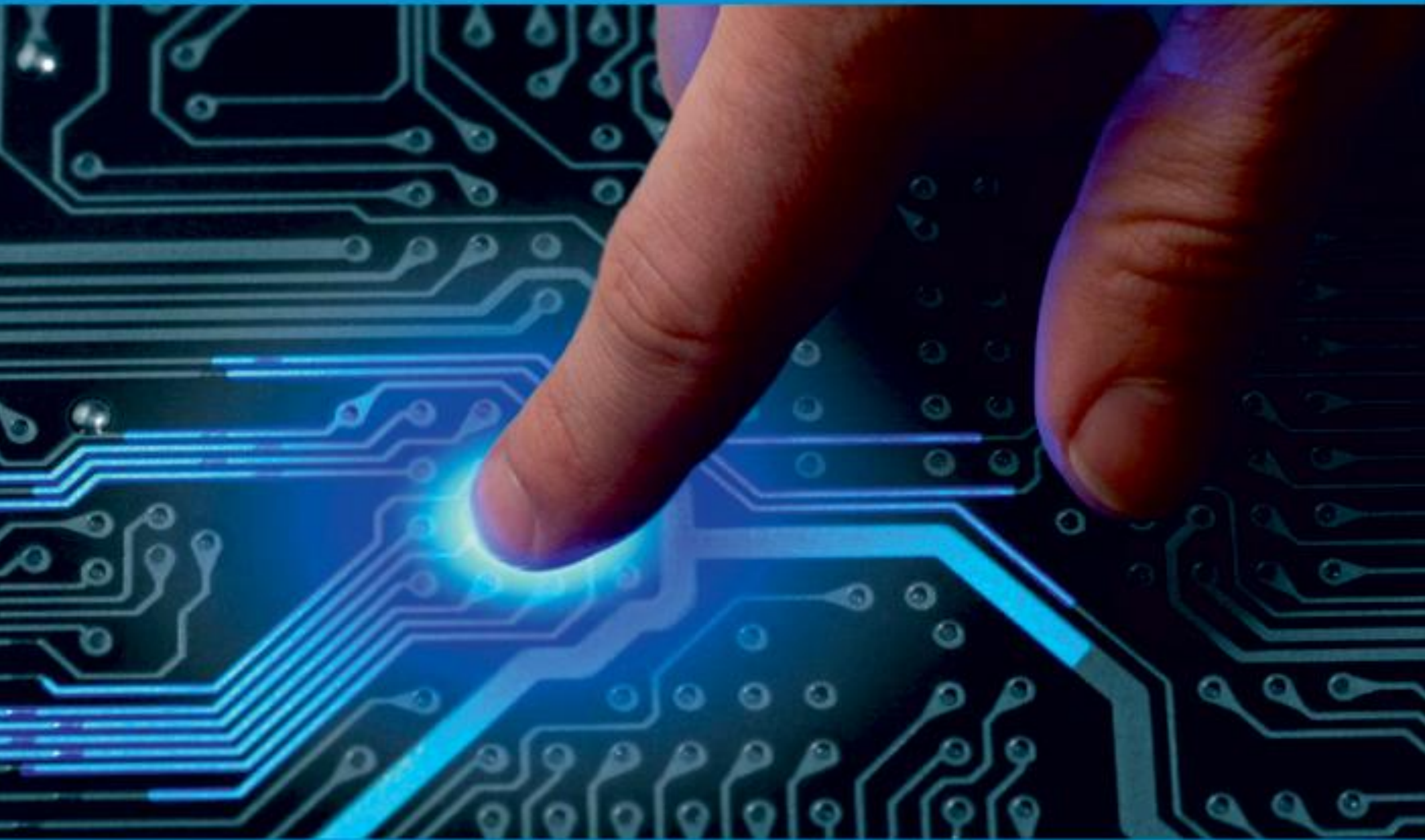
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Special Issue 1, March 2024

1st International Conference on Machine Learning, Optimization and Data Science

Organized by

Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, India

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 & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Predicting Cardiac Arrest Using Machine Learning with Zero Coding

Zeba Vishwakarma[1], Abhishek Singh[2]

Assistant Professor, Dept. of CSE, BGIEM, Jabalpur, India[1]

Associate Professor, Dept. of CSE, BGIEM, Jabalpur, India[2]

