Disease Prediction using Machine Learning

Karthik – 20BCD7013

Ganesh Jasti – 20BCD7123

Kamalnath Reddy – 20BCD7039

Sasank – 20BCI7330

ABSTRACT

Disease Prediction system is based on predictive modelling predicts the disease of the user on the basis of the symptoms that user provides as an input to the system. The system analyses the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the
Decision Tree Algorithm. Decision Tree Algorithm calculates the probability of the disease. With big data growth in biomedical and health care communities, accurate analysis of medical data benefits early disease detection, patient care. By using linear regression and decision tree we are predicting diseases like Diabetes, Malaria, Jaundice, Dengue, and Tuberculosis.

INTRODUCTION

Machine learning is programming computers to optimize a performance using example data or past data. Machine learning is study of computer systems that learn from data and experience . Machine learning algorithm has two passes: Training, Testing .Prediction of a disease by using patient's symptoms and history machine learning technology is struggling from past decades. Machine Learning technology gives a good platform in medical field, so that a healthcare issues can be solved efficiently.

We are applying machine learning to maintained complete hospital data Machine learning technology which allows building models to get quickly analyse data and deliver results faster, with the use of machine learning technology doctors can make good decision for patient diagnoses and treatment options, which leads to improvement of patient healthcare services. Healthcare is the most prime example of how machine learning is use in medical field.

To improve the accuracy from a large data, the existing work will be done on unstructured and textual data . For prediction of diseases the existing will be done on linear KNN, Decision Tree algorithm. The order of reference in the running text should match with the list of references at the end of the paper.

OBJECTIVE

There is a need to study and make a system which will make it easy for an end users to predict the chronic diseases without visiting physician or doctor for diagnosis. To detect the Various Diseases through the examining Symptoms of patient's using different techniques of Machine Learning Models. To Handle Text data and Structured data is no Proper method. The Proposed system will consider both structure and unstructured data. The Predictions Accuracy will Increase using Machine Learning.

EXISTING SYSTEM

The system predicts the chronic diseases which is for particular region and for the particular community. The Prediction of Diseases is done only for particular diseases. In this System Big Data & CNN Algorithm is used for Diseases risk prediction. For S type data, system is using Machine Learning algorithm i.e K-nearest Neighbours, Decision Tree, Naïve Bayesian. The accuracy of the System is upto 94.8%.

Existing paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real
life hospital data collected from central China. We propose a new convolutional neural net-work based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital.

PROPOSED SYSTEM

This system is used to predict most of the chronic diseases. It accepts the structured and textual type of data as input to the machine learning model. This system is used by end users. System will predict disease on the basis of symptoms. This system uses Machine Learning Technology. For predicting diseases Decision Tree Algorithm, for clustering KNN algorithm, final output will be in the form of 0 or 1 for which Logistic tree is used.

DATASET AND MODEL DESCRIPTION

In this we describe dataset which is being used to train the machine learning model. The dataset will contain symptoms of various diseases.

DATASET OF HOSPITAL

The hospital data will be in the form of textual format or in the structural format. The dataset used in this project is real life data. The structural data contains symptoms of patients while unstructured data consist of textual format. The dataset used is contains real-life hospital data, and data stored in data center. The data provided by the hospital contains symptoms of the patients

EVALUATION METHOD

To calculate performance evaluation in experiment, first we denote TP, TN, FP and FN as true positive (the number of results correctly predicted as required), true negative (the number of results not required), false positive (the number of results incorrectly predicted as

required), false negative (the number of results incorrectly predicted as not required) respectively. We can obtain four measurements: recall, precision, accuracy and F1 measure as follows:

$$\text{accuracy -:}$$

$$\frac{TruePositive+TrueNegative}{TruePositive+TrueNegative+FalsePositive+FalseNegative}$$

