

Heart Disease Risk Predictor

Rashmi Mishra, Pooja Saharan, Amrita Jyoti

Abstract: Cardiovascular disease is one of the focused areas in medical area because its origins, sickness and death amongst the population of the entire world. Data mining techniques play an important role to convert the large amount of raw data into meaningful information which will help in prediction and decision of Cardiovascular disease. The prediction models were technologically advanced using diverse amalgamation structures and sorting techniques such as k-NN, Naive Bayes, LR, SVM, Neural Network, Decision Tree. It is very necessary for the recital of the prediction models to choose the exact amalgamation of momentous features. The main aim of the proposed system is to develop an intelligent system using data mining modeling technique. The proposed system retrieves the data set and compares the data set with the predefined trained data set. The existing decision support system cannot predict the complex question for diagnosing the heart disease but the proposed system predicts the complex queries which will help and assist the healthcare practitioners to take appropriate decisions. This proposed system aims to provide a web platform to predict the occurrences of disease on the basis of various symptoms. The user can select various symptoms and can find the diseases with their probabilistic figures.

Index Terms: Naive Bayes, Heart disease, Data Mining, JDK, Thal

I. INTRODUCTION

Heart is the furthestmost vital part of human body and if heart gets affected due to any reason (life style, type of food, tension, cholesterol, diabetes, Hypertension) leads to another problem or affects the other parts of the body. Nowadays people want a luxurious life for this they work like a machine in order to earn lots of money hence they forget to take maintenance of their well-being. The proposed system manages complete record of the patient in a single database. Doctors use the same database to consult the patients and keep track of the records. Main focus of the proposed system is to overcome the limitation of existing systems i.e., over-time wage and gradually increase the number of patients. In the present scenario's hospital inventory management can use simple questions to diagnose heart diseases such as "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Given patient records, predict the probability of patients getting a heart disease." Main features of the proposed system are:

Revised Manuscript Received on August 5, 2019

Rashmi Mishra, Computer Science & Engineering, ABESEC, Ghaziabad, India

Pooja Saharan, Computer Science & Engineering, ABESEC, Ghaziabad, India

Amrita Jyoti, Computer Science & Engineering, ABESEC, Ghaziabad, India

Web based application (ability to communicate easily and conveniently in real-time with the legal and registered users), Time saver (Heart Disease Prediction will provide automated medical diagnosis of the patient in no time), Platform independent (The system will not be platform dependent). It will work in any operating system provided that there is an Internet connection.

II. SYSTEM INTERFACES

There is a System interface in our proposed system just to ensure the system is compatible to all the browsers and servers.

Table 1: Interfaces

Login	Log into the system as a User
Take Test	Patient will input health parameters
View Result	Health report will be displayed
Decide	Whether to check up or find your medical history

Hardware Interfaces: Minimum requirement of hardware used in the proposed system is to use 2 GB RAM, Pentium processor, Server to connect to the database. **Software Interfaces:** For the requirement of implementation of any system is to use Operating System, Android Development Kit (ADK) is suited for the reverse engineering process and for modifying etc, it will increase the efficiency very efficiently, Java development kit (JDK 8), Oracle software for the database history, Coding done in java and Oracle. **Communications Interfaces:** The system is communication software which includes many communicative interfaces like send a message write a message to user registered. **Memory Constraints** of the proposed system is 1GB+ RAM, monitor with minimum resolution of 1024x768, keyboard and mouse. **Product Functions** of the proposed system:

- Log In- User will login in into the page by entering their credentials.
- Collect Attributes- user will provide the required attributes to the website by filling up the form for further processing.
- Data Processing-The Values are used by Naïve Bayes algorithm for Calculating the probability of having a disease to the patient or not.
- Result-Then the result is displayed to the user in yes or no that whether he/she is suffering from heart disease.

- Log Out-- Log Out to checkout from your account.
Assumptions and dependencies: User must have prior details of attributes (Thala, Chest pain, resting blood pressure, cholesterol level, height, weight etc.) required for data processing.

