International Journal of Innovative Research in Engineering
Volume 4, Issue 3(May-June 2023), PP: 90-95.
www.theijire.com ***ISSN No: 2582-8746***

Multiple Human Disease Detection System

Shreya Agrahari[1], Shruti Pandey[2], Vijendra Pratap Singh[3]
[1,2]Computer Science and Engineering, Institute of Technology and Management, Gorakhpur, India.
[3]Assistant Professor Department of Computer Science and Engineering, Institute of Technology and Management, Gorakhpur, India.

How to cite this paper:
Shreya Agrahari[1], ShrutiPandey[2], Vijendra Pratap Singh[3], "Multiple Human Disease Detection System", IJIRE-V4I03-90-95.

Abstract: *This project is an attempt to help one to predict the disease he/she is having through the symptoms and the correct readings of the bodily vitals needed. There are times when people keep on ignoring health issues due to high medical fees. This may lead to severe issues later and even death. If not covered by insurance, medical bills can be a menace. This website is an approach in reducing the effort of a normal person by estimating the kind of disease one has and its severity. We have designed a disease prediction system using multiple machine learning algorithms. Based on the symptoms, age, and gender of an individual, the diagnosis system gives the output giving the information about whether the user is suffering from that particular disease or not. It provides a simple yet effective approach for predicting the disease, if the provided values of vitals are accurate. The user will experience a simple yet effective User Interface and pleasing design.*

Key Word: *Disease Prediction; Machine Learning Algorithm; Supervised Learning; Diagnosis System; User Interface.*

I.INTRODUCTION

Human life is evolving every single day, but is the health of the generation improving or declining? Life is full of uncertainty. Every now and then we come across many people suffering from fatal health issues due to late identification of diseases. The study says, one in two Indian diabetics are unaware of their condition. Nearly 463 million people in the world have diabetes. One in four deaths in India are now because of CVDs with ischemic heart disease and stroke responsible for more than 80% of this burden. The study estimates more than 50 million people in the world, considering the adult population, would be affected with chronic liver disease. But, it can be prevented by identifying the disease in its early stage. The project "Disease Prediction using Machine Learning" is developed to identify general disease in earlier stages. Now-a-days, people put health as a secondary priority, which leads to various problems. According to research, 40% of people ignore the symptoms, due to fear of facing financial issues or other generic reasons. Many cannot afford to consult a doctor or some are very busy and have a tight schedule, but ignoring the recurring symptoms for a long period of time may have severe consequences to their health. According to research 70% of people in India suffer from common diseases and the mortality rate is 25%, mostly due to ignorance in early stages. The main motive to develop this project is that a user can conveniently have a check-up of their health, if they have any of the symptoms.

II. MATERIALS AND METHODS

Related Work

In this paper [1] "Multiple Human Disease prediction using Machine Learning over Big Data". Big data is the fastest concept in the current trend, so this concept is applied in more fields. Big data is most widely used in every field because it is very large. Big data is applied in the medical field. Both sides develop better growth in both fields, that is, big data is applied in medical fields and the medical fields at the same time increases the growth in the big data field. Big data helps to achieve better growth in the medical and health care sectors. It additionally, provides more merits gives, (i) medical data analysis with accuracy, (ii) early prediction for disease, (iii) patient-oriented data with accuracy, (iv) the medical data, is securely stored and used in many places, (v) incomplete regional data are reduced and give the accurate result. The goal of the concept is to choose the region and collect the hospital data or medical data of the particular selected region, this process is using the machine learning algorithm. The advantages of the concept is, better feature description and better accuracy, and the disadvantages of this system is, this feature is only applicable for the structured data so it is not good in disease description.

Authors in this paper [2] have proposed the concept of machine learning-based disease prediction using big data to overcome the machine learning drawbacks. The smooth progress of big data is moved in the biomedical and healthcare communities in hospitals for accurate results in any experiment result. This concept is (a) reduces the unfinished data and (b) effective disease prediction.

The paper [3] author has presented the info mining concept "Multiple Human Disease prediction using Machine Learning". The best growth of the stage is developing that technique into the healthcare basis, the data analysis is an important part of every field. Data mining predicts the information for healthcare is called rapid growth of the medical care field. The existing one is designed for the purpose of (i) analyzing, (ii) managing, (iii) predicting healthcare data, it is to describe the overall healthcare systems.