$$\text{Precision} = \frac{TruePositive}{TruePositive+FalsePositive}$$

$$\text{Recall} = \frac{TruePositive}{TruePositive+FalseNegative}$$

$$\text{F1-Measure} = \frac{2\times precision\times recall}{precision+recall}$$

ALGORITHM

KNN

K Nearest Neighbour (KNN) could be a terribly easy, simple to grasp, versatile and one amongst the uppermost machine learning algorithms. In Healthcare System, user will predict the disease. In this system, user can predict whether disease will detect or not. In propose system, classifying disease in various classes that shows which disease will happen on the basis of symptoms. KNN rule used for each classification and regression issues. KNN algorithm based on feature similarity approach. A case is classed by a majority vote of its neighbours, with the case being assigned to the class most common amongst its K nearest neighbours measured by a distance function. If K = 1, then the case is just assigned to the category of its nearest neighbour

$$Euclideandistance = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

It ought to even be noted that every one 3 distance measures square measure solely valid for continuous variables. In the instance of categorical variables, the Hamming distance must be used. It conjointly brings up the difficulty of standardization of the numerical variables between zero and one once there's a combination of numerical and categorical variables within the dataset.

$$HammingDistance = \sum_{i=1}^{k} |x_i - y_i|$$

Support Vector Machine

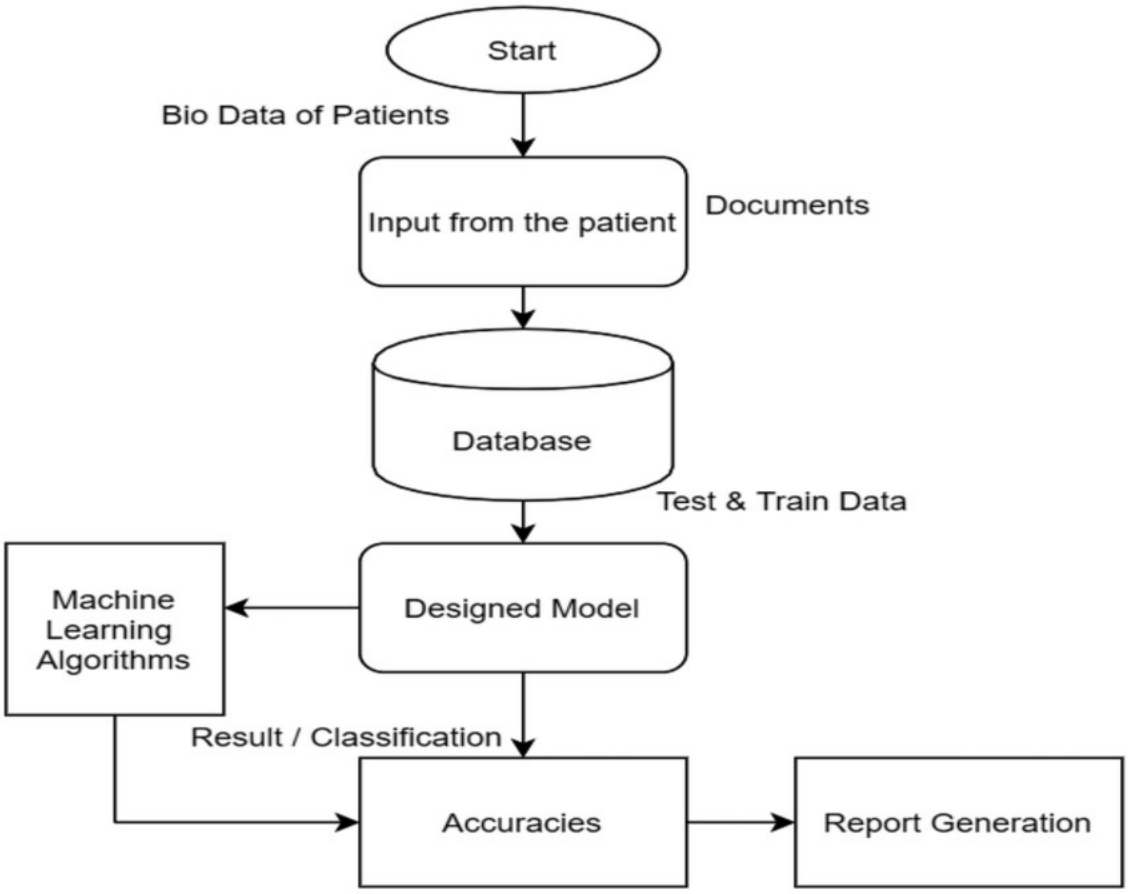
Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers' detection. It is effective when the number of attributes is greater than the number of samples but when that number increases significantly, this method is likely to give poor performance.