A. Use Case Diagrams

Use Case diagram is used when the preliminary task is accomplished then use case diagram are modelled are demonstrated to current the external opinion. The main purpose of use case diagram is to imprisonment the active feature of a system. Use case diagram is similar to the activity diagram, sequence diagram, collaboration diagram and state chart. But some unique property of Use Case diagram differentiates it from other diagram such as it gathers the internal and external influences. When a system is ready to gather the functionalities then use case is prepared and actors are defined.

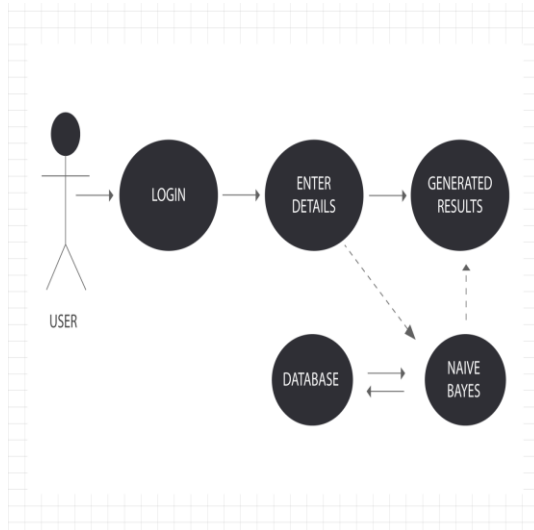


Figure 1: Use case Diagram

B. Sequence diagrams

It shows the interaction diagram that shows how the thing communicate with each other in a time sequence manner. It shows the process in between object and classes that shows the functionality of the message. If logical view is considered Sequence diagram is associated to the Use case diagram.

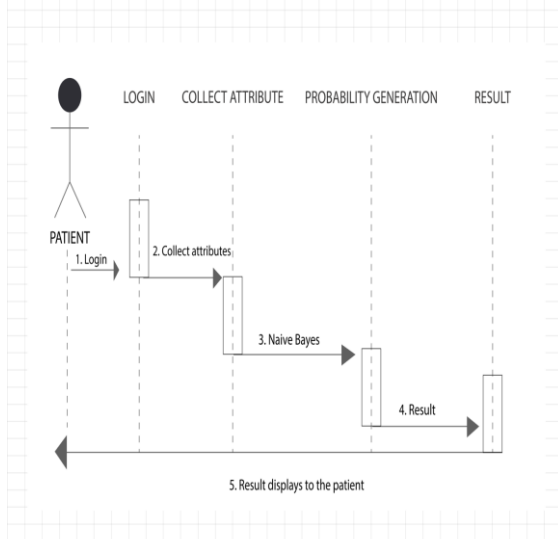


Figure 2: Activity Diagram

C. Architecture Diagram

Architecture of our system wherein there is a model which is trained using the test data and then that trained model is used to classify the data on the set of values provided by the user.

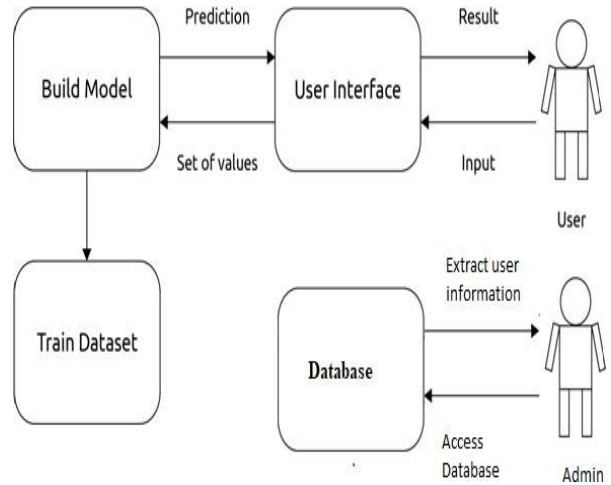


Figure 3: Architecture Diagram

III. SYSTEM DESIGN

The Proposed system consists of the following modules:

- Module – 1: - Registration and Data collection
- Module – 2: - Finding the Probability
- Module – 3: - Generating the Remedy

In Registration and Data collection module user need to register his account by entering their personal details and also gives the answers of questionnaires contained in the form. By Using the information entered by the user Database is created. This database is used by the doctors to analyses whether the patient is in risk of any Heart disease or not and generate the probability factors using the system. Depending upon the information gained by the database further questioner will generate. This questioner based on the symptoms will filled by the patient. Further probability and the symptoms generated by the system doctor generate the remedy and will displayed on the user screen.