ABSTRACT: Heart disease cases are rising rapidly day by day, thus it's very difficult to predict any likely illness in advance. But to predict a heart ailment in advance is not an easy task. The consequences of several medical and pathological tests have to be combined in a complicated manner in order to diagnose cardiac disease. Doctors, researchers, and academicians are extremely interested in how Artificial Intelligence and Machine Learning can be used to anticipate heart illnesses because of the ominous complexity of the problem and the outstanding improvements in machine learning technology. ML has previously generated forecasts that were fast, precise, and efficient. However, the main obstacle to using machine learning in this field appears to be a shortage of coding skills. And here comes the solution: Alteryx Predictive Modeling. The cardiac disease patient dataset will be carefully and interactively analyzed as part of this research project, It also utilizes Alteryx for automated data processing. The dataset contains the parameters which are diagnosed as major medical conditions contributing to cardiac arrest. Then, the machine model is trained and predictions are made with XGBoost Classifier. We've utilized Predictive Modeling with Alteryx, which works kind of like a semi-automated machine learning process because it allows choosing certain parameters manually and handles many machine learning tasks automatically, it really speeds up creating models. Plus, you don't need to write any code at all.

KEYWORDS: Heart disease, Alteryx, AutoML, Predictive Modeling, XGBoost Classifier, Zero Coding.

I. INTRODUCTION

1.1 Heart disease: A threat like never before.

Cardiovascular disease (CVD), commonly referred to as heart disease, encompasses a wide range of conditions that affects the heart [1]. It was listed as the underlying cause of death, accounting for 874,613 deaths in the United States in 2019. On average in 2019, someone died of a stroke every 3 minutes and 30 seconds in the United States [2]. It's interesting to note that most of the time, unhealthy lifestyles, cigarette use, and alcohol use lead to heart failure rather than diseases of the heart itself. AMI, stroke, heart failure, and cardiac arrhythmias are all conditions brought on by these causes raising blood pressure, glucose levels, lipids, and other variables. In this research work, the dataset we are using consists of 14 such conditions (variables) which are more closely related to cardiac arrest.

1.2 Machine Learning: A boon to the medical field.

In the field of cardiovascular medicine, machine learning (ML), a subset of artificial intelligence (AI), is being used more and more. In essence, it describes how computers interpret data and categorize or decide on a job, whether or not human oversight is involved. Models that take in input data (such as photos or text) and predict results using a combination of mathematical optimization and statistical analysis are the foundation of machine learning (ML) theory (e.g., favorable, unfavorable, or neutral). Clinicians may find this information useful in figuring out the best algorithms to use for their dataset and in data interpretation.

1.3 Machine learning using Alteryx

With its AI Intelligent Suite, the analytics-focused data analytics platform Alteryx offers machine learning technologies. Regression, classification, and clustering approaches are used to develop prediction models for diverse machine-learning issues. Alteryx is the ideal option for anyone, no matter how non-technical a person is, to experience the field of machine learning algorithm development without writing a single line of code because of its versatility and a simple drag-and-drop interface. Every step of the machine learning process, including data preparation, feature selection, dataset splitting, model training, and deployment, can be completed using the tools at hand. Additionally, it includes a library of machine-learning algorithms that may be utilized for testing.

II. RELATED WORK

Bardhwaj et al., (2017), Shailaja et al., (2018), Sun et al., (2019), and Lee & Yoon, (2017) studied a broad overview of machine learning techniques used in healthcare for various diseases. They provided insights into the potential value of medical big data that can be used for clinical decision support, diagnostics, treatment decisions, fraud detection, and prevention. They briefly summarized the nine-step data mining process along with focusing on why efficient decision support was required by the healthcare system. The results from their experiment showed that machine learning models can be used for the early diagnosis of diseases. Their research is applicable to this project to an extent; however, their research is less focused on the diagnosis of heart diseases. Therefore, we move forward to review the literature that aligns with our project objective which is how machine learning algorithms can be used in the diagnosis of heart disease. A comprehensive review by Tripoliti et al., (2017) focused on machine learning methodologies evaluating heart failure. They researched severity estimation of heart failure and the prediction of re-hospitalization, mortality, and destabilizations. They performed an extensive study on related works of heart failure. A study by J. & S., (2019) used two supervised classifiers called Naïve Bayes Classifier and Decision Tree Classifiers to predict heart diseases on a dataset. 8 Their Decision Tree model predicted the heart disease patients with an accuracy of 91 percent and the Naïve Bayes Classifier had an accuracy of 87 percent. A study by Kamal kant et al.(2014) proposed a model using the Naïve Bayes algorithm to predict heart diseases. The naïve Bayes algorithm is used to assign no dependency between the features. Their study concluded that the Naïve Bayes algorithm is the most effective for heart disease prediction after that Neural Networks and Decision Trees.

