IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# MACHINE LEARNING FOR HEART DISEASE DETECTION

**Pooja Chavan, Rakshanda Jumbad, Aasha Shinde, Urmila Wadikar**

**Prof. T.B. Tambe**

Department of computer Engineering, P K Technical Campus,

Chakan, Maharashtra, India

**Abstract** - With the rapid advancement of technology and data, the healthcare domain is one of the most significant study fields in the contemporary era. The enormous amount of patient data is tough to manage. Big Data Analytics makes it easier to manage this information.

There are numerous ways for treating various ailments all throughout the world. Machine Learning is a new approach that aids in disease prediction and diagnosis. Machine learning is used to predict disease based on symptoms in this article.
On the presented dataset, machine learning algorithms such as Naive Bayes, Decision Tree, and Random Forest are used to predict the disease. It is implemented using the Python programming language. The study demonstrates which algorithm is the most accurate. The performance of an algorithm on a given dataset determines its accuracy.

## INTRODUCTION

When someone is currently afflicted with an illness, they must visit a doctor, which is both time demanding and expensive. It may also be tough for the user if they are outside of the reach of doctors and hospitals because the condition cannot be diagnosed. So, if the following procedure can be conducted using an automated programme that saves both time and money, the patient's life may be made simpler.This can help with the procedure Other Heart Disorders Data mining techniques are used to analyse the level of risk in the prediction system.Heart Disease Predictor is a web-based programme that predicts a user's heart disease based on their symptoms. Data sets from many health-related websites have been collected for the heart disease prediction system. The user will be able to determine the likelihood of a disease based on the symptoms provided using the heart Disease Predictor. People are always keen to learn new things, especially since the use of the internet increases every day.When a problem emerges, people try to look it up on the internet. Hospitals and doctors have less access to the internet than the general public. When a person is suffering from a sickness, they do not have many options. Because people have access to the internet 24 hours a day, this method can be beneficial to them.

## MOTIVATION

The present techniques for predicting heart disease use a short dataset. The goal of our system is to work with a larger dataset in order to improve overall system efficiency. Our technology is simple to use and provides quick results. The heart disease prediction function is based on natural language processing, which allows users to report their health problems.

## OBJECTIVES

• To implement a Nave Bayes Classifier that classifies diseases based on the user's input.

• To create a web-based platform for illness prediction in the heart.

• To provide immediate output to the user (heart predict the disease Quickly)

### System Architecture

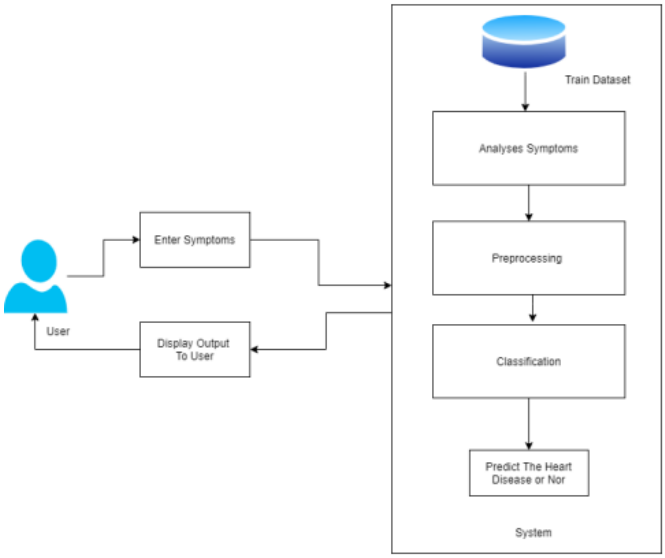



Figure 5.1: System Architecturel

## LITERATURE SURVEY

1.Paper Name: Design And Implementing Heart Disease Prediction Using Naives Bayesian Author:Anjan Nikhil Repaka,Sai Deepak Ravikanti. Abstract :—Data mining, a great developing technique that revolves around exploring and digging out significant information from massive collection of data which can be further beneficial in examining and drawing out patterns for making business related decisions. Talking about the Medical domain, implementation of data mining in this field can yield in discovering and withdrawing valuable patterns and information which can prove beneficial in performing clinical diagnosis. The research focuses on heart disease diagnosis by considering previous data and information. To achieve this SHDP (Smart Heart Disease Prediction) is built via Navies Bayesian in order to predict risk factors concerning heart disease. The speedy advancement of technology has led to remarkable rise in mobile health technology that being one of the web application. The required data is assembled in a standardized form. For predicting the chances of heart disease in a patient, the following attributes are being fetched from the medical profiles, these include: age, BP, cholesterol, sex, blood sugar etc... The collected attributes acts as input for the Navies Bayesian classification for predicting heart disease. The dataset utilized is split into two sections, 80% dataset is utilized for training and rest 20% is utilized for testing. The proposed approach includes following stages: dataset collection, user registration and login (Application based), classification via Navies Bayesian, prediction and secure data transfer by employing AES (Advanced Encryption Standard). Thereafter result is produced. The research elaborates and presents multiple knowledge abstraction techniques by making use of data mining methods which are adopted for heart disease prediction. The output reveals that the established diagnostic system effectively assists in predicting risk factors concerning heart diseases.

