



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)



# Heart Disease Prediction and Prescription Using ML

Mrunal Shinde <sup>1</sup>, Aishwarya Sawant<sup>2</sup>, Vaibhav Wakchaure <sup>3</sup>, Shubham Mali <sup>4</sup>, Prof . R .B. Pawar<sup>5</sup>

U.G. Student, Department of Information Technology Engineering, Amrutvahini College of Engineering,  
Sangamner, India<sup>1234</sup>

Associate Professor, Department of Information Technology Engineering, Amrutvahini College Of Engineering  
Sangamner, India<sup>5</sup>

**ABSTRACT:** Heart plays significant role in living organisms. Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day. To deal with the problem there is essential need of prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence (AI), it provides prestigious support in predicting any kind of event which take training from natural events. In this system, we calculate accuracy of machine learning algorithms for predicting heart disease, for this algorithms are k-nearest neighbor, decision tree, linear regression and support vector machine(SVM). We also calculate the desired prescription after detecting heart disease. For implementation, Python programming with OpenCV is best tool, which have many type of library, header file, that make the work more accurate and precise.

**KEYWORDS:** Heart Disease, Prediction, Prescription Machine Learning, Healthcare.

## I. INTRODUCTION

Health-care information systems tend to capture data in databases for research and analysis in order to assist in making medical decisions. As a result, medical information systems in hospitals and medical institutions become larger and larger

and the process of extracting useful information becomes more difficult. Traditional manual data analysis has become inefficient and methods for efficient computer-based analysis are needed. To this aim, many approaches to computerized data analysis have been considered and examined. Data mining represents a significant advance in the type of analytically tools. It has been proven that the benefits of introducing data mining into medical analysis are to increase diagnostic accuracy, to reduce costs and to save human resources. Heart Disease is destructive diseases in our society. Every year numerous people die out of cancer. The Agency for Healthcare Research and Quality (AHRQ) says that medical cost for cancer in the year 2011 in the United States was 88.7 billion dollars [1]. And out of various types of cancer, breast cancer has been one of the significant types over the past years.

Machine Learning techniques has become great potential for the health-care industry to predict health deceases by using systematic data and analytics' to identify inefficiencies and best practices that improve care and reduce costs. These techniques are fast in nature and take less time for the prediction system to improve the heart decease with more accuracy.

## II. RELATED WORK

Paper[1]"An Optimization Approach to Improve Classification Performance in Cancer and Diabetes Prediction"  
Some classification algorithms are experimented. Some optimization attempts are made to improve the algorithms performances. Detecting diseases like cancer and diabetes might be helpful for the patients as well as the doctors. From the doctors perspective, they can help the patients to identify their next step by identifying the vulnerability of cancer or prevalence of diabetes in a patient. That is how the doctors may find a way to determine the patient's condition and also if someone is at a high risk of cancer the doctors can decide on the medication and a lifestyle to help them live a better life. Techniques used for Cancer and diabetes prediction. Abundant literature has been dedicated to the classifiers of data mining and tremendous progress has been made ranging from efficient and scalable algorithm for different datasets. Authors believe that mining research has substantially broadened the scope of data analysis and will have deep impact on mining methodologies and applications in the future. However, there are still some challenging research



issues that need to be solved in searching using concept of various classifiers of data mining in the research of diabetes prediction Paper[2] "Type 2 diabetes mellitus prediction model based on data mining" WEKA toolkit and use the same Pima Indian Diabetes Dataset. realistic dataset provided by Dr. Schorling was used to test verify the model. K-means Algorithm and Logistic Regression: A novel model based on data mining techniques for predicting type 2 diabetes mellitus (T2DM). Based on a series of pre-processing procedures, the model is comprised of two parts, the improved K-means algorithm and the logistic regression algorithm. The Pima Indians Diabetes Dataset and the Waikato Environment for Knowledge Analysis toolkit were utilized to compare results with the results from other researchers.