IV. Methodology

The proposed system is using Naive Bayes Data mining technique. The training dataset with the medical attribute was obtained from the Cleveland Heart Disease database. With the training dataset heart diseases prediction are extracted. Diagnosis attribute are set to “0” or “1”. For the value “0” means user does not have any heart problem and “1” indicate user have heart disease. Using this system some of the problems also resolved such as mislaid data, identical, unpredictable data and inconsistent data. Naive base attributes used in proposed system are listed below:

Table 2: Database Description

S. N	Field	Description	Range and Values
1	Age	Age of the patient	0-100 in years
2	Sex	Gender of the patient	0-1, 1: Male, 0:Female
3	Chest Pain	Type of chest pain	Typical Angina, typical Angina Non-anginal, Asymptotic
4	Resting Blood Pressure	Blood pressure during rest	mm Hg
5	Cholesterol	Serum Cholesterol	mg / dl
6	Fasting Blood Sugar	Blood sugar content before food intake if >120 mg/dl	0: False 1: True
7	ECG	Resting Electrocardiographic results	0: Normal 1: Having ST-T wave
8	Max Heart Rate	Maximum heart beat rate.	Beats/min
9	Exercise Induced Angina	Has Pain been induced by exercise	0: No 1: Yes
10	Old Peak	ST depression induced by exercise relative to rest	0-4
11	Slope of Peak Exercise	Slope of the peak exercise ST segment	Up sloping, Flat, Down sloping
12	Ca	Number of vessels colored by fluoroscopy	0-3
13	Thal	Defect Type	3- normal, 6-Fixed Defect

The proposed system is based on the JAVA Application. In architecture of proposed system, the users need to enter the answers of the predefined questioner, then the database is created using the information entered by the user, data is categorized based on the probabilities by the system then users are able to check whether the user is prone to heart diseases or not. After the probabilities generated by the system some set of questioners are asked by the doctors, remedy is generated. System divided into further parts: Data collection, Probability generation

A. Data Collection

In data collection, user need to entered the details (personal, complaints, etc.) displayed on the screen, answers the questioner, probabilities of symptoms are generated. Using data entered by the user’s database is created. Symptoms are classified based on the naïve based attribute and matched with the trained data set. Conclusion of the classification user suffer from heart disease or not is predicted.

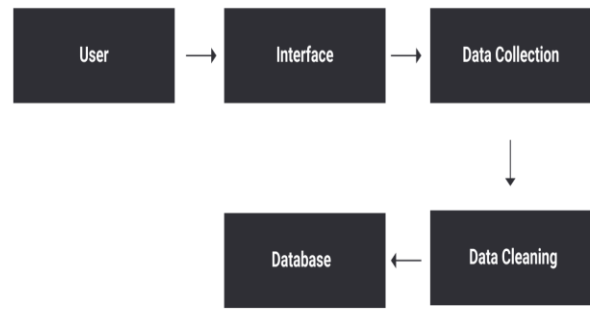


Figure 4: Data Collection

B. Classification of Symptoms

Classification of symptoms is based probability generated by the questioner filled by the users.

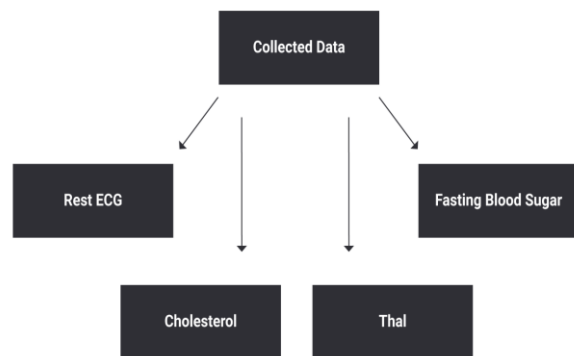


Figure 5: Classification of Symptoms

C. Probability Generation

Probability Generation, the patient will enter his records and the algorithm will generate the probability depending upon the following steps. The values in data set will be filtered in order to get discrete values. Then probability of each Naive base attribute will be calculated and then the combined probability will be shown. Bayes’ theorem is used to calculate the probability.