Authors, presents the survey paper [4] for "Multiple Human Disease prediction using Machine Learning over Big Data". Can develop the medical specialty basis this concept is applied to produce the medical data into mass medical data, which means the data which is enlarged. The goal of this concept is targeted: the simplest data is stored into the space of medical massive data analysis, called "medical data analysis in massive collection". It produces the accuracy and it reaches the 4.8% speed faster than the CNN-UDRP. It only focuses on these three data, (a) structured data, (b) text data, (c) structured and text data. In this proposed system it improves the medical data-oriented term.

In this paper [5] the author has presented, "personalized disease prediction care from harm using big data", for healthcare analysis. This concept describes the medical field as a rich data industry because it holds healthcare records, also. The daily treatment records are increased every day, that is it includes the number of transactions, and the patient information is stored and retrieved from the database. The medical treatment records are updated every day because every day improves the patient's health improvements based on treatment. It gives the right solutions for various sorts of diseases.

Procedure methodology

Disease Prediction system based on predictive modeling predicts the disease of the user on the basis of the symptoms that user provides s an input to the system. Disease Prediction is done by implementing the various machine learning algorithms. In this project, we are proposing a system which is used to predict more than one disease and we are concentrating on providing immediate and accurate output.

Our work aims to develop a web application to detect diseases like Parkinson's disease, diabetes, and heart diseases, etc., using machine learning models. The machine learning methodologies used in our proposed work are as follows:

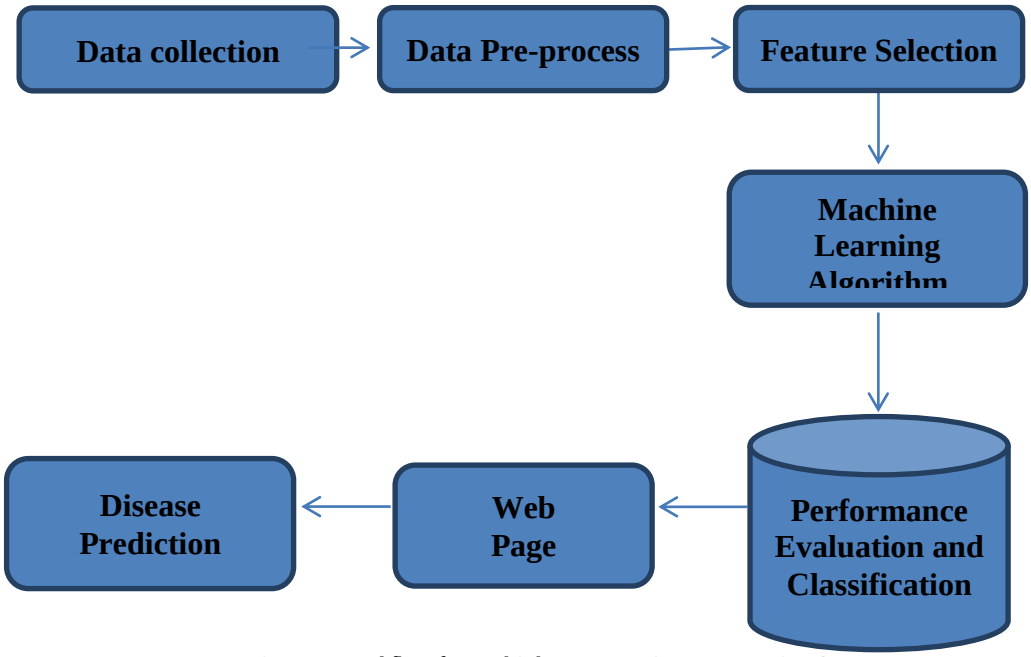


Figure1: Workflow for Multiple Human Disease Detection System

Step 1: Data Set Collection

Data collection has been done from the internet to identify the disease here the real symptoms of the disease are collected i.e., no dummy values are entered. The symptoms of the disease are collected from different health related websites. Preparing and choosing the right dataset is one of the most crucial steps in training an AI/ML model. It can be the determinant between the success and failure of the AI/ML development project.

Step 2: Data Preprocessing

Data Preprocessing includes the steps we need to follow to transform data so that it may be easily parsed by the machine. The main agenda for a model to be accurate and precise in predictions is that the algorithm should be able to easily interpret data's features.

Step 3: Feature Selection

Feature selection is a process that chooses a subset of features from the original features so that the feature space is optimally reduced according to a certain criterion. It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve. We do this by including or excluding important features without changing them. It helps in cutting down the noise in our data and reducing the size of our input data.