2.Paper Name:Application of Machine Learning in Disease Prediction Author: Pahulpreet Singh Kohli,Shriya Arora Abstract :The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

3.Paper Name:Disease phenotype similarity improves the prediction of novel disease-associated microRNAs Author:Duc-Hau Le Description : —Many studies have shown roles of miRNAs (microRNAs) on human disease and a number of computational methods have been proposed to predict such associations by ranking candidate microRNAs according to their relevance to a disease. Among them, network-based methods are becoming dominant since they well exploit the “disease module” principle in miRNA functional similarity networks. Of which, Random Walk with Restart (RWR) algorithmbased method on a miRNA functional similarity network, namely RWRMDA, is state-of-theart one. The use of this algorithm was inspired from its success in predicting disease gene because “disease module” principle also exists in protein interaction networks. Besides, many other algorithms were also designed for prediction of disease genes. However, they have not yet

been utilized for disease microRNA prediction. In this study, we proposed a method, namely RWRHMDA, for prediction of disease-associated miRNAs. This method was based on RWRH algorithm, which was successfully proposed for disease gene prediction on a heterogeneous network of genes and disease phenotypes. In particular, we used this algorithm to rank disease candidate miRNAs on a heterogeneous network of phenotypes and miRNAs, which was constructed by integrating a shared target gene-based microRNA functional similarity network and a disease phenotype similarity network. Comparing the prediction performance of RWRHMDA with that of RWRMDA on a set of 35 disease phenotypes, we found that RWRHMDA significantly outperformed RWRMDA irrespective of parameter settings since it better exploited "disease module" principle. In addition, using RWRHMDA method, we identified eight novel Alzheimer's diseaseassociated miRNAs

4.Paper Name:Efficient Heart Disease Prediction System using Decision Tree Author: Purushottam,Prof. (Dr.) Kanak Saxena,Richa Sharma Description :— Cardiovascular disease (CVD) is a big reason of morbidity and mortality in the current living style. Identification of Cardiovascular disease is an important but a complex task that needs to be performed very minutely, efficiently and the correct automation would be very desirable. Every human being can not be equally skillful and so as doctors. All doctors cannot be equally skilled in every sub specialty and at many places we don't have skilled and specialist doctors available easily. An automated system in medical diagnosis would enhance medical care and it can also reduce costs. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the user's requirement. The performance of the system is evaluated in terms of classification accuracy and the results shows that the system has great potential in predicting the heart disease risk level more accurately

5.Paper Name:Diabetes Disease Prediction Using Data Mining Author:Deeraj Shetty,Kishor Rit,Sohail Shaikh. Abstract:Data mining is a subfield in the subject of software engineering. It is the methodical procedure of finding examples in huge data sets including techniques at the crossing point of manufactured intelligence, machine learning, insights, and database systems. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

6.paper Name:Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis Author:Shang-Ming Zhou1 *, Fabiola Fernandez-Gutierrez1 , Jonathan Kennedy1 , Roxanne Cooksey1 , Mark Atkinson1 , Spiros Denaxas2 , Stefan Siebert3 , William G. Dixon4 , Terence W. O'Neill4 , Ernest Choy5 , Cathie Sudlow6 , UK Biobank Follow-up and Outcomes Group7¶, Sinead Brophy. Abstract:1) To use data-driven method to examine clinical codes (risk factors) of a medical condition in primary care electronic health records (EHRs) that can accurately predict a diagnosis of the condition in secondary care EHRs. 2) To develop and validate a disease phenotyping algorithm for rheumatoid arthritis using primary care EHRs.This study linked routine primary and secondary care EHRs in Wales, UK. A machine learning based scheme was used to identify patients with rheumatoid arthritis from primary care EHRs via the following steps: i) selection of variables by comparing relative frequencies of Read codes in the primary care dataset associated with disease case compared to nondisease control (disease/non-disease based on the secondary care diagnosis); ii) reduction of predictors/associated variables using a Random Forest method, iii) induction of decision rules from decision tree model. The proposed method was then extensively validated on an independent dataset, and compared for performance with two existing deterministic algorithms for RA which had been developed using expert clinical knowledge.