Decision Tree Algorithm

The methodology used in the Decision tree is a commonly used data mining method for establishing classification and prediction systems based on multiple explanatory parameters for developing prediction models for a target instance. This path classifies a population into branch-like segments in a tree that construct an inverted tree with a root node, internal nodes, and leaf nodes. A decision tree is a non-parametric algorithm which can efficiently deal with huge, complicated data sets without involving multiple parametric structures. If the sample size is large enough, study data can be divided into training and validation data sets. Using the training data set to build a decision tree model and a validation data set decide on the appropriate tree size to achieve the optimal final model.

The Decision tree works with the underlying symptoms and predicts a disease. Initially, we get the user's top five symptoms and put it in an array with the value assigned as 1 across these values. This is passed as an input to the model for predicting the disease. This array matches the disease data collection and ends at a common leaf node with the highest degree of trust.

Flow Diagram



Advantages of disease prediction using machine learning

Personalized Medicine: ML models can analyze individual patient data, including genetic information, medical history, and lifestyle factors, to develop personalized treatment plans. This can lead to more targeted and effective interventions, minimizing adverse effects and improving patient outcomes.

Improved Accuracy: ML algorithms can process vast amounts of medical data and identify complex relationships between variables. This can result in more accurate predictions and diagnoses, reducing the risk of misdiagnosis and unnecessary treatments.

Proactive Healthcare: ML-based disease prediction systems can enable proactive healthcare by identifying individuals at risk of developing certain diseases. This allows for early interventions, lifestyle modifications, and preventive measures to mitigate the risk and prevent disease progression.

Disadvantages of disease prediction using machine learning

Data Limitations: ML models rely on large, high-quality datasets to achieve accurate predictions. Limited or biased data can lead to flawed predictions and inaccurate results. Data collection and quality assurance are crucial for the success of ML-based disease prediction systems.

Ethical Concerns: ML-based disease prediction systems raise ethical considerations related to privacy, consent, and potential bias in the data. Patient data must be handled with utmost care and in compliance with privacy regulations to protect individuals' rights and maintain trust in the healthcare system.

Human Expertise: ML models should be viewed as tools to support healthcare professionals rather than replace them. The expertise and clinical judgment of healthcare providers are essential for interpreting and contextualizing ML predictions and making informed decisions about patient care.

Applications

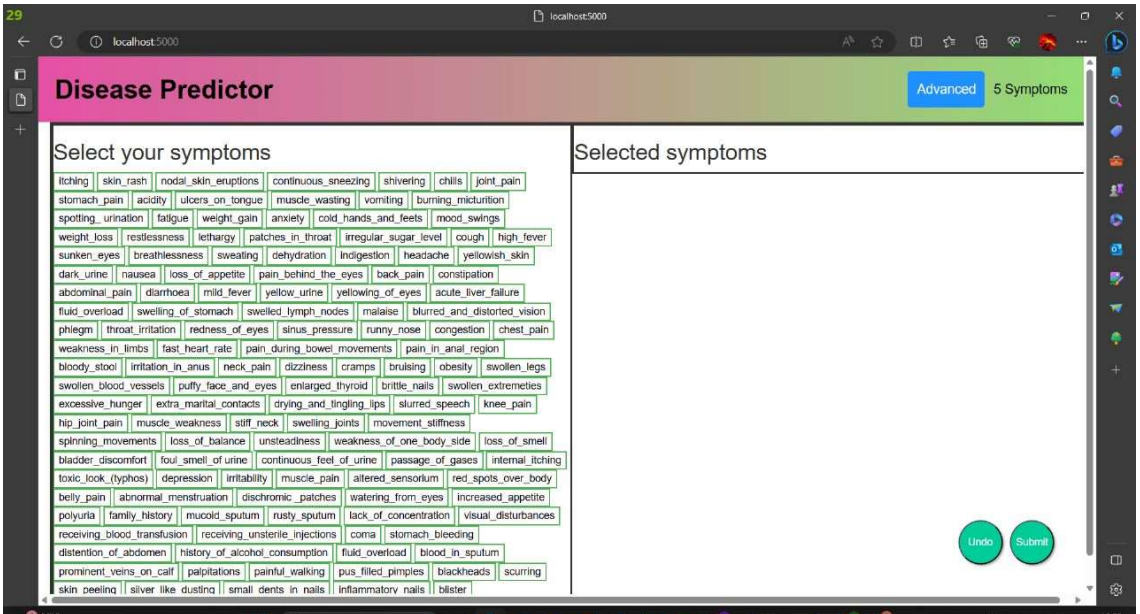
Early Detection of Diseases: ML algorithms can analyse large datasets of patient information, including medical records, lab results, and genetic data, to identify patterns and risk factors associated with various diseases. This can help in early detection of diseases such as cancer, cardiovascular conditions, diabetes, and neurological disorders, allowing for timely intervention and improved outcomes.