Paper[3] An expert Personal Health System to monitor patients affected by Gestational Diabetes Mellitus. The graphs and charts provided in this study were of simple nature, as a future development, more elaborated charts representation of the data that puts physiological patterns in relation may allow the medical doctors to have a better understanding of the health of the patient, thus improving their ability to effectively modify the treatment plan. Temporal Abstraction with Data Mining: Hemodialysis patients might suffer from unhealthy care behaviours or longterm dialysis treatments and need to be hospitalized. If the hospitalization rate of a hemodialysis center is high, its service quality will be low. Therefore, decreasing hospitalization rate is a crucial problem for health care centers. This study combines temporal abstraction with data mining techniques for analyzing dialysis patients biochemical data to develop a decision support system. The mined temporal patterns are helpful for clinicians to predict hospitalization of hemodialysis patients and to suggest immediate treatments to avoid hospitalization. Paper[4] An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus. In this study, 54 UCI datasets are used to evaluate the performance of various classification algorithms. These datasets are characterized according to data sizes, features, classes. Electromagnetism-like Mechanism: The use of artificial intelligence based data mining techniques for massive medical data classification and diagnosis has gained its popularity, whereas the effectiveness and efficiency by feature selection is worthy to further investigate. They Presented a novel method for feature selection with the use of opposite sign test (OST) as a local search for the electromagnetism-like mechanism (EM) algorithm, denoted as improved electromagnetism-like mechanism (IEM) algorithm. Nearest neighbour algorithm is served as a classifier for the wrapper method. The proposed IEM algorithm is compared with nine popular feature selection and classification methods. Forty-six datasets from the UCI repository and eight gene expression micro array datasets are collected for comprehensive evaluation. Non-parametric statistical tests are conducted to justify the performance of the methods in terms of classification accuracy and Kappa index. The results confirm that the proposed IEM method is superior to the common state-of- art methods.

Paper[5] Comparison of three data mining models for predicting diabetes or prediabetes by risk factors Author: Xue-Hui Meng Year : 2012. The characteristics of participants and Pearson Chi-square test results between two groups. The sensitivity analysis performed, logistic regression model, ANN model , C5.0 decision tree model, to determine the order detailed predictions produced from the training and testing datasets are presented in the form of confusion matrices. Pre-Diabetes Detection by Risk Factors: Purpose of this was to compare the performance of logistic regression, artificial neural networks (ANNs) and decision tree models for predicting diabetes or pre-diabetes using common risk factors. A standard questionnaire was administered to obtain information on demographic characteristics, family diabetes history, anthropomorphic measurements and lifestyle risk factors. Then 8 "Improve the Performance of Cancer and Diabetes Detection by Using Novel Technique of Machine Learning" He developed three predictive models using 12 input variables and one output variable from the questionnaire information we evaluated the three models in terms of their accuracy, sensitivity and specificity. The logistic regression model achieved classification accuracy.

### III. METHODOLOGY

The process starts with data manipulation. Next, four models will be investigated for finding a prediction model. Then, accuracy of each model will be calculated and compared for seeking the best model. Detecting diseases like heart disease might be helpful for the patients as well as the doctors. From the doctors' perspective, they can help the patients to identify their next step by identifying the vulnerability of cancer or prevalence of diabetes in a patient. The study ends up with creating a web application.

