# **Hybrid Machine Learning Approaches for Multi Disease Prediction Utilizing Relief Feature Selection**

Many existing machine learning models for health care analysis focus on a single disease. For example, one analysis could be for diabetes, another for cancer, and another for skin problems. There is no universal approach that can forecast multiple diseases using a single analysis. By utilising the Flask API, this project proposes a system that can forecast different diseases. Diabetes, heart disease, and breast cancer were all examined in this investigation. Other disorders, such as skin conditions, fever analysis, and a variety of others, can be added later. Using machine learning methods, tensor flow, and the Flask API, numerous illness analyses was implemented. The model behaviour is saved using Python pickling, and it is loaded using Python unpickling.

​​​As we are using various kinds of vast health data sets, this leads to need for a cloud for accessing the data So here we use cloud for our data access, then the remaining process get held by using the various Machine Learning Algorithms

**Existing system:**

“Disease Prediction System” used Decision tree, Random Forest and Naïve Bayes algorithms to predict a disease on the basis of systems and to enable synchronized and well-versed medical systems ensuring maximum patient satisfaction

Heart Disease Prediction with Machine Learning Approaches" made use of LR, NB,KNN,SVM,DT and RF algorithms for prediction of heart disease with proper data processing and implementation of ML algorithm with different parameters and among all Machine Learning algorithms, the highest accuracy is achieved by KNN with 87%.

”Heart Attack Prediction By Using Machine Learning Techniques" has compared various Machine Learning models with the help of performance metrics and to detect heart related problems with highest accuracy of 89.34% by SVM.

”Disease Prediction Using Machine Learning over Big Data” has proposed a CNN-MDRP algorithm which combines structured and unstructured data and proved that CNN-MDRP is more accurate than previous prediction algorithm.

**Disadvantages:**

* Existing system works for only single disease and for each type of disease use should visit multiple websites to get predict disease
* Existing methods use same algorithms for all disease predictons.

**Proposed System:**

We are proposing such a system that will flaunt a simple, cost effective , elegant User Interface and also be time efficient . Our proposed system bridges the gap between doctors and patients which will help both classes of users to achieve their goal. This system is used to predict below mentioned diseases

Diabetes

- Breast Cancer

- Heart Disease

- Kidney Disease

- Liver Disease

- Malaria

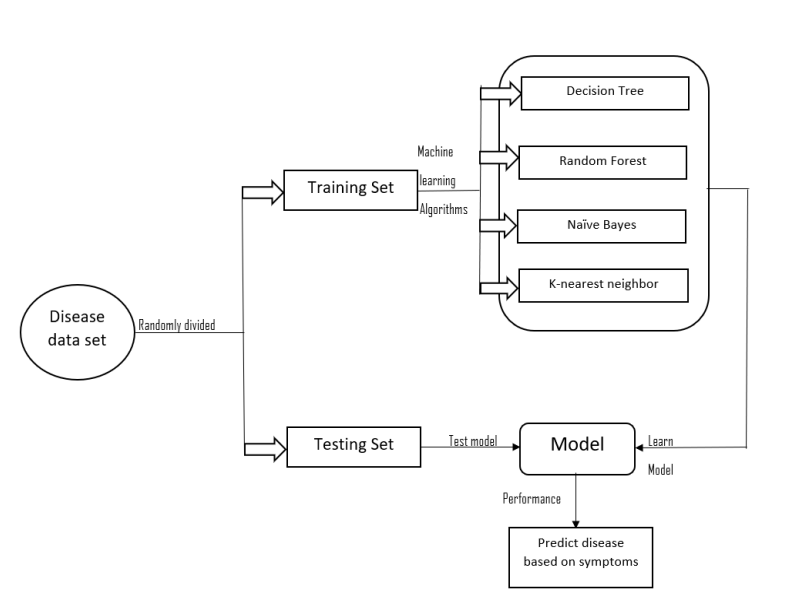
- Pneumonia

. In this proposed system we are going to take down six disease dataset from the Kaggle website and evaluate them by applying algorithms such as Decision Tree, Random Forest, Naïve bayes and KNN which will help in getting accurate prediction .Our system will explore and merge more datasets which includes large diversity of population to get more effective results and thus our system will improve and enhances the accuracy of the results. Along with the increased accuracy rate, we will proliferate the reliability of our system for this job and can gain the trust of patient in this system. Apart from all these, our system will comprise of a Database for storing the data entered by the users and the name of the disease the patient is suffering from which can be used as a reference in future for further treatment. Hence this system will contribute in easier health management with better satisfaction to the users.