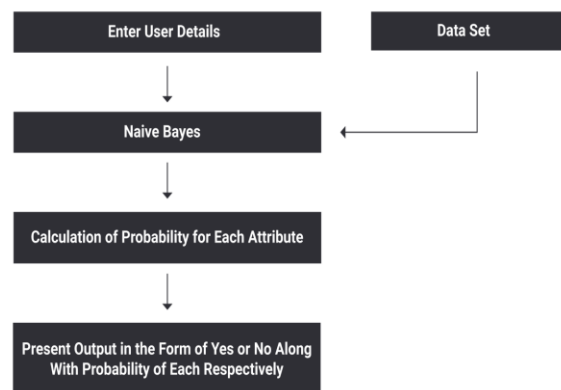


Figure 6: Probability Generation

Heart Disease Risk Predictor

D. Activity Diagram

Activity diagram is form of flowchart that shows the transition between the operation in sequence, parallel, branched and concurrent manner. It uses fork, join, etc. to deal with all the types of flow control.

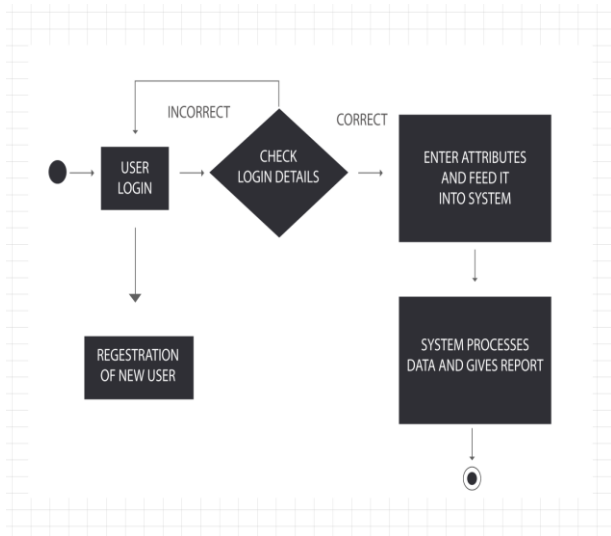


Figure 7: Activity Diagram

V. IMPLEMENTATION AND RESULTS

A. Home: First user needs to register if he is not registered using the form shown in the figure and login using their ID and password.