III. PROPOSED ALGORITHM

XGBoost, an acronym for eXtreme Gradient Boosting, is a potent algorithm well-known for its efficiency in problems involving regression and classification. This is a suggested algorithm for heart disease prediction that makes use of XGBoost:

Set the XGBoost model's hyperparameters first. These include the learning rate, the maximum tree depth, the number of boosting rounds, and the regularization parameters. The model's performance can be greatly affected by these parameters, thus it is important to properly select them through testing or hyperparameter tweaking methods like random or grid search.

The dataset of patients with heart disease should be preprocessed in order to manage missing values, encode categorical variables, and scale or normalize numerical features as needed. Ascertain that the dataset is divided into training and testing sets in order to precisely evaluate the model's performance. Training: Match the training dataset to the XGBoost model. Using a gradient boosting architecture, XGBoost steadily improves the overall predictive performance by correcting flaws in prior model iterations, or trees. The model gains the capacity to forecast the likelihood of heart disease from the input features during training.

Prediction: Apply the learned XGBoost model to the testing dataset to generate predictions. Every event is given a probability score by the model that indicates the chance of heart disease. The accuracy and dependability of the model can then be evaluated by analyzing these predictions.

Model Evaluation: Utilize relevant metrics, such as accuracy, precision, recall, and F1-score, to assess the efficacy of the XGBoost model.

Optimization: If required, fine-tune the XGBoost model by modifying the hyperparameters or investigating more complex strategies like feature engineering or ensemble methods. To get the best potential results on the job of predicting cardiac disease, iteratively improve the model.

Deployment: Take the model to production for practical use after you're happy with its performance. Track its performance over time and think about retraining it sometimes to adjust to shifting requirements or data trends.

IV. PSEUDO CODE

1.Open Alteryx and import the required datasets and libraries.

2. Prepare the dataset:
    a. Deal with null values.
    b. For categorical variables, encode them.
    c. Scale or normalize numerical characteristics.
3. Create training and testing sets from the dataset.
4. Select pertinent characteristics to forecast cardiac disease.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |
| "1st International Conference on Machine Learning, Optimization and Data Science"|
Organized by
| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|
|| Volume 12, Special Issue 1, March 2024 ||

5. Set up predictive modeling with Alteryx.
6. Set up XGBoost Classifier as the model in Alteryx Predictive Modeling.
7. Use the training dataset to train the model:
    a. The model is automatically fitted to the training data by Alteryx Predictive Modeling.
8. Predict using the testing dataset and the trained model:
    a. Alteryx Predictive Modeling uses XGBoost to automatically produce predictions.
9. Assess the model's functionality:
    a. Examine and contrast expected and actual values.
    b. The F1-score, accuracy, precision, and recall are automatically determined by Alteryx Predictive Modeling.
10. Show performance metrics of the model.