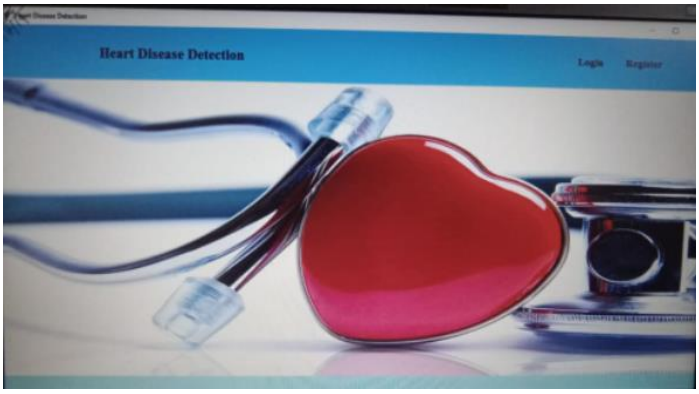
7. Paper Name:Prediction of Disease Infection of Welsh Onions by Rust Fungus Based on Temperature and Wetness Duration Author:Hiroyuki Takanashi, Hiromitsu Furuya and Seiji Chonan Abstract:The style of agriculture practiced in Japan and other countries in Asia is small scale compared to that in North America. While systematic production and management systems have been maintained in Europe and America, Japanese agricultural style tends to depend on past experiences, and application of agricultural chemicals is guided by the calendar and past experiences. Japan is also advanced in the field of plant disease prediction. This paper focuses on a prediction model of disease infection for a foliar parasite on Welsh onions. Rust fungus disease is the most typical disease on Welsh onions, and the Weibull probability density function is appropriate for approximating the infection rate of the disease. The model utilizes temperature and wetness duration to predict the infection of Welsh onions by rust fungus. Producers, then, can use the model to determine the day on which the infection rate will rapidly increase, then carry out appropriate countermeasures to the disease. The proposed prediction method is applicable to several infections found throughout Asia.

8.Paper Name:Neurodegenerative disease prediction based on gait analysis signals acquired with force-sensitive Author:ger Selzler 1 , James R. Green 1 , Rafik Goubran. Abstract:Neurodegenerative diseases such as Parkinson's Disease (PD), Huntington Disease (HD), and Amyotrophic Lateral Sclerosis (ALS) affect the lives of thousands of people around the world. One of the consequences of such diseases occurs in the motor neurons of the patients, resulting in problems in movement, causing a change in gait pattern. Force sensitive resistors can be used to measure the force/pressure between the shoe and the patient's foot, providing information about the gait dynamics when the patient walks. This project uses signals from the Gait Dynamics in Neuro-Degenerative Disease database to extract features for classification of neurodegenerative diseases (NDD). Manually labelled features from the database are used for comparison with previous studies. Time series signals is also used, where algorithms for signal reliability, feature extraction and feature selection are implemented, allowing real-time signal processing and classification. Multiple feature sets are used for classification with algorithms such as K-nearest neighbor, Support Vector Machines, and Decision Trees, and the performance of these algorithms are then reported. This study presents a realtime system with accuracy exceeding 82% for the aforementioned diseases. Finally, a discussion about possible improvements for future studies are presented.

