Disease Prediction using Machine Learning Algorithms

Ms S.Surekha, M Poojitha, T Sahithi, K Saitejaswini

[1]Assistant professor, Electronics and Communication Engineering, BRECW

[2,3,4]B.Tech Students, Department of Electronics and Communication Engineering, BRECW

ABSTRACT:

The development and exploitation of several prominent Data mining techniques in numerous real-world application areas (e.g. Industry, Healthcare and Bio science) has led to the utilization of such techniques in machine learning environments, in order to extract useful pieces of information of the specified data in healthcare communities, biomedical fields etc. The accurate analysis of medical database benefits in early disease prediction, patient care and community services. The techniques of machine learning have been successfully employed in assorted applications including Disease prediction. The aim of developing classifier system using machine learning algorithms is to immensely help to solve the health-related issues by assisting the physicians to predict and diagnose diseases at an early stage. A Sample data of 4920 patients' records diagnosed with 41 diseases was selected for analysis. A dependent variable was composed of 41 diseases. 95 of 132 independent variables(symptoms) closely related to diseases were selected and optimized. This research work carried out demonstrates the disease prediction system developed using Machine learning algorithms such as Decision Tree classifier, Random Forest classifier, and Naïve Bayes classifier. The paper presents the comparative study of the results of the above algorithms used.

*Keywords**:** Machine Learning, Data mining, Decision Tree classifier, Random forest classifier, Naive Bayes classifier.*

1-INTRODUCTION

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of computer system under which the Machine Learning model learn from data and experience. The machine learning algorithm has two phases: 1) Training & 2) Testing. To predict the disease from a patient's symptoms and from the history of the patient, machine learning technology is struggling from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology.

We are applying complete machine learning concepts to keep the track of patient's health. ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services.

To introduce machine learning in the medical field, healthcare is the prime example. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm. Specialists find it difficult to make decisions about the illnesses because they may not have skills in all areas. To address this issue, it is necessary to develop a disease prediction system that combines medical knowledge with an integrated system to produce the biggest results and can help society

2-LITERATURE SURVEY

[1] **M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu**, "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," J. Am Med Inform Assoc, vol. 18, no. 5, pp. 601–606, 2011.

[2] **M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang**,"Disease prediction by machine learning over big data from healthcare communities" IEEE Access, vol. 5, no.1, pp.8869– 8879, 2017.

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this 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 in 2013–2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction.

[3] **Sayali Ambekar, Rashmi Phalnika**r, "Disease RiskPrediction by Using Convolutional Neural Network" IEEE, 978-1-5386-5257-2/18, 2018.

Data analysis plays a significant role in handling a large amount of data in the healthcare. The previous medical researches based on handling and assimilate a huge amount of hospital data instead of prediction. Due to an enormous amount of data growth in the biomedical and healthcare field