iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48562

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 I Jan 2023- Available at www.ijraset.com

Heart Disease Prediction Using Machine Learning

E. Nikleshwar[1], D. Sindhuja[2], T. Sai charan[3], Dr. Mohammed Mahabbob Basha[4]
[1, 2, 3]*Student,* [4]*Professor, Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Ghatkesar, Hyderabad, India.*

Abstract: ***Cardiovascular disease is a major health burden worldwide in the 21st century. Human services consumption is overpowering national and corporate spending plans because of asymptomatic infections including cardiovascular ailments. Consequently, there's an urgent requirement for early location and treatment of such ailments. The information which is gathered by data analysis of hospitals is utilized by applying different blends of calculations and algorithms for the early-stage prediction of Cardiovascular ailments. Machine Learning is one among the slanting innovations utilized in numerous circles far and wide including the medicinal services application for predicting illnesses. The proposed project is predicated on a typical machine learning algorithm k-nearest algorithm.***
Keywords: Heart Disease, KNN

I. INTRODUCTION

Heart disease (HD) is one of the critical health issues and various people are suffering from this disease around the globe. The HD occurs with common symptoms of breath shortness, human body weakness and feet are swollen. Researchers attempt to encounter an efficient technique for the detection of heart disease, because the current diagnosis techniques of heart condition aren't much effective in early time identification thanks to several reasons, such as accuracy and execution time . The diagnosis and treatment of heart conditions is extremely difficult when modern technology and doctors aren't available. The effective diagnosis and proper treatment can save the lives of many people. Consistent to the EU Society of Cardiology, 26 million approximately people of HD were diagnosed and 3.6 million annually. Most of the people in the US are suffering from heart disease. Diagnosis of HD is traditionally done by the analysis of the medical records of the patient, physical examination report and analysis of concerned symptoms by a physician. But the results obtained from this diagnosis method aren't accurate in identifying the tolerant HD. Moreover, it is expensive and computationally difficult to research. Thus, to develop a non-invasive diagnosis system supported classifiers of machine learning (ML) to resolve these issues.

Machine learning predictive models need proper data for training and testing. The performance of machine learning models are often increased if a balanced dataset is used for training and testing of the model. Furthermore, the predictive capabilities are often improved by using only balanced data.Therefore, data balancing is significantly important for model performance improvement.

II. OBJECTIVE

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of heart diseases.

III. LITERATURE REVIEW

As opposed to classical ML, which depends exclusively on the loss of accuracy of the model, the author of paper [1] offered a model with a superior approach that gains more accuracy and confidence. We tested our methodology using the Pima Indian Diabetes dataset (PIDD) and breast cancer, and with ensemble learning, we grow confident in the dependability of our models. For a very long time, the diagnosis of any disease has relied heavily on the prediction and early detection of diseases. Algorithms for machine learning (ML) have shown to be quite effective at identifying diseases and making health care decisions. Even though the majority of ML algorithms had outstanding accuracy, domain adaptability and resilience remain major issues.

While certain algorithms do well on some data sets, they struggle on others. Updated ML algorithms are required since the performance of these algorithm scan frequently alter in the future due to data variance. In paper , the author suggests a methodology for predicting HD and T2D in chronic diseases. The outlier identification method for the proposed study was random forest mixed with DBSCAN, while the data balancing method was SMOTE-ENN.

©IJRASET: All Rights are Reserved

The model was developed using three HD datasets (Stat log and Cleveland) and one T2D dataset (NHIS Korea), and the output was compared with several machine learning (ML) algorithms, such as GNB, LR, MLP, DT, and SVM. In this study, k-fold (10) cross-validation and a number of performance metrics, such as accuracy, precision, f-measure, and recall, are used to assess the model's performance. The accuracy rates for the Stat log HD dataset, Cleveland HD dataset, and NHIS T2D dataset for the model we suggested are, respectively, 97.63%, 97.69%, and 94.85%,outperforming other classification algorithms and earlier studies.

