

# Flask Based Multiple Disease Predictor Using ML and DL

**Ritik Chaudhary**

Dept .of Information Technology  
Noida Institute of Engineering and  
Technology  
Greater Noida, India  
ritikchaudhary36@gmail.com

**Rohit Kumar**

Dept .of Information Technology  
Noida Institute of Engineering and  
Technology  
Greater Noida, India  
rohitchaudhary1107@gmail.com

**Yash Yadav**

Dept .of Information Technology  
Noida Institute of Engineering and  
Technology  
Greater Noida, India  
yashyada2017@gmail.com

**Mr. Abdul Khalid**

Dept .of Information Technology  
Noida Institute of Engineering and  
Technology  
Greater Noida, India  
khalid.ahmad246@gmail.com

**Pranshu Vikram Singh**

Dept .of Information Technology  
Noida Institute of Engineering and  
Technology  
Greater Noida, India  
Pranshuvikrams@gmail.com

*Abstract—Healthcare is a broad area in which computer technology continuously subsume into numerous technologies, mainly Machine Learning algorithms and hospital-generated datasets. Supervised Machine Learning algorithms are exoneration in the healthcare industry. With the help of this forecast, we will identify the illness at the premature phase and deal with the required treatment. We are testing the precision of different models using the given dataset. In our opinion, during the analysis of medical data on a larger scale, no previous work has dealt with both types of data. The purpose of this literature, the aim is to acknowledge trends among different types of supervised ML models in disease detection by examining the performance metrics. The most discussed ML algorithms were Naive Bayes (NB), Decision Trees (DT), K-Nearest Neighbor (KNN). As per records, Support Vector Machine (SVM) is the most accurate at detecting kidney dis eases.*

*Index Terms—ML, Healthcare, Decision Tree, Prediction, supervised learning*

## INTRODUCTION

Machine learning is computer programming to optimize performance using sample or historical data. Machine learning is the study of whole system of computer that learn from data and experience. Machine learning algorithms contains two section: training and testing. Predicting disease using symptoms and patient history Machine learning technology has been at work for decades. Machine learning technology provides an unfathomable foundation in the medical field to effectively solve health problems. We apply machine learning to maintain comprehensive hospital data. leading to a reference in the current text must correspond to the list of references at the end of the document. In this review, we review studies that used multiple supervised ML models for each disease identification problem. This approach provides more comprehensiveness and accuracy because evaluating the performance of a single algorithm in different research contexts causes biases that produce inaccurate results. The analysis of ML models will be performed on a number of local diseases of the heart,

kidney, and breast. To detect the disease, many methods will be evaluated such as KNN, DT, CNN and SVM. At the end of this document, the ML models that work best for each disease will be concluded.

## I. IMPLEMENTATION

### A. Population and Sample

The dataset we are referring to is Kaggle's. Kaggle is a leading platform that provides free open-source datasets for learning purposes, and these datasets are accurate for models. Each survey asking people about their experience of illness can help us predict and makes our model more accurate. Currently, we are using patient sample data set to predict disease, diabetes.

- ❖ Diabetes- For diabetes analysis, originally the Pima Indian Diabetes Dataset. Data sets obtained from a hospital in Frankfurt, Germany are used.
- ❖ Heart disease- For heart disease analysis, data sets of heart disease patients from Cleveland, Hungary and Switzerland were used.

### B. Data and Sources of Data

We make disease predictions based on user-entered symptoms. For our proposed model, we use Decision Tree Prediction, Random Forest Algorithm as well as Naïve Bayes Algorithm. We went through many steps while building our model. MEDLINE, EMBASE, CINAHL, ProQuest, Scopus, Web of Science, Cochrane Library, INSPEC and ACM Digital Library were searched.

## II. METHODOLOGY

### A. Data Collection

We receive data in the form of .xlsx files in Microsoft Excel or text files. We collect data from various sources for our model to improve the prediction and accuracy of our model. The more data, the more likely the accuracy of the model is, and the accuracy of the model depends on the quality of the data.

