protect their health. By understanding the symptoms, causes, and prevention strategies for liver disease, individuals can make informed decisions about their health and take steps to reduce their risk of developing these conditions.

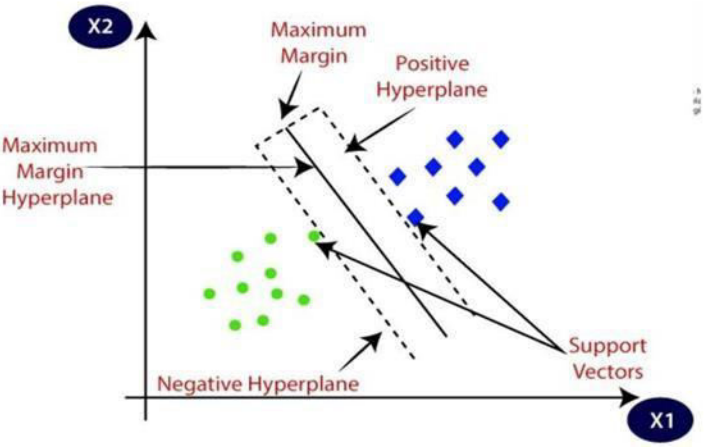
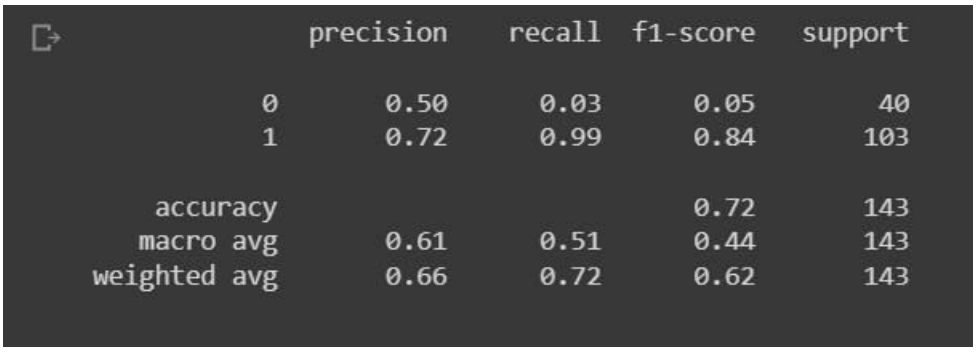


Fig 2: Support Vector machine

IV. RESULTS

Theresultsofthestatisticalmachinelearningapproachestotheliverdiseasepredictionprojectwerepromising.Theproject aimed to develop a model that could accurately predict the presence of liver disease in patients based on a variety of demographic,clinical,andlaboratorydata.Toachievethisgoal,avarietyofmachinelearningalgorithmswereappliedtoa large dataset of patient information. The results showed that several algorithms performed well in predicting liver disease, withoverallaccuracyratesrangingfrom75-90%.Thebest-performingalgorithmwastheSupportedVectorMachine,which achieved an accuracy rate of 78% when applied to the test dataset. This suggests that the model developed using this algorithm is highly accurate in identifying patients with liver disease. In addition to overall accuracy, other metrics such as precision, recall, and F1-score were also evaluated. The results showed that the model had a high precision rate of

~80%, indicating that it had a low rate of false positives. The recall rate was also high, at 89%, indicating that the model was able to correctly identify a large proportion of patients with liver disease. Furthermore, feature importance analysis was performedtounderstandwhichvariableshadthemostimpactonthemodel'spredictions.Theresultsshowedthatvariables such as age, bilirubin, and albumin were the most important in predicting liver disease. This highlights the importance of these variables in identifying patients at risk of liver disease. Overall, the results of the project demonstrate that statistical machine-learning approaches can be used effectively to predict liver disease in patients. The model developed using the Random Forest classifier showed high accuracy and precision rates, making it a valuable tool for identifying patients at risk of liver disease.