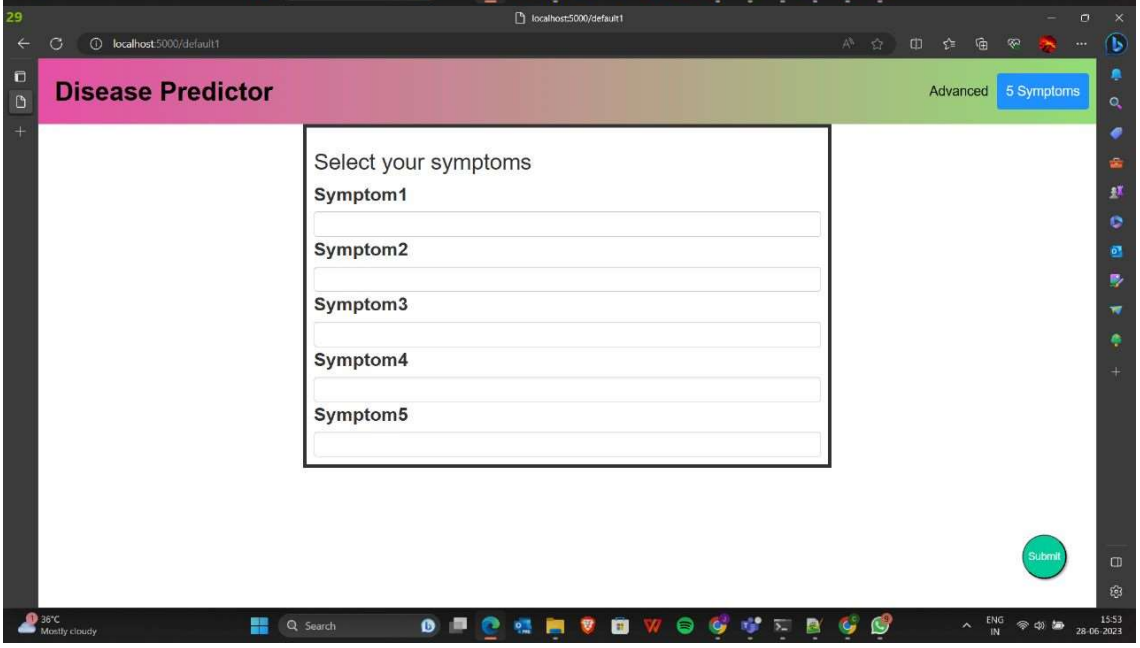
Personalized Medicine: ML models can analyse a patient's genetic makeup, medical history, lifestyle factors, and treatment outcomes to develop personalized treatment plans. By predicting how an individual is likely to respond to specific therapies, ML can help healthcare professionals optimize treatment strategies, minimize adverse effects, and improve overall patient care.