### B. Data Processing

The type of data used in any analytical technique is important. It is time consuming and tedious to satisfy the data quality and then build a model to solve the missing

data and external use problem. Exploratory analytics is a deeper approach to exploring the nuances of data to expand feed content.

### C. Training a model

This step involves determining the best method and submitting the data in the form of the model. The cleaned data is divided into two parts: training and testing (ratio varies according to requirements), where the first (training data) is used to build the model. The second part (test data) serves as a guide.

### D. Evaluation of the model

We need to test our model based on algorithms. Working and using algorithms and testing them to use the most efficient algorithm for our data set based on usage and requirements.

### E. Improving the performance

Use algorithms that provide the most efficient, best results, and highest accuracy for a given data set. Thus, the developed model will be sufficient and adaptable for data storage and processing.

## III. USED ALGORITHMS

### A. Decision Tree

It is a supervised algorithmic learning program mainly used to solve classification problems. It works for both categorical and nonstop dependent variables, which is surprising. We tend to divide the population into two or more homogeneous groups in this algorithmic approach. This is done by using the most important attributes/independent factors to form as many separate teams as possible. In the real world, the tree has many analogs, and it seems to have influenced a wide area of machine learning, including classification and regression. A choice tree is often used in speech analysis to visually and clearly represent choices and higher cognitive processes. As the name suggests, it uses a decision tree model. Although it is often used in data mining to create a plan to achieve a specific goal, it is also used in machine learning. We use the trained model to predict whether the scales will tip to the right, left, or balance after the decision tree classifier has finished modelling.

### B. Random Forest

Random Forest is a great way to train early in the model building process to evaluate its performance, and its simplicity makes it difficult to design a "poor" random forest. This rule is also a good choice when you need to create a model in a short time. In addition, it provides an accurate representation of the weight it places on your choice.

Random forest is quite hard to judge performance wise. Plus, they'll handle a lot of different types of features, including binary, categorical, and numeric. Random forest is a tool that is (for the most part) fast, simple, and adaptable, even with its limitations. Random forest is a synchronous learning method for classification, regression and other tasks that works by building multiple decision trees during training and then generating class which is the mode of categories (classification) or predict the mean

(regression) of each tree. Random call forests correct the features of call trees that are too well-suited to their training set.

### C. Naïve Bayes Algorithm

Naive Bayes algorithm is an algorithm that studies the probability of belonging to a certain group/class of an object with given characteristics. For example, if you are looking for a fruit that is orange in color, shape, and taste, it is most likely an orange. All these characteristics increase the probability that this fruit is an orange, which is why it is called "simple". The Bayesian section refers to the statistician and philosopher Thomas Bayes and the Bayes theorem named after him, which serves as the basis of the New Bayes algorithm.

## IV. EXISTING THEORETICAL

Existing Theoretical Many of present evaluation worried analysing precise sickness. When a person desires to examine diabetes wishes to apply one evaluation and equal person desires to examine coronary heart sickness then person has to apply one greater version. This is a time taking process. And additionally, if any person having multiple sickness however in present machine if it could be expecting simplest one sickness then there may be a danger of mortality charge growth because of now no longer capable of are expecting the alternative sickness in advance.

## V. METHODOLOGICAL CONTRIBUTIONS TO THE PROPOSED WORK

Methodological contributions to the proposed work in multi sickness version prediction, it's far viable to are expecting multiple sickness at a time. So, person no want to traverse many fashions to are expecting the sicknesses. It will lessen time and additionally because of predicting a couple of sicknesses at a time there may be a danger of decreasing mortality charge.

## VI. RESULTS

This section will show the User Interface of Web application along with diseases it can predict faster and with better accuracy.



Fig. 1. Home page of Web Application

Fig.1 shows the basic details to the user about the Web Application with a simple User Interface. We can also see the model accuracies at the home page so that use can trust the result of our model.