IV. IMPLEMETATION

A. *Block Diagram*

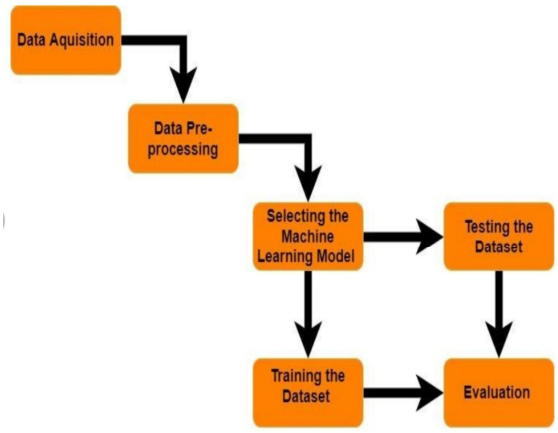


Fig.4.1.1 System Architecture

Fig: Use case Diagram

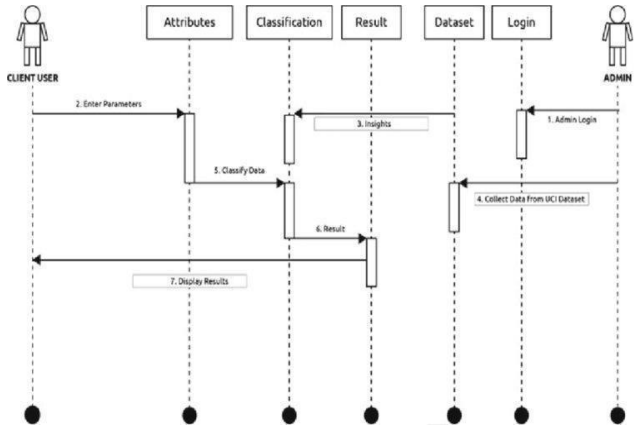


Fig 1. Block Diagram

B. *Data Flow Diagram*

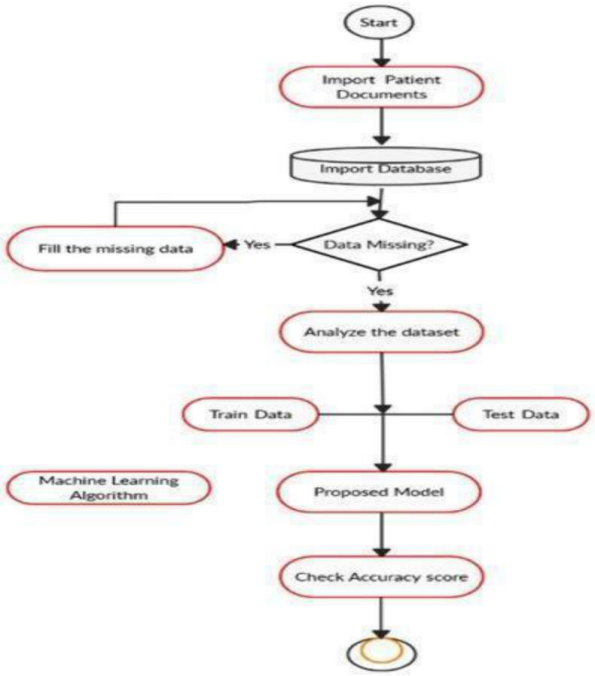


Fig.4.2.1 Data Flow Diagram

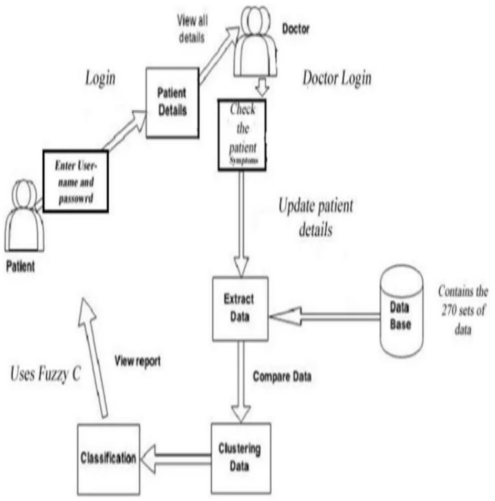


Fig.4.3.2.1 Sequence Diagram

V. METHODOGY

Various machine learning algorithms are analysed in this paper, including K nearest neighbours (KNN),logistic regression, and random forest classifiers, which can help practitioners or medical analysis diagnose heart disease accurately. This paper work entails looking over journals, papers that have been published, and recent cardiovascular disease statistics. The proposed model has a framework provided by methodology [13].

The methodology is a procedure that involves stages that convert provided data into acknowledged data patterns for the consumers' awareness. The proposed methodology consists of three steps: data collection in the first stage, substantial value extraction in the second stage, and data exploration in the third stage of pre-processing.

Depending on the algorithms utilised, data pre-processing addresses missing values, data cleaning, and normalisation . Pre-processed data are then classified using a classifier. The classifiers employed in the proposed model are KNN, Logistic Regression, and Random Forest Classifier. The suggested model is then put into practise, and its accuracy and performance are assessed using a variety of performance indicators. Using several classifiers, an efficient Heart Disease Prediction System (EHDPS)has been created in this model. For prediction, this model incorporates 13 medical variables, including age, sex, blood pressure, cholesterol, blood sugar during fasting, and chest pain.

Where yi is the ith case of the example sample and y is the prediction (outcome) of the query point. In contrast to regression, in classification problems, KNN predictions are based on a voting scheme in which the winner is used to label the query. So far we have discussed KNN analysis without paying any attention to the relative distance of the K nearest examples to the query point. In other words, we let the K neighbours have equivalent influence on predictions irrespective of their relative distance from the query point.