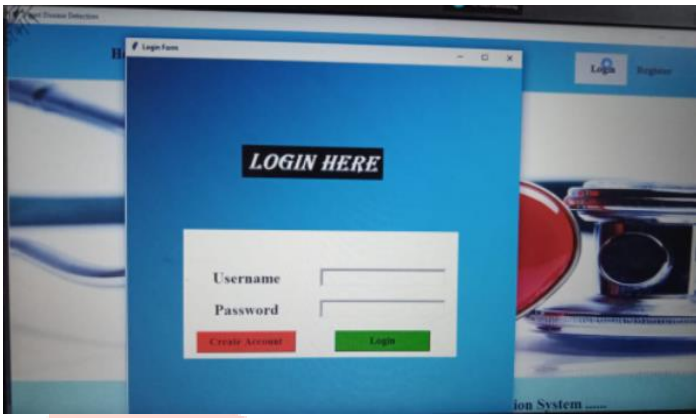
## CONCLUSION

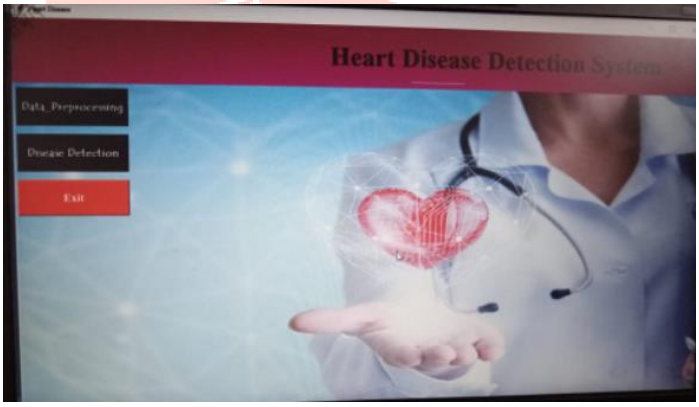
The goal of this project is to predict heart disease based on symptoms. The project is set up so that the system takes the user's symptoms as input and creates an output, such as predicting heart disease. An average accuracy probability of 55 was achieved. The grails framework was used to successfully develop the heart disease predictor.

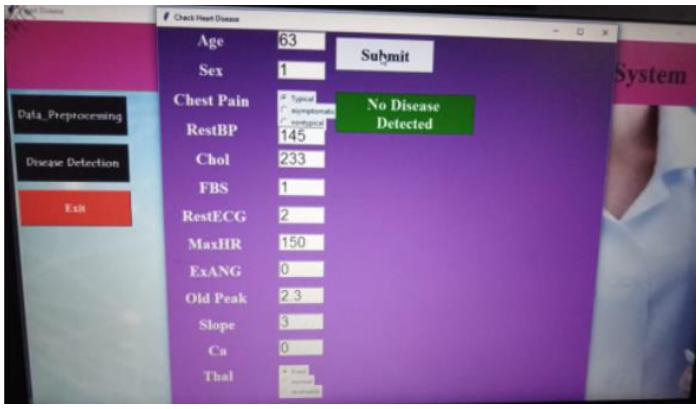
## RESULTS








Patient Report

Heart Disease Report

Patient Name : sagar jadhav

Address : chakan Mobile No : 88888888

| Sr.No | Parameters | Values |
|---|---|---|
| 1 | Age | 25 |
| 2 | Sex | Female |
| 3 | Chest Pian | 2 |
| 4 | Rest_BP | 34 |
| 5 | Chol | 34 |
| 6 | FBS | 53 |
| 7 | Rest ECG | 32 |
| 8 | Max_HR | 32 |
| 9 | Ex_ANG | 23 |
| 10 | Old Peak | 32 |
| 11 | Slop | 23 |
| 12 | Ca | 23 |
| 13 | Thal | 2 |

Result:

As per input data and system model **Heart Disease Detected** for Respective Patient.

***Kindly Follow Medicatins***

Patient Report

Heart Disease Report

Patient Name : sagar jadhav

Address : chakan Mobile No : 88888888

| Sr.No | Parameters | Values |
|---|---|---|
| 1 | Age | 25 |
| 2 | Sex | Male |
| 3 | Chest Pian | 2 |
| 4 | Rest_BP | 0 |
| 5 | Chol | 00 |
| 6 | FBS | 0 |
| 7 | Rest ECG | 0 |
| 8 | Max_HR | 0 |
| 9 | Ex_ANG | 0 |
| 10 | Old Peak | 0 |
| 11 | Slop | 0 |
| 12 | Ca | 0 |
| 13 | Thal | 2 |

Result:

As per input data and system model **Heart Disease Not Detected** for Respective Patient.

***Kindly Follow Medicatins***

## REFERENCES

1. D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3033–3049, 2015.

2. M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017.

3. M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring," ACM/Springer Mobile Networks and Applications, Vol. 21, No. 5, pp. 825C845, 2016

4. M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," IEEE Access, DOI: 10.1109/ACCESS.2016.2641480, 2016.

5. J. Wang, M. Qiu, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform," Journal of Systems Architecture, vol. 72, pp. 69–79, 2017.

6. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Healthcps: Healthcare cyber-physical system assisted by cloud and big data," IEEE Systems Journal, 2015.

7. K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the

vehicular networks," IEEE Transactions on Industrial Informatics, 2016

8. K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for iot localization in smart buildings," IEEE Transactions on Automation Science and Engineering, vol. 13, no. 3, pp. 1294–1307, 2016.

9. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

10. S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson, and P. J. O'Connor, "Data mining for censored time-to-event data: a bayesian network model for predicting cardiovascular risk from electronic health record data," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1033–1069, 2015.

11. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

12. J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas and A. Vasilakos, "A Manufacturing Big Data Solution for Active Preventive Maintenance", IEEE Transactions on Industrial Informatics, DOI: 10.1109/TII.2017.2670505, 2017.

13. N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, "Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp. 855–864.