Step 4: Machine Learning Algorithm

ML is a branch of artificial intelligence that enables computers to think like human beings and make their own decisions without human interference. ML has much progress in detecting various forms of disease due to the rapid growth

of Artificial Intelligent. A machine learning algorithm also provides us with more precise predictions and performance.

III.RESULT

Statistical models for estimation that are not capable to produce good performance results have flooded the assessment area. Statistical models are unsuccessful to hold categorical data, deal with missing values and large data points. All these reasons arise the importance of MLT. ML plays a vital role in many applications, example image detection, data mining, natural language processing, and disease diagnostics. In all these domains, ML offers possible solutions. This paper provides the survey of different machine learning techniques for diagnosis of different diseases such as heart disease, diabetes disease, liver disease, dengue and hepatitis disease. Many algorithms have shown good results because they identify the attribute accurately. From previous study, it is observed that for the detection of heart disease, SVM provides improved accuracy of 94.60%. Diabetes disease is accurately diagnosed by Naive Bayes. Survey highlights the advantages and disadvantages of these algorithms. Improvement graphs of machine learning algorithms for prediction of diseases are presented in detail. From analysis, it can be clearly observed that these algorithms provide enhanced accuracy on different diseases. These tools are very useful for the analysis of such problems and also provide opportunity for the improved decision-making process.

The datasets used for the experiment are collected from the Kaggle.com .

Dataset Used

1. Heart Disease Dataset

For predicting the occurrence of heart disease we have used the “Heart Disease Dataset” by K. This dataset consists of 13 medical predictor features and one target feature. The attributes are as follows chol, cp, trestbps, age, fbs, sex, restecg, exang, slope, thal, ca, old peak, thalach

2. Diabetes Disease Dataset

For predicting the occurrence of Diabetics disease we have used the “Pima Indians Diabetes Dataset” by Kaggle. This dataset contains 8 medical predictor features and one target feature. The attributes are as follows Blood pressure, Pregnancies, Glucose, Skin Thickness, BMI, Insulin, Age, and Diabetes Predigree.

3. Parkinson’s Disease Dataset

For predicting the occurrence of Parkinson’s disease we have used the “Parkinson’s Disease (Diagnostic) Data Set” by Kaggle. This dataset contains 31 medical predictor features and one target feature. Some of the important attributes are as follows diagnosis, id, radius-mean, texture-mean, perimeter-mean, area-mean, smoothness-mean, shimmer, MDV, HNR, NHR, spread1, spread2, MDVP(%), DFA, PPE, D2, RPDE.

4. Liver Disease

For predicting the occurrence of Liver disease we have used the “Indians Liver Disease Record” by Kaggle. This dataset contains 8 medical predictor features and one target feature.