V. SIMULATION RESULTS

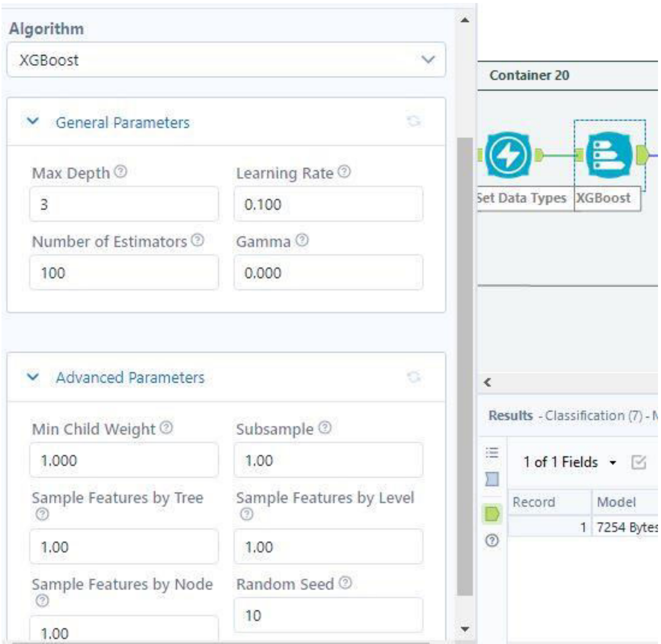


Fig.1 Selecting the XGBoost using Classifier tool in Alteryx

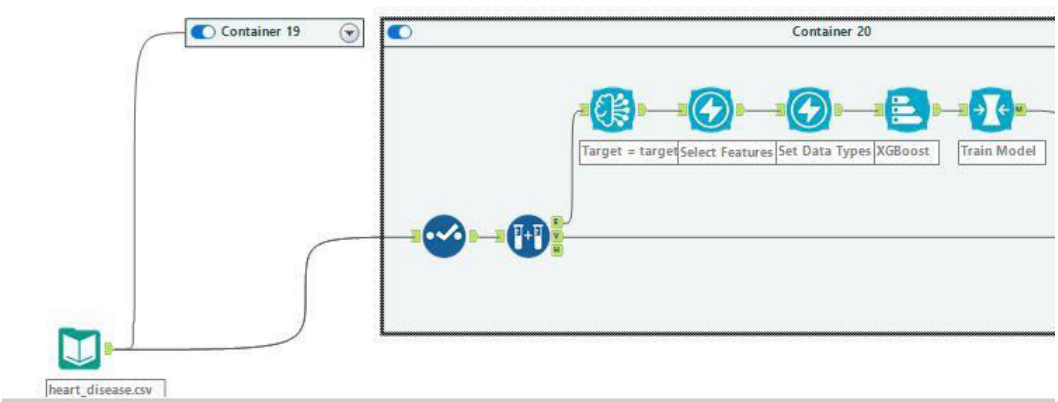


Fig.2 Train the model using the fit tool in Alteryx

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

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

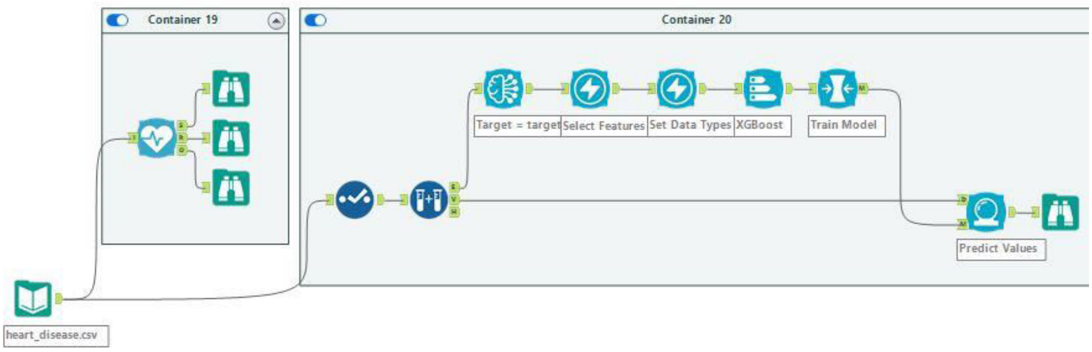


Fig.3 Predicting the Actual vs. Predicted values using the Predict Tool in Alteryx