A. System Design:

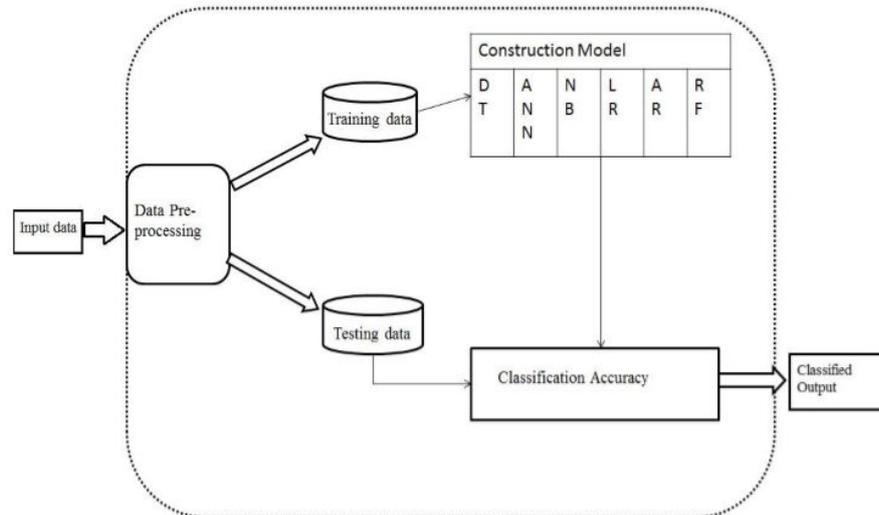


Fig 1: System Architecture

B. Modules:

1. Data Preprocessing

First, each attribute’s correlation to DM is analysed. e.g. number of pregnancies is transformed into a nominal attribute. The value 0 indicates on-pregnant and 1 indicates pregnant. The complexity of the dataset was reduced by this process. second, some missing and incorrect values in the dataset due to errors are removed. for example, body mass index could not be 0, which indicates that the real value was missing. To reduce the influence of meaningless values, the training data is used to replace all missing values.

2. Data Classification

The Model Consists Of Double-Level Algorithms. In The First Level, Authors Used The Improved K-Means Algorithm To Remove Incorrectly Clustered Data. The Optimized Dataset Was Used As Input For Next Level. Then, They Used The Logistic Regression Algorithm To Classify The Remaining Data.

3. Improved K-Means Clustering Algorithm Cluster analysis aims at partitioning the observations into disparate clusters so that observations within the same cluster are more closely related to each other than those assigned to different clusters. The K-means is one of the most popular cluster algorithms. It is a typical distance-based cluster algorithm, and the distance is used as a measure of similarity, i.e., the smaller distance between objects shows the greater similarity.

C. Algorithms

1) Algorithm for Heart Disease Prediction SVM :

Input: D dataset, on-demand features, aggregation-based features,

Output: Classification of Application

for each application App-id in D do

Get on-demand features and stored on vector x for App-id

x.add ( Get-Features(app-id));

end for

for each application in x vector do

Fetch first feature and stored in b, and other features in w.

$hw, b(x) = g(z)$  here  $z = (w^T x + b)$

if  $(z > 0)$

assign  $g(z)=1$ ;

else  $g(z)=-1$ ;

end if

end



2) FOR PREDICTION DECISION TREE

1. Check if algorithm satisfies termination criteria
2. Computer information-theoretic criteria for all attributes
3. Choose best attribute according to the information-theoretic criteria
4. Create a decision node based on the best attribute in step
5. Induce (i.e. split) the dataset based on newly created decision node in step 4
6. For all sub-dataset in step 5, call C4.5 algorithm to get a sub-tree (recursive call)
7. Attach the tree obtained in step 6 to the decision node in step 4
8. Return tree