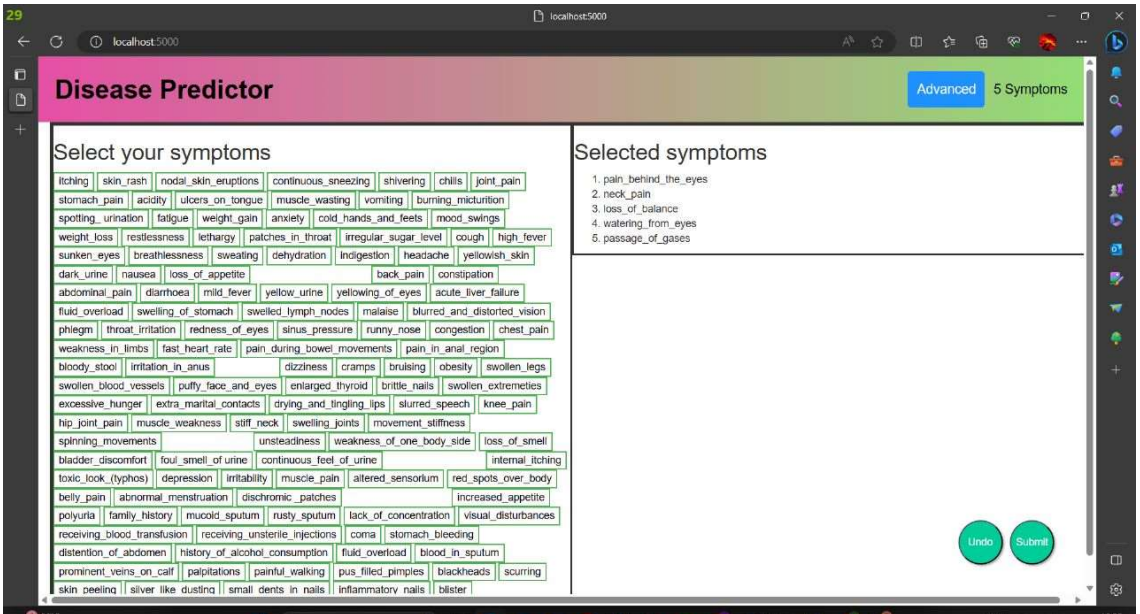
Risk Stratification: ML can be used to assess an individual's risk of developing certain diseases based on their demographic information, lifestyle choices, and medical history. By identifying high-risk individuals, healthcare providers can implement preventive measures, such as lifestyle interventions or targeted screening programs, to mitigate the risk and prevent the onset of diseases.

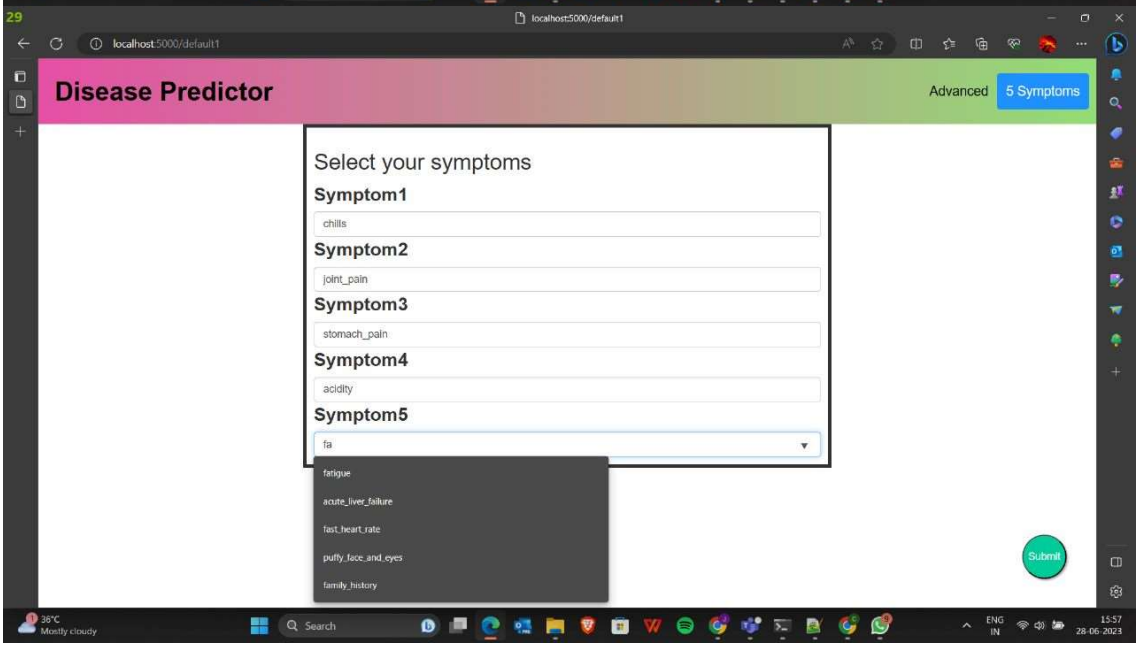
Prognostic Modelling: ML can be used to develop prognostic models that predict the future progression of diseases and estimate patient outcomes. These models can assist in treatment planning, resource allocation, and counselling patients and their families about potential outcomes.