An alternative approach is to use arbitrarily large values of K (if not the whole prototype sample) with more importance given to cases closest to the query point.

This is achieved using so-called 'distance weighting'. K-Nearest Neighbour is one among the most popular machine learning algorithms. The K-nearest neighbour classifier may be the simplest algorithm for prediction of any dataset which includes computing distance using the Euclidean distance formula. Here K means number of neighbours, so if we take k=1 then it gives the very nearest neighbour as output. It works on a voting based system meaning while prediction it takes nearest neighbours votes. This output gets more votes that will be the predicted output value.

In a simulation and result the data set from kengle website has been taken. Simulation has been done on Jupiter notebook. From these results we can see that although most of the researchers are using different algorithms such as SVC, Decision tree for the detection of patients diagnosed with Heart disease, KNN, Random Forest Classifier and Logistic regression yield a better result to out rule them .

The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from

VI. SIMULATION AND RESULT

In a simulation and result the data set from kengle website has been taken. Simulation has been done on Jupiter notebook. From these results we can see that although most of the researchers are using different algorithms such as SVC, Decision tree for the detection of patients diagnosed with Heart disease ,KNN, Random Forest Classifier and Logistic regression yield a better result to out rule them.

The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from previous researches. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took.

Our project also tells us that Logistic Regression and KNN outperforms Random Forest Classifier in the prediction of the patient diagnosed with a heart Disease.

This proves that KNN and Logistic Regression are better in diagnosis of a heart disease. The following figure shows a plot of the number of Different feature of data set and patients that are been segregated and predicted by the classifier depending upon the age group, Resting Blood Pressure, Sex, Chest Pain:

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

A. KNN Algorithm

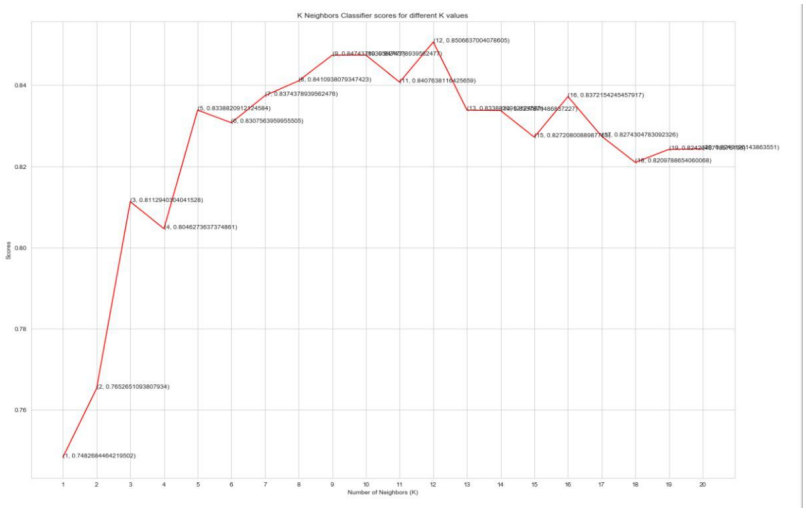
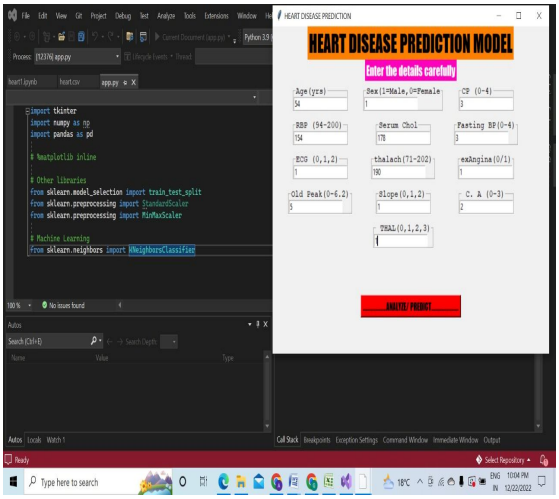
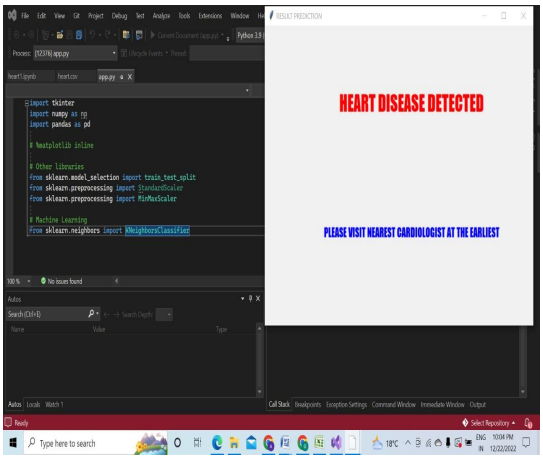


Fig: KNN algorithm prediction percentage.

B. Result





C. Advantages And Applications

1) Advantages

a) High performance

b) Process data and gives result in very less time compared to existing system.

c) Knn is widely used in various domains with high rates of success.

d) High accuracy rate

VII. CONCLUSION

Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. preventative measures are adopted as soon as possible.

VIII. FUTURE SCOPE

Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. KNN proved to be quite accurate in the prediction of heart disease.

The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature-selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

REFERENCES

[1] M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate", Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255-260, 2016.

[2] L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, et al., "Decision making in advanced heart failure: A scientific statement from the American heart association", Circulation, vol. 125, no. 15, pp. 1928-1952, 2012.

[3] L. Bui, T. B. Horwich and G. C. Fonarow, "Epidemiology and risk profile of heart failure", Nature Rev. Cardiol., vol. 8, pp. 30, 2011.