**Advantages:**

* Single website is used to predict six type of diseases.
* Each disease is trained with different algorithms
* Time taken for training and prediction is less with high actuary**.**

**System Architecture:**

****

**MODULES:**

### Data COLLECTION:

Heart, kidney, diabetic, liver disease datasets are collected form Kaggle website which are in the form of csv format. These datasets have features and labels based on type of disease dataset we are using features and labels are changed.

**Understanding features of dataset.**

**Cancer Dataset:**

**Features:**

id radius\_mean texture\_mean perimeter\_mean area\_mean smoothness\_mean compactness\_mean concavity\_mean concave points\_mean symmetry\_mean fractal\_dimension\_mean radius\_se texture\_se perimeter\_se area\_se smoothness\_se compactness\_se concavity\_se concave points\_se symmetry\_se fractal\_dimension\_se radius\_worst texture\_worst perimeter\_worst area\_worst smoothness\_worst compactness\_worst concavity\_worst concave points\_worst symmetry\_worst fractal\_dimension\_worst

Labels:

Diagnosis

**Diabetic Disease:**

**Features:**

Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age Outcome

**Labels:**

Outcome

### Pre-processing:

In this stage data analysis of each dataset is performed to check relation between features and labels with graphical representation. Null values are removed from the dataset and balanced dataset is prepared for all diseases datasets.

### Split Data:

Data set is split in to two parts using test train split function ( 80 and 20 ) as test and train datasets. Train features are called as train x and labels as train y. These values are used to train algorithm and test data is used to check accuracy of each disease dataset**.**

**Apply ML algorithms**

In this stage pre processed dataset is taken as input of each disease dataset and trained features and labels are given as input to fit function to train model and model is saved in to system in the form of pkl file. The model is used in web application for prediction results based on user given input.

**Accuracy results**

After training is done test set is given and input to algorithm to test accuracy of each dataset.

**Flask Web framework:**

For this project web application is developed using flask framework which takes trained model as input and html, css for web page design. Using this application own input is given to webpage and disease is predicted.

**Software Requirement:**

### HARDWARE REQUIREMENTS:

* System : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
* Hard Disk : 1 TB.
* Input Devices : Keyboard, Mouse
* Ram : 4 GB.

### SOFTWARE REQUIREMENTS:

* Operating system : Windows XP/7/10.
* Coding Language : Python
* Tool : Anaconda
* Interface : OPENCV

**References;**

[1] Khurana, Sarthak . , Jain, Atishay ., Kataria ,Shikhar. ,Bhasin ,Kunal . , Arora ,Sunny . ,& Gupta , Dr.Akhilesh . Das. (2019). Disease Prediction System.International Research Journal Of Engineering and Technology , 6(5) , 5178-5184.

[2] Kamboj ,Mgha. (2020).Heart Disease Prediction with Machine Learning Approaches.International Journal Of Science and Research , 9(7) , 1454-1458.

[3]Ware,Miss.Sangya . , Rakesh,Mrs.Shanu. K.,&Choudhary,Mr.Bharat . (2020). Heart Attack Prediction By Using Machine Learning Techniques. International Journal Of Recent Technology and Engineering , 8(5), 1577- 1580. Turkish Journal of Computer and Mathematics Education \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 4023 Research Article Vol.12 No.6 (2021), 4013-4023

[4] Shirsath ,Shraddha.Subhash .,& Patil , Prof. Shubhangi . (2018).Disease Prediction Using Machine Learning over Big Data .International Journal Of Innovative Research in Science and Technology , 7(6), 6752-6757

[5]Marimuthu , M. , Abinaya, M. ,Hariesh,K.S., Madhan,K.,& Pavithra, Kumar. V.(2018).A Review of Heart Disease Prediction Using Machine Learning and Data Analytics Approach .International Journal of Computer Application , 181(18), 20-25.

[6] Battineni ,Gopi. , Sagaro,Getu.Gamo. ,Chinatalapudi, Nalini. ,&Amenta,Francesco. (2020). Application Of Machine Learning Predictive Models in the Chronic Disease .International of PersonalisedMedicine , 10(21), 1- 11