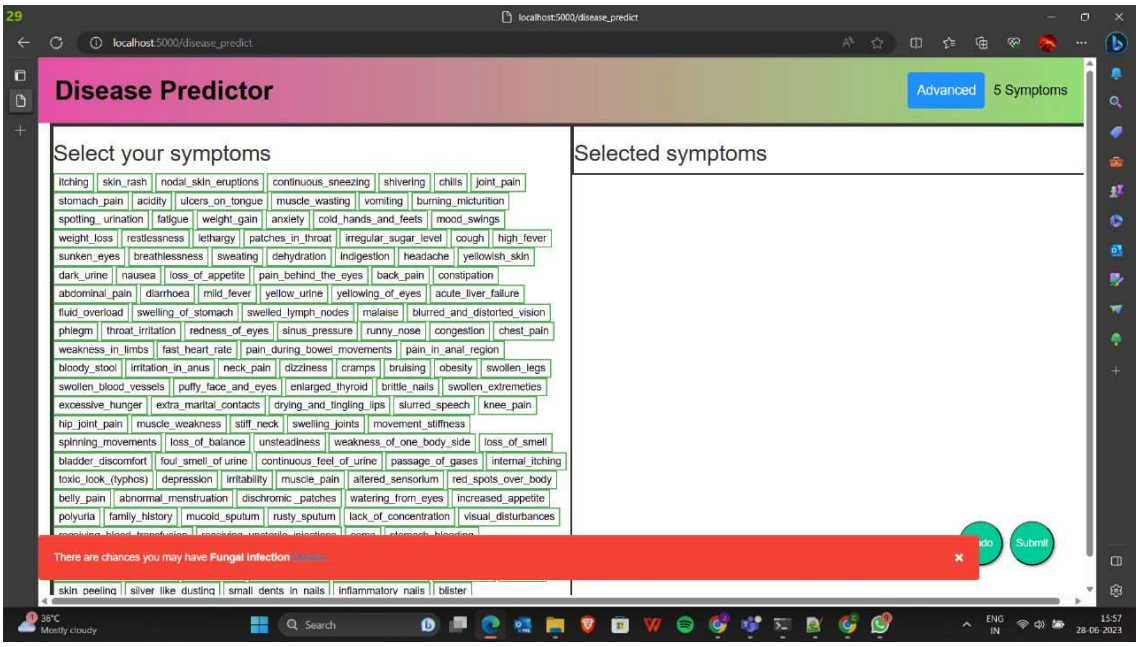
Screenshots


localhost:5000
Disease Predictor
Advanced
5 Symptoms
Select your symptoms
itching skin_rash nodal_skin_eruptions continuous_sneezing shivering chills joint_pain
stomach_pain acidity ulcers_on_tongue muscle_wasting vomiting burning_micturition
spotting_ urination fatigue weight_gain anxiety cold_hands_and_feets mood_swings
weight_loss restlessness lethargy patches_in_throat irregular_sugar_level cough high_fever
sunken_eyes breathlessness sweating dehydration indigestion headache yellowish_skin
dark_urine nausea loss_of_appetite pain_behind_the_eyes back_pain constipation
abdominal_pain diarrhoea mild_fever yellow_urine yellowing_of_eyes acute_liver_failure
fluid_overload swelling_of_stomach swelled_lymph_nodes malaise blurred_and_distorted_vision
phlegm throat_irritation redness_of_eyes sinus_pressure runny_nose congestion chest_pain
weakness_in_limbs fast_heart_rate pain_during_bowel_movements pain_in_anal_region
bloody_stool irritation_in_anus neck_pain dizziness cramps bruising obesity swollen_legs
swollen_blood_vessels puffy_face_and_eyes enlarged_thyroid brittle_nails swollen_extremeties
excessive_hunger extra_marital_contacts drying_and_tingling_lips slurred_speech knee_pain
hip_joint_pain muscle_weakness stiff_neck swelling_joints movement_stiffness
spinning_movements loss_of_balance unsteadiness weakness_of_one_body_side loss_of_smell
bladder_discomfort foul_smell_of urine continuous_feel_of_urine passage_of_gases internal_itching
toxic_look_(typhos) depression irritability muscle_pain altered_sensorium red_spots_over_body
belly_pain abnormal_menstruation dischromic _patches watering_from_eyes increased_appetite
polyuria family_history mucoid_sputum rusty_sputum lack_of_concentration visual_disturbances
receiving_blood_transfusion receiving_unsterile_injections coma stomach_bleeding
distention_of_abdomen history_of_alcohol_consumption fluid_overload blood_in_sputum
prominent_veins_on_calf palpitations painful_walking pus_filled_pimples blackheads scurring
skin_peeling silver_like_dusting small_dents_in_nails inflammatory_nails blister
Selected symptoms
Undo
Submit
Search
ENG
IN
15:53
28-06-2023


localhost:5000/default1
Disease Predictor
Advanced