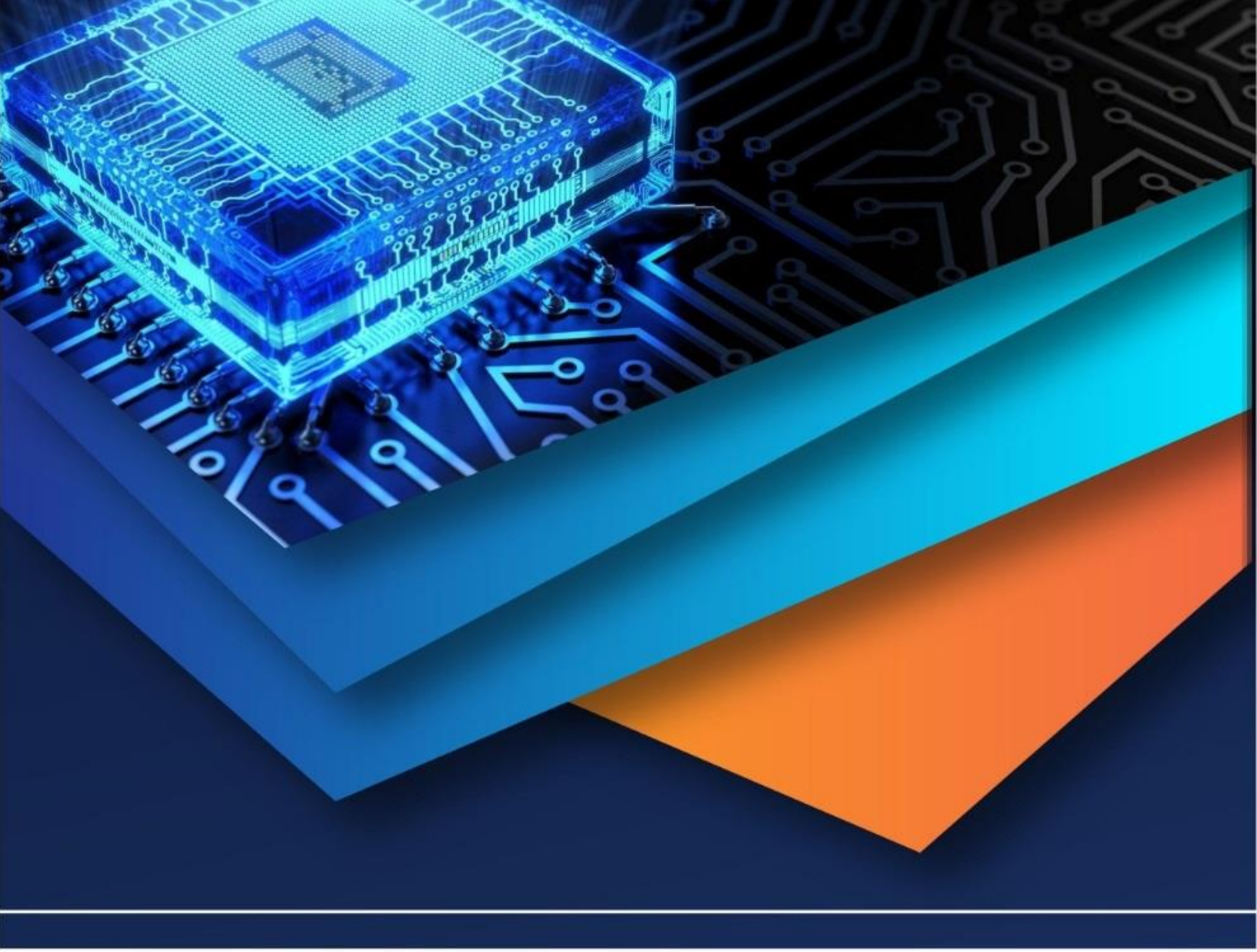
[4] S. Ghwanmeh, A. Mohammad and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart disease diagnosis", J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013.

[5] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis", Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150-154, 2011.

[6] J. Lopez-Sendon, "The heart failure epidemic", Medicographia, vol. 33, no. 4, pp. 363-369, 2011.

[7] P. A. Heidenreich,J.G.Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, et al.,"Forecasting the future of cardiovascular disease in the united states 2011.

©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)