IV. EXPERIMENTAL RESULTS

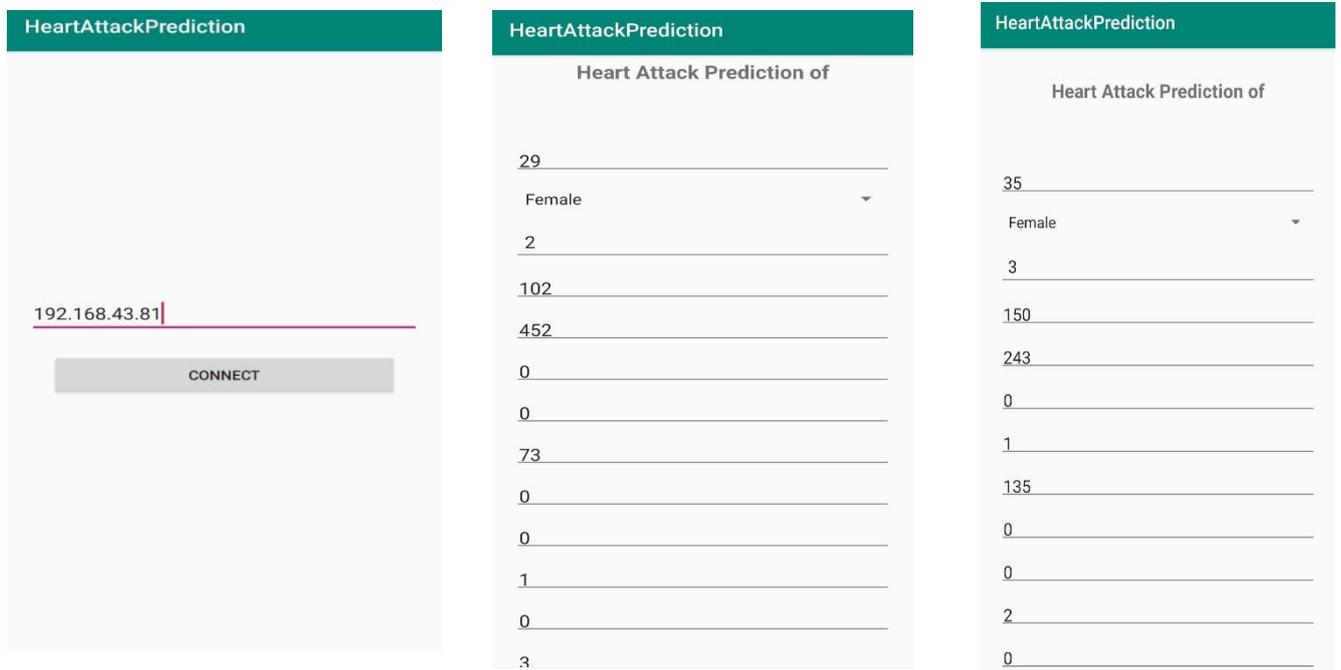


Fig: Connection is done with Server (a), Input Parameters can be seen in (b), Input Variables can be seen in 9(c)