5 Symptoms
Select your symptoms
Symptom1
Symptom2
Symptom3
Symptom4
Symptom5
Submit
Search
ENG
IN
15:53
28-06-2023


localhost:5000
Disease Predictor
Advanced
5 Symptoms
Select your symptoms
itching skin_rash nodal_skin_eruptions continuous_sneezing shivering chills joint_pain
stomach_pain acidity ulcers_on_tongue muscle_wasting vomiting burning_micturition
spotting_ urination fatigue weight_gain anxiety cold_hands_and_feets mood_swings
weight_loss restlessness lethargy patches_in_throat irregular_sugar_level cough high_fever
sunken_eyes breathlessness sweating dehydration indigestion headache yellowish_skin
dark_urine nausea loss_of_appetite back_pain constipation
abdominal_pain diarrhoea mild_fever yellow_urine yellowing_of_eyes acute_liver_failure
fluid_overload swelling_of_stomach swelled_lymph_nodes malaise blurred_and_distorted_vision
phlegm throat_irritation redness_of_eyes sinus_pressure runny_nose congestion chest_pain
weakness_in_limbs fast_heart_rate pain_during_bowel_movements pain_in_anal_region
bloody_stool irritation_in_anus dizziness cramps bruising obesity swollen_legs
swollen_blood_vessels puffy_face_and_eyes enlarged_thyroid brittle_nails swollen_extremeties
excessive_hunger extra_marital_contacts drying_and_tingling_lips slurred_speech knee_pain
hip_joint_pain muscle_weakness stiff_neck swelling_joints movement_stiffness
spinning_movements unsteadiness weakness_of_one_body_side loss_of_smell
bladder_discomfort foul_smell_of urine continuous_feel_of_urine internal_itching
toxic_look_(typhos) depression irritability muscle_pain altered_sensorium red_spots_over_body
belly_pain abnormal_menstruation dischromic _patches increased_appetite
polyuria family_history mucoid_sputum rusty_sputum lack_of_concentration visual_disturbances
receiving_blood_transfusion receiving_unsterile_injections coma stomach_bleeding
distention_of_abdomen history_of_alcohol_consumption fluid_overload blood_in_sputum
prominent_veins_on_calf palpitations painful_walking pus_filled_pimples blackheads scurring
skin_peeling silver_like_dusting small_dents_in_nails inflammatory_nails blister
Selected symptoms
1. pain_behind_the_eyes
2. neck_pain
3. loss_of_balance
4. watering_from_eyes
5. passage_of_gases
Undo
Submit