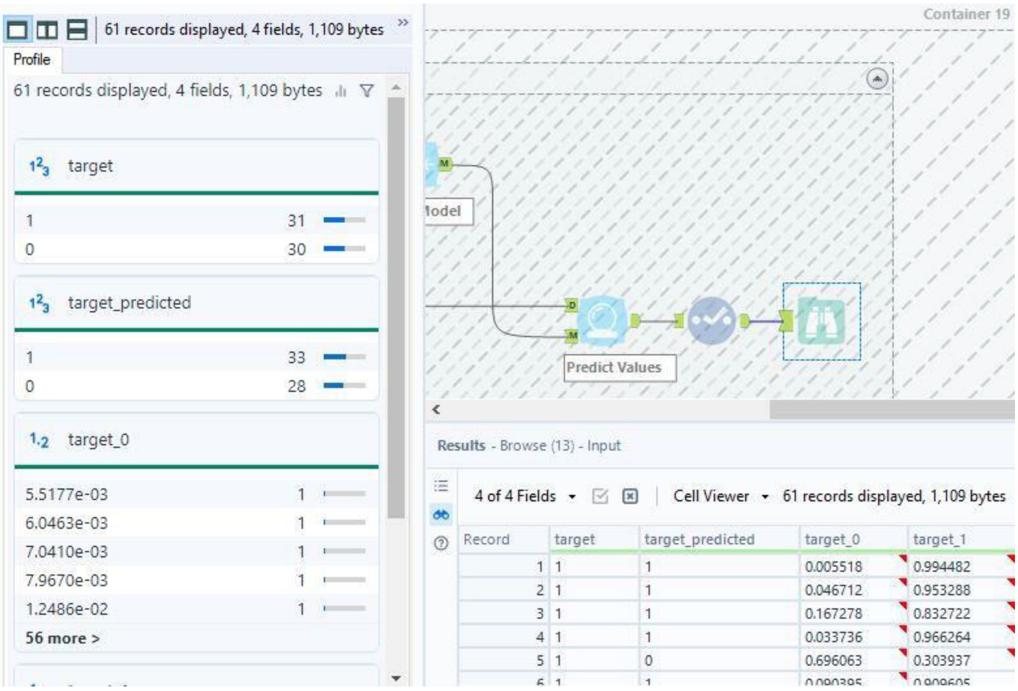


Fig.4 Comparing the actual vs. predicted output using Browse tool in Alteryx

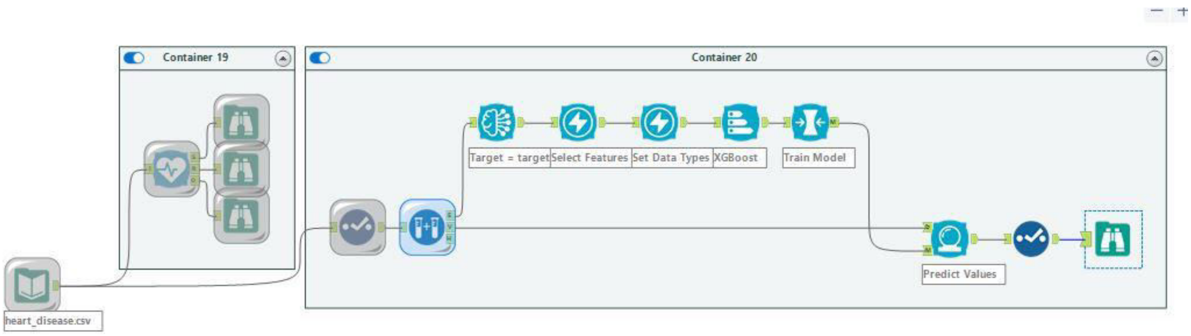


Fig.5 Complete machine model to predict the heart disease using Alteryx

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

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

VI. CONCLUSION AND FUTURE WORK

The machine learning model has been created using Alteryx Predictive Modeling using zero coding successfully. The model is giving an accuracy of approx. 93% in predicting heart disease based on certain parameters. The model used is XGBoost Classifier.

Future research and development for the aforementioned project may focus on a number of areas, including:

Feature engineering is the process of adding additional pertinent parameters or derived features to the feature set in order to increase the model's capability for prediction. To make sure the model includes all pertinent parts of the data, feature engineering is essential.

Data Collection and Quality: Make sure the model is trained using a representative and complete dataset. This could entail gathering more information from various sources in order to record a wider range of patient characteristics, medical background, and diagnostic factors.

Model tuning: To possibly enhance model performance, adjust the XGBoost Classifier's hyperparameters or investigate alternative machine learning techniques. This could entail hyperparameter optimization methods like grid search or random search.

REFERENCES

[1] ChaimaaBoukhatem; HebaYahia Youssef Department of Electrical Engineering, University of Sharjah; Ali BouNassif, Heart Disease Prediction Using Machine Learning, INSPEC Accession Number: 21668689, DOI: 10.1109/ASET53988.2022.9734880 IEEE Conference, Dubai, United Arab Emirates

[2] Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A. on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee (2022). Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. *Circulation, 145*(4), e153–e639.

[3] Janosi, Andras, Steinbrunn, William, Pfisterer, Matthias, Detrano, Robert & M.D., M.D. (1988). Heart Disease. UCI Machine Learning Repository. https://www.kaggle.com/ronitf/heart-disease-uci

[4] Victor Chang a, Vallabhanent Rupa Bhavani, An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics.Volume: 2(1).Page: 100016.Publication Year: January 2022.DOI: 10.1016/j.health.2022.100016.License: CC BY-NC-ND 4.0.

[5] https://www.nvidia.com/en-us/glossary/data-science/xgboost/

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