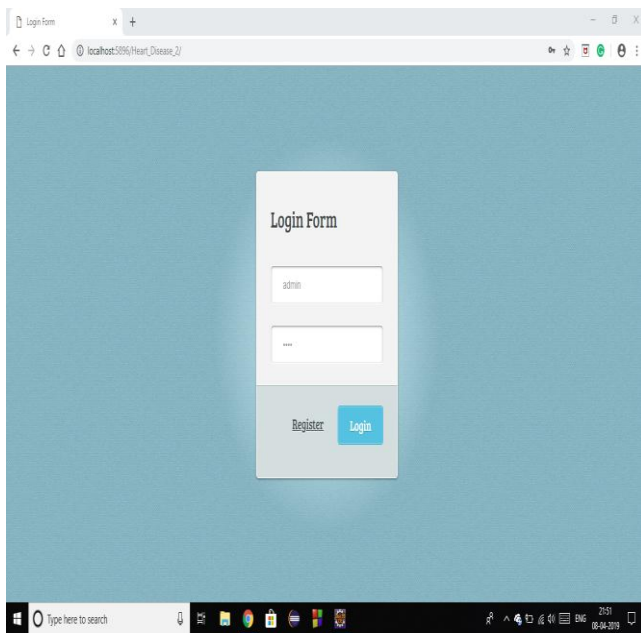


Figure 8: Home

B. Take Test: User need to enter the answers of the questioner displayed on the Patient’s detail page shown below

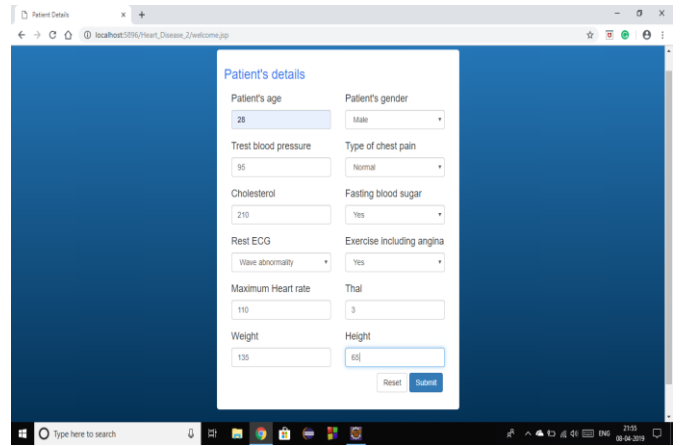


Figure 9: Take Test

C. Decide: Based on the questioner filled by the user probability of having heart disease.

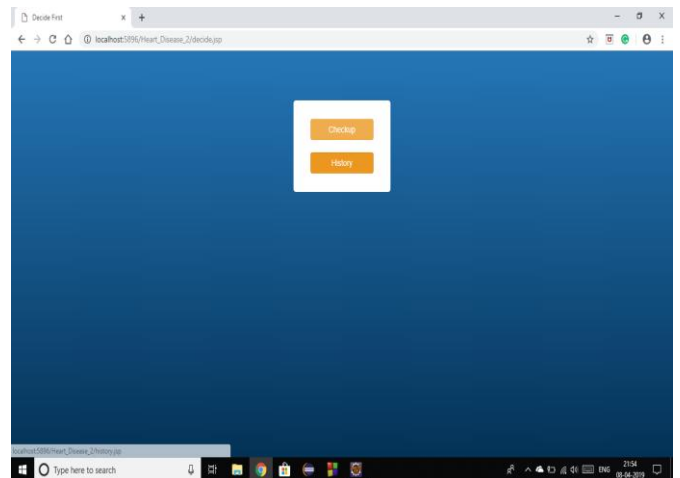


Figure 10: Decision

D. Result: Result shows that patient suffering from Heart diseases or not based on questioner asked by the doctor. Two attributes are used for prediction of heart disease. “0” means user does not have any heart problem and “1” indicate user have heart disease.

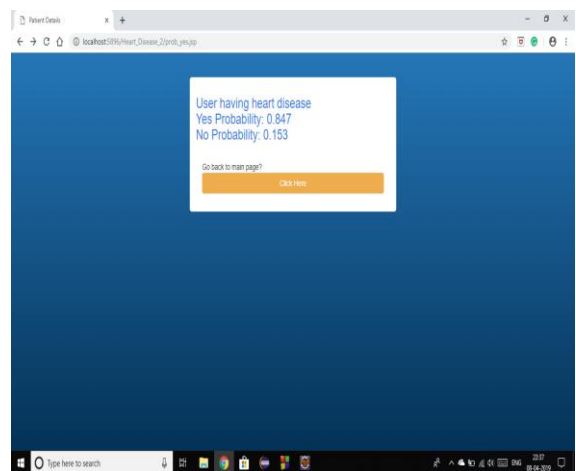


Figure 11: Result

VI. CONCLUSION

Feature of the proposed system is to minimize the paper work and data will be Safed in secure manner and provide automated medical diagnosis of the patient which store patient’s medical history into Database. It will also be generating different Diagnosis using Data Mining technique to calculate probabilities for a patient about the chances of Heart Disease. To provide remedies depending upon the symptoms specified by the patient.