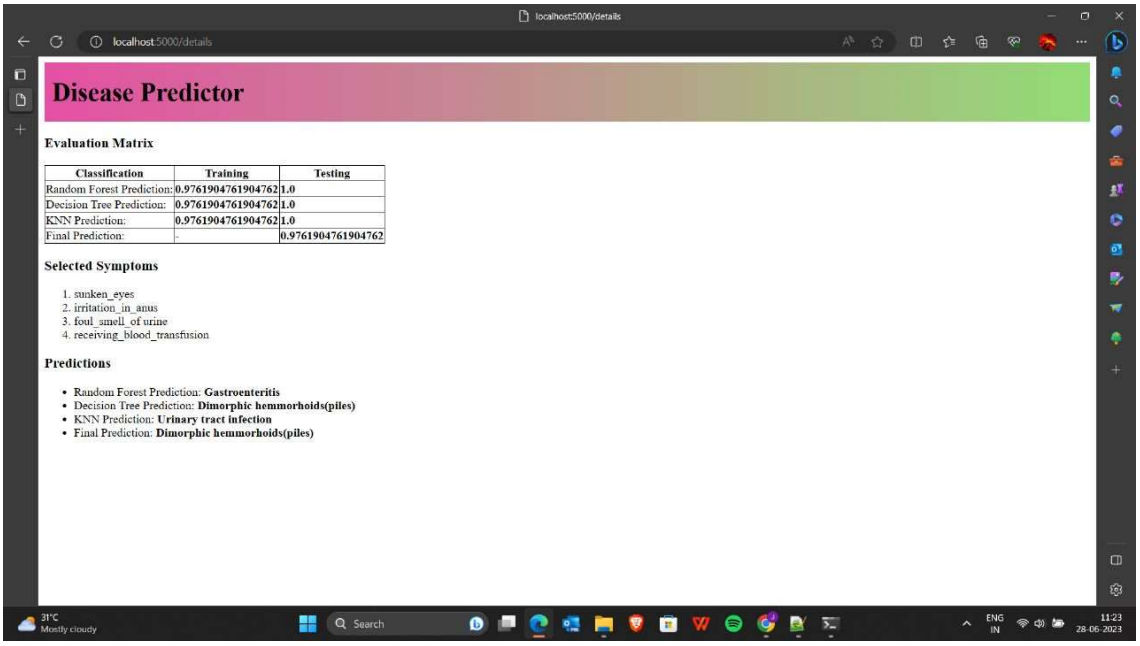

localhost:5000/default1
Disease Predictor
Advanced
5 Symptoms
Select your symptoms
Symptom1
chills
Symptom2
joint_pain
Symptom3
stomach_pain
Symptom4
acidity
Symptom5
fa
fatigue
acute_liver_failure
fast_heart_rate
puffy_face_and_eyes
family_history
Submit



RESULTS AND DISCUSSION

The Prediction Engine provides an optimal performance with the right dataset and efficient training of the classifier models considering all aspects and a lot of learning from the previous experiences. The implemented Prediction Engine is capable of predicting the presence of Diabetes with an accuracy of,

- Random Forest: 0.9761904761904762
- Decision Tree: 0.9761904761904762
- KNN: 0.9761904761904762



CONCLUSIONS

We developed a Prediction Engine which enables the user to check whether he/she has diabetes or heart disease. The user interacts with the Prediction Engine by filling a form which holds the parameter set provided as an input to the trained models.

The Prediction engine provides an optimal performance compared to other state of art approaches. The Prediction Engine makes use of three algorithms to predict the presence of a disease namely: Support Vector Machine (SVM), K-Nearest Neighbours (KNN) and Decision Tree.The reason to choose these three algorithms are:

- They are effective, if the training data is large.
- A single dataset can be provided as an input to all these 3 algorithms with minimal or no modification.
- A common scalar can be used to normalize the input provided to these 3 algorithms.

Future Scope

- To enhance the functionality of the prediction engine providing the details of 5 nearest hospitals or medical facilities to the user input location.
- Provide a user account which allows the user to keep track of their medical test data and get suggestions or support to meet the right specialists or the tests to be taken
- Provide admin controls to upload, delete the dataset which will be used to train the model.
- Automate the process of training the model and extracting pickle files of the trained models which will be consumed by the API's to predict the disease.
- Mail the detailed report of the prediction engine results along with the information of 5 nearest medical facilities details having location and contact information.