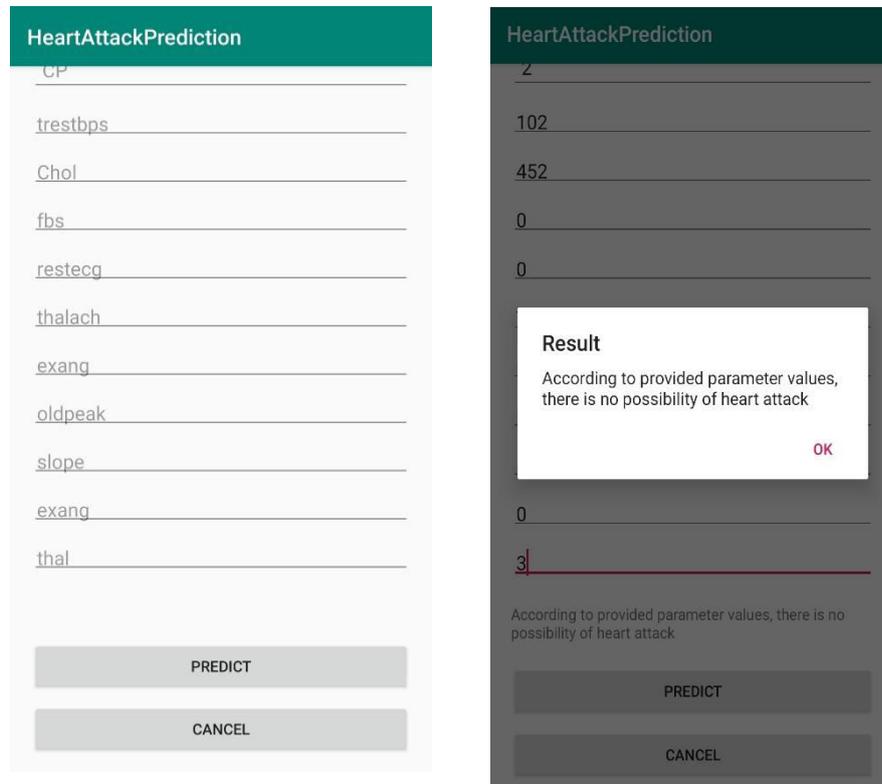


Fig: Prediction Of heart Disease is done on Value Inputed

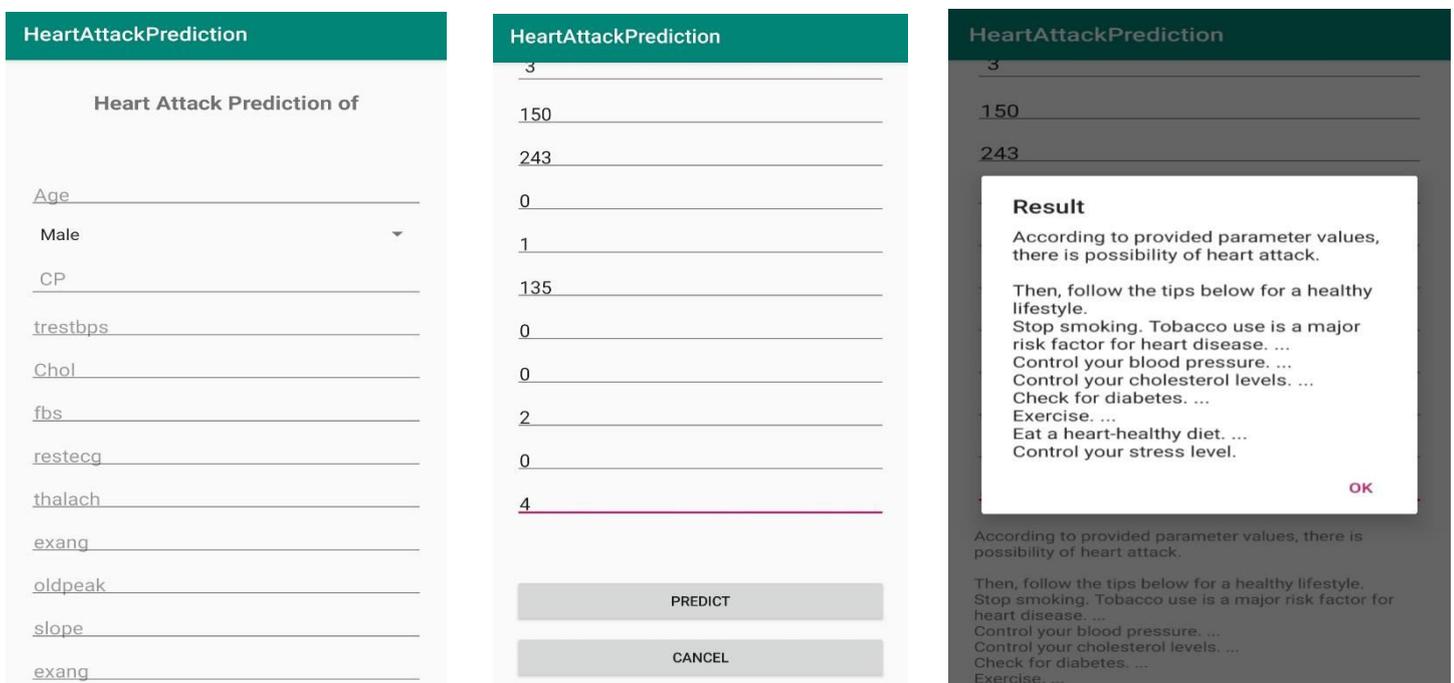


Fig: Prediction Of heart Disease is done on Value Inputted And Prescription is given



## V. CONCLUSION

The proper filters were utilized to improve the validity and rationality of the dataset. The proposed model that consisted of both cluster and class method ensured the enhancement of prediction accuracy. One of the proposed model's benefits is that it avoids deleting overmuch original data. The benefit is that the model can apply in the Pima Indian Diabetes Dataset as well as other various datasets. While the limitation is that it consumes more time during the part of preprocessing. It assures less time consuming and maximum retention of original data. The main problems solved are improving accuracy of prediction model and making the model to adapt to different datasets.

## REFERENCES

- [1] Mustakim Al Helal, Atiqul Islam Chowdhury, Ashraful Islam, Eshtiaq Ahmed, Md. Swakshar Mahmud, Sabrina Hossain, "An Optimization Approach to Improve Classification Performance in Cancer and Diabetes Prediction", International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019 (Base Paper)
- [2] Stefano Bromuri, Serban Puricel, Rene Schumann, Johannes Krampf, Juan Ruiz and Michael Schumacher, "An expert Personal Health System to monitor patients affected by Gestational Diabetes Mellitus: A feasibility study", Journal of Ambient Intelligence and Smart Environments 8(2016) 219–237.
- [3] Gyorgy J. Simon, Pedro J. Caraballo, Terry M. Therneau, Steven S. Cha, M. Regina Castro and Peter W. Li, "Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING.
- [4] Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", Informatics in Medicine Unlocked 10 (2018) 100–107.
- [5] Xue-Hui Meng a, Yi-Xiang Huang a, Dong-Ping Rao b, Qiu Zhang a, Qing Liu b, "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors", Kaohsiung Journal of Medical Sciences (2013) 29, 93e99.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details