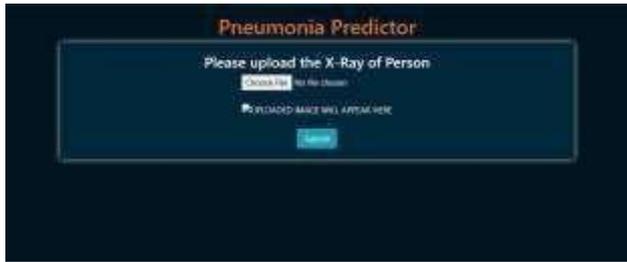


Fig. 1. Input screen of Pneumonia Predictor

Fig.2 shows the screen where user is going to give the input in form of images and text depending upon the disease they want to predict.



Fig. 3. Prediction made by the model.

Fig.3 shows the output the user has the disease or not.

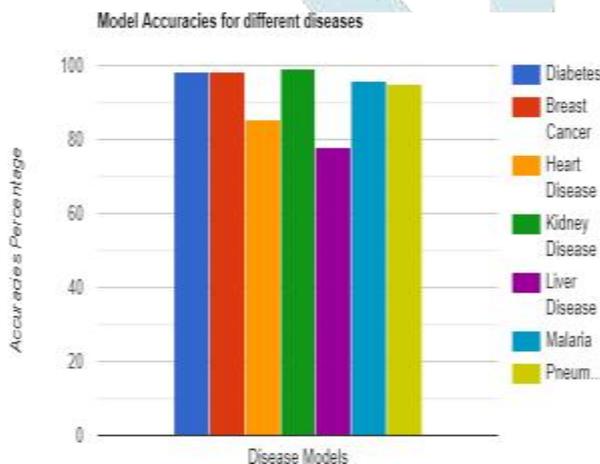


Fig. 4. Accuracy Graph

Fig.4 shows the accuracy graph of our trained models using Machine learning.

## VII. CONCLUSION

Multi Disease Prediction Model is used to predict multiple diseases at a time. Here based on the user input disease will be predicted. The choice will be given to user. If the user wants to predict particular disease or if the user doesn't enter any disease type, then based on user entered Inputs corresponding disease model will be invoked and predicted. The advantage of Multi Disease Prediction

Model in advance can prognosticate the probability of circumstance of colorful complaint and also can reduce Mortality rate.

## REFERENCES

- [1] Bracket and opinion of Diabetes norms of Medical Care in Diabetes — 2018 American Diabetes Association Diabetes Care 2018; 41( Supplement 1) S13 – S27. <https://doi.org/10.2337/dc18-S002>.
- [2] M.M.Mehdy,E.E.Shair andP.Y.Ng," Artificial Neural Networks in Image Processing for before Discovery of bone Cancer", Hindawi, Computational and Mathematical styles in Medicine, Volume 2017, Article ID 2610628.
- [3] Diabetes Prevention Program Research Group Long- term goods of life intervention or metformin on diabetes development and icrovascular complications over 15- time follow- up the Diabetes Prevention Program issuesStudy.Lancet Diabetes Endocrinol. 2015; 3 866- 875
- [4] DhafarHamed, JwanK. Alwan, Mohamed Ibrahim, MohammadB. Naeem" The Utilisation of Machine Learning Approaches for Medical Data Bracket" in Annual Conference on New Trends in Information & Dispatches Technology Applications-march2017A.J. Jenkins,M.V. Joglekar,A.A. Hardikar,A.C. Keech,D.N.O'Neal, andA.S. Januszewski, " Biomarkers in diabetic retinopathy, " The Review of Diabetic Studies,vol. 12,no. 1- 2,pp. 159 – 195, 2015.
- [5] JamilianM.etal.The influences of vitamin D and omega- 3 cosupplementation on clinical, metabolic and inheritable parameters in women with polycystic ovarysyndrome.J Affect Disord. 2018; 238 32- 38" International clinical diabetic retinopathy complaint inflexibility scale detailed table", Tech.Rep., 2002.
- [6] ChaitraliS. Dangare, SulabhaS. Apte," Improved Study of Heart Disease Prediction System using Data Mining Bracket ways", International Journal of Computer Applications( 0975 888) Volume 47No.10, June 2012.