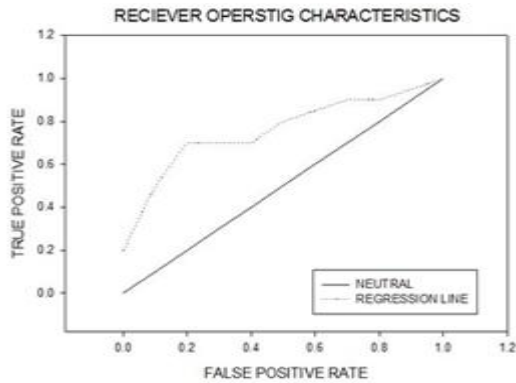


Figure 12: Naïve Bayes ROC Curve

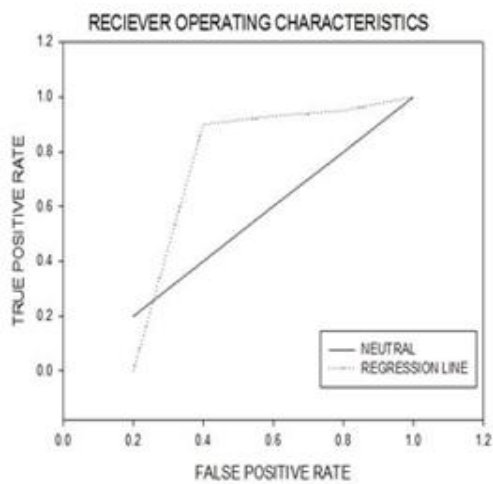


Figure 13: Decision Tree ROC Curve

A. Performance Evaluation: Proposed system is based on the JAVA application system uses data mining technique. Performance is evaluated by extracting the hidden knowledge from the database. We have examined the performance of two Data mining techniques such as Naïve Bayes and Decision Tree. As we have used Naïve Bayes Algorithm, it gives better performance as compare to the Decision Tree.

ALGORITHMS	ACCURACY
NAÏVE BAYES	80.73%
DECISION TREE	80.54%

Table 2: Comparison table

REFERENCES

1. Sana S, Sawant A, Paradkar S, Patil K,2015, Electronic Recording System- Heart Disease Prediction System, International Conference on Technologies for Sustainable Development, Feb. 04 – 06.
2. Sellappan P, Rafiah A, Intelligent Heart Disease Prediction System Using Data Mining Techniques, IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2015.
3. Tomar D, Agarwal S, Feature selection based least square twin support vector machine for diagnosis of heart disease. Int. J. Bio-Sci. Bio-Technol. 6 (2), 69–82, 2014.
4. Verma, L., Srivastava, S., Negi, P.C , A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data, J. Med. Syst. 40(7), 1–7,2016.
5. Wiharto W, Kusananto H, Herianto H, Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis. Int. J. Electr. Computer. Eng. (IJECE) 7 (2), 1023–1031, 2017.

AUTHORS PROFILE



Rashmi Mishra received the B.Tech degree in Information Technology from Institute of Technology and Management, Gida, Gorakhpur (U.P. Technical Univ., Lucknow) India, in 2009 and received M.Tech in Information Security from Ambedkar Institute of Technology, Govt. of NCT Delhi, Geeta Colony, New Delhi (Guru Govind Singh Indraprastha University, New Delhi), India and perusing Ph.D from Delhi Technological University, Delhi, India. Her current research interest includes Wireless Sensor Network, MANET, Secure Semantic Web Services, Cryptography and Network Security, Cyber Security, Blockchain. She has more than 8 years teaching experience. She is working as Assistant Professor in ABES Engineering college, Ghaziabad, Uttar Pradesh, India. She is the author/co-author of more than 16 publication in in International/National journals and conferences.



Pooja Saharan received the B.Tech degree in Computer Science and Engineering Information from University Institute of Engineering and Technology, Rohtak, Haryana (M.D.U Rohtak) India, in 2009 and received M.Tech in Computer Science and Engineering from University Institute of Engineering and Technology, Rohtak, Haryana (M.D.U Rohtak) India, in 2011 and pursuing Ph.D from Ambedkar Institute of Technology, Govt. of NCT Delhi, Geeta Colony, New Delhi (Guru Govind Singh Indraprastha University, New Delhi), India. Her current research interest includes Data mining, clustering, recommender system, blockchain, Bigdata. She has more than 8 years teaching experience. She is working as Assistant Professor in ABES Engineering college, Ghaziabad, Uttar Pradesh, India. She is the author/co-author of more than 6 publication in in International/National journals and conferences.



Amrita Jyoti , presently working as Associate professor in ABES Engineering College, Ghaziabad ,Uttar Pradesh, India. She received her B.Tech. degree in Information Technology from Kurukshetra University , India in 2003 and M.tech in Computer Science & Engineering from the Dr.APJ Abdul Kalam Technical University (Formerly as UPTU), Uttar Pradesh, India in 2011. She is the author of a Book “ Title: Data compression” which has covered all the techniques of compression in text, audio and video. Her current research area include software Engineering, software testing, data compression, data structure, JAVA, cloud computing and Blockchain. She published many research paper in National.