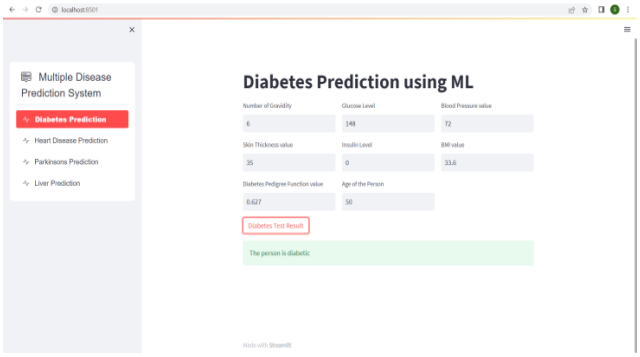


Figure 2 Results predicting a person is diabetic

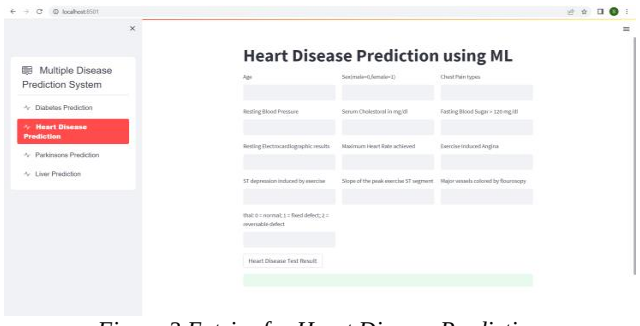


Figure 3 Entries for Heart Disease Prediction

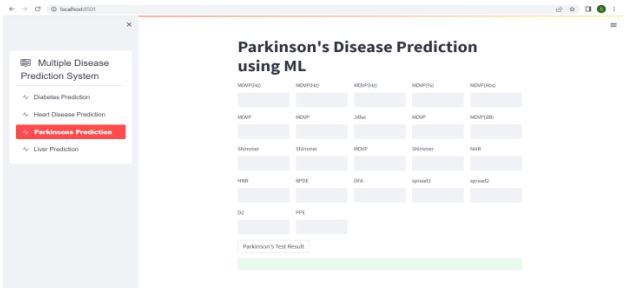


Figure 4 Entries for Parkinson's Disease Prediction

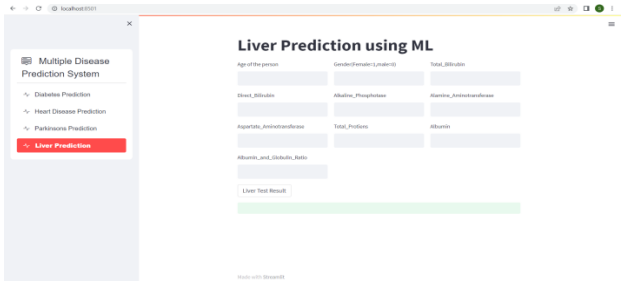


Figure 5 Entries for Liver Disease Prediction

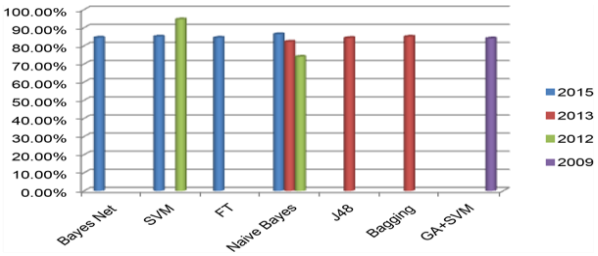
IV.DISCUSSION

Heart Disease

In existing literature, SVM offers highest accuracy of 94.60% in 2012. In many application areas, SVM shows good performance result. Attribute or features used by Parthiban and Srivatsa in 2012 are correctly responded by SVM. In 2015, Otoom et al. used SVM variant called SMO. It also uses FS technique to find best features. SVM responds to these features and offers the accuracy of 85.1% but it is comparatively low as in 2012. Training and testing set of both data sets are different, as well as, data types are different.

Table no 1: Comprehensive view of machine learning techniques for heart disease diagnosis.

Machine Learning Techniques	Author	Year	Disease	Resources of Data Set	Tool	Accuracy
Bayes Net	Otoom et al.	2015	CAD(Coronary artery disease)	UCI	WEKA	84.5%
SVM						85.1%
FT						84.5%
Naive Bayes	Vembandasamy et al.	2015	Heart Disease	Diabetic Research Institute in Chennai	WEKA	86.419%
Naive Bayes	Chaurasia and Pal	2013	Heart Disease	UCI	WEKA	82.31%
J48						84.35%
Bagging						85.03%
SVM	Parthiban and Srivatsa	2012	Heart disease	Research institute in Chennai	WEKA	94.60%
Naive Bayes						74%
Hybrid Technique (GA + SVM)	Tan et al.	2009	Heart disease	UCI	LIBSVM and WEKA	84.07%



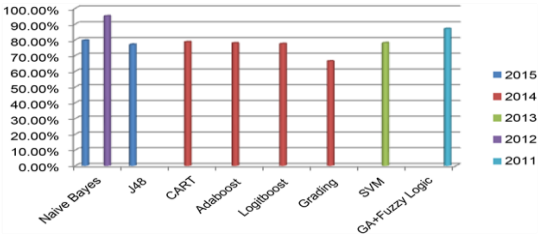
Diabetes Disease

Naive Bayes based system is helpful for diagnosis of Diabetes disease. Naive Bayes offers highest accuracy of 95%

in 2012. The results show that this system can do good prediction with minimum error and also this technique is important to diagnose diabetes disease. But in 2015, accuracy offered by Naive Bayes is low. It presents 79.5652% or 79.57% accuracy. This proposed model for detection of Diabetes disease would require more training data for creation and testing. Figure shows the Accuracy graph of Algorithms for the diagnosis of Diabetes disease according to time.

Table no 2: Comprehensive view of machine learning techniques for diabetes disease diagnosis.