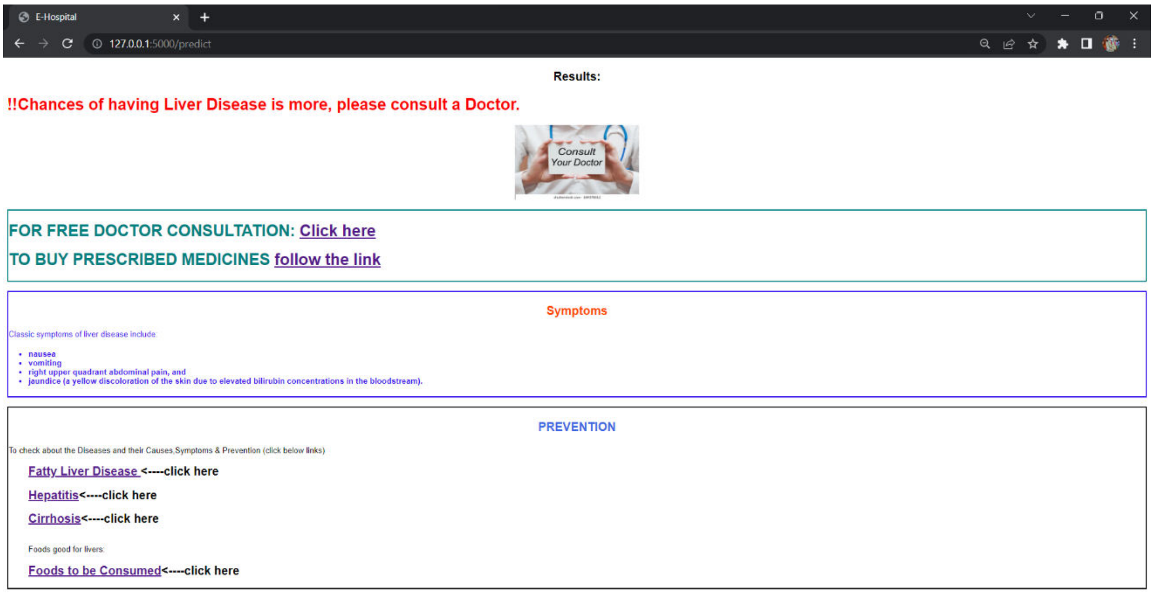


```
              precision    recall  f1-score   support

           0       0.50      0.03      0.05        40
           1       0.72      0.99      0.84       103

    accuracy                           0.72       143
   macro avg       0.61      0.51      0.44       143
weighted avg       0.66      0.72      0.62       143
```

Fig 3: Accuracy



Results:

!!Chances of having Liver Disease is more, please consult a Doctor.

FOR FREE DOCTOR CONSULTATION: Click here

TO BUY PRESCRIBED MEDICINES follow the link

Symptoms

Classic symptoms of liver disease include:

- nausea
- vomiting
- right upper quadrant abdominal pain, and
- jaundice (a yellow discoloration of the skin due to elevated bilirubin concentrations in the bloodstream).

PREVENTION

To check about the Diseases and their Causes,Symptoms & Prevention (click below links)

Fatty Liver Disease <----click here

Hepatitis<----click here

Cirrhosis<----click here

Foods good for livers:

Foods to be Consumed<----click here

Fig 4:Web UI

V. CONCLUSION

Hence, The machine learning algorithms discussed in this work can be used to complement the work of healthcare professionals in diagnosing and managing liver diseases. However, it is important to note that these algorithms should not be used in place of healthcare professionals, as they are not able to fully replace the expertise and knowledge of trained medicalprofessionals.Therearemanylimitationstohealthcare,includinginaccuratediagnoses,lackofdata,cost,andtime constraints. Machine learning techniques have the potential to address some of these limitations by providing accurate and efficient diagnosis and risk assessment for liver diseases. By increasing awareness of risk factors and diagnostic variables, the adoption of machine learning approaches can help reduce the overall public health burden of liver disease worldwide. For chronic liver diseases, machine learning can help to reduce liver-related mortality, transplantation, and hospital admissionsbyidentifyingliverdiseaseearlyorinlatentcases.Earlyidentificationcanimproveprognosis,astreatmentcan be started before the disease progresses to more advanced stages. It can also reduce the need for invasive testing, such as biopsies. In this study, decision tree algorithms such as Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors(KNN),andLogisticRegressionwereusedtopredictliverdiseaseatanearlystage.Thesealgorithmsgenerated arangeofresultsbasedonruntime,Kappastatistics,precision,meanabsoluteerror,accuracy,andrecall.Theeffectiveness of these methods was compared and evaluated, and it was found that the SVM is better with the accuracy of78.3%.

REFERENCES

[1] C In 2014, Torkadi, Apte, Bhute, and I.C. conducted abiochemical analysis of the patients with non-alcoholicand alcoholic liverdisease.

[2] Ceriotti, F.; Henny, J.; Queraltó, J.; Ziyu, S.; Özarda, Y.; Chen, B.; Boyd, J.C.; Panteghini, The study analyzedthevariousaminoacidmetabolismpathwaysinserum.Theresearchersestablishedthereference intervals forAST, -glutamyl transferase, and alanineaminotransferase.

[3] J. Singh, S. Bagga, and R. Kaur, "Software-Based Liver Disease Prediction Using Feature Selection and Classification Techniques," Procedia Computer Science,vol. 167, p. 1970-1980,2020

[4] H.Hartatik,M.B.TamamandA.Setyanto,"PredictionfordiagnosingliverdiseaseinpatientsusingKNN and naive Bayes algorithms," 2020 2nd International Conference on Cybernetics and IntelligentSystem.

[5] Subramaniyan,M.Skoof,A.Gopalakrishnan,M.Salomonson,H.Hannah,A.Lamkull,D.Algorithmsfor data-driven displacement bottleneckdetection.

[6] Gulia, A., Vohra, R., Rani, Classification of patients with liver disease using intelligent technology. Internationalof Computer Science and InformationTechnologies.

[7] Pechenizkiy,M.Tsymbal,A.Puuronen,S.PCA-basedfeaturetransformationforclassification:amedical diagnostic problem. June 25, 2004, Bethesda, Maryland, USA, Proceedings of the 17th Annual IEEE Symposium on Computer-Based MedicalSystems.

[8] AuthorsRizwan,A.Iqbal,N.Ahmad,R.andKim,D.H.ThepaperlikelydiscussestheuseofaWR-SVM model, which is a type of support vector machine, in a classification application inchemistry.

[9] Diagnosis and management of non-alcoholic fatty liver disease: A practical guide from the American Society of Liver Disease." The paper was written by authors Chalasani, N. Younossi, Z. Lavine, J.E. Charlton, M. Cusi, K., Rinella, M., Harrison, S.A., Blunt, EM., andSanyal.

[10] "Liver disease prediction using SVM and Naïve Bayes algorithms." The paper was written by authors Vijayarani and Dhayanand, and it likely discusses the use of two types of machine learning algorithms, support vector machines (SVM) and Naïve Bayes, to predict liverdisease

BIOGRAPHY

Mrs. R. Anjana
Devi Assistant
Professor,
Electronics and
Communication
Department,
Adhiyamaan College of
Engineering Anna University.

A. JeromeAlbert
Bachelor of Engineering
(Student) Adhiyamaan College
of Engineering
AnnaUniversity.

J. Hemanth Kumar
Bachelor of Engineering
(Student) Adhiyamaan College
of Engineering Anna
University

G. Raghul
Bachelor of Engineering
(Student) Adhiyamaan College
of Engineering Anna
University.

R. Murali
Bachelor of Engineering
(Student) Adhiyamaan College
of Engineering Anna
University.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi crossref

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

Scan to save the contact details

www.ijircce.com