Machine Learning Techniques	Author	Year	Disease	Resource of Data Set	Tool	Accuracy
Naive Bayes	Iyer et al.	2015	Diabetes Disease	Pima Indian Diabetes dataset	WEKA	79.5652%
J48						76.9565%
CART	Sen and Dash	2014	Diabetes Disease	Pima Indian Diabetes dataset from UCI	WEKA	78.646%
Adaboost						77.864%
Logiboost						77.479%
Grading						66.406%
SVM	Kumari and Chitra	2013	Diabetes Disease	UCI	MATLAB 2010a	78%
Naive Bayes	Sarwar and Sharma	2012	Diabetes type-2	Different Sectors of Society in India	MATLAB with SQL Server	95%
GA + Fuzzy Logic	Ephzibah	2011	Diabetes disease	UCI	MATLAB	87%

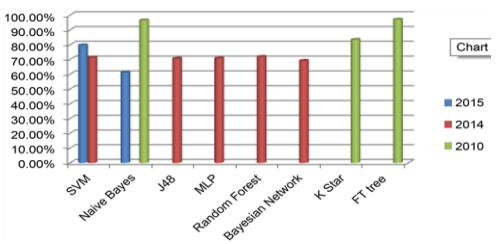


Liver Disease

To diagnose liver disease, FT Tree Algorithm provides the highest result as compare to the other algorithms. When FT tree algorithm is applied on the dataset of liver disease, time taken for result or building the model is fast as compared to other algorithms. According to its attribute, it shows the improved performance. This algorithm fully classified the attributes and offers 97.10% correctness. From the results, this Algorithm plays an important role in determining enhanced classification accuracy of data set. Accuracy graph of algorithms are shown below

Table no 3: Comprehensive view of machine learning techniques for liver disease diagnosis

Machine Learning Techniques	Author	Year	Disease	Resource of Data Set	Tool	Accuracy
SVM	Vijayarani and Dhayanand	2015	Liver Disease	ILPD from UCI	MATLAB	79.66%
Naive Bayes						61.28%
J48	Gulia et al.	2014	Liver Disease	UCI	WEKA	70.669%
MLP						70.8405%
Random Forest						71.8696%
SVM						71.3551%
Bayesian Network						69.1252%
Naive Bayes	Rajeswari and Reena	2010	Liver Disease	UCI	WEKA	96.52%
K Star						83.47%
FT tree						97.10%



V.CONCLUSION

Intelligent data processing is a social necessity for identifying, as soon as possible, of useful and robust disease detections to provide patients with appropriate care within the shortest possible time. This detection has been carried out in recent decades by detecting exciting patterns in databases. Smart data processing is emerging as a requirement for effective and robust diseases to be found by society. Detection of patients providing the necessary treatment as soon as possible within the shortest possible period. This identification has been achieved in recent decades through the method of identifying exciting patterns in databases. A comprehensive overview of intelligent data analysis tools in the medical sector is given in this paper. Some examples of some algorithms used in these medical field areas are also presented, examining potential patterns based on the target searched, the methodology used, and the application field. Given the pace at which new works emerge in this emerging field, a systematic analysis such as the one we have just presented may become obsolete in a short period.

The aim of this project is to predict disease based on symptoms. The project is set up in such a way that the device takes the user's symptoms as input and generates an output, which is disease prediction. A prediction accuracy probability of 95% is obtained on average. The Stream lit is used to successfully incorporate Disease Predictor.

References

[1]. Shraddha Subhash Shirsath, Prof. Shubhangi Patil Disease Prediction Using Machine Learn.Over Big Data". International Journal of Innovative Research in Science, Engineering and Technology, [2018]. ISSN (Online):2319-8753, ISSN (Print): 2347-6710.

[2]. Vinitha S, Sweetlin S, Vinusha H, Sajini S. "Disease Prediction Using Machine Learning Over Big Data". Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, [2018]. DOI:10.5121/cseij.2018.8101.

[3]. Sayali Ambekar and Dr.Rashmi Phalnikar. "Disease Prediction by using Machine Learning". International journal of computer engineering and applications, Volume XII, special issue, May 18. ISSN: 2321-3469.

[4]. Lohith S Y, Dr. Mohamed Rafi. "Prediction of Disease Using Learning over Big Data - Survey". International Journal on Future Revolution in Computer Science & Communication Engineering. ISSN: 2454-4248.

[5]. J. Senthil Kumar, S. Appavu. "The Personalized Disease Prediction Care from Harm using Big Data Analytics in Healthcare". Indian Journal of Science and Technology, vol 9(8), DOI:10.17485/ijst/2016/v9i8/87846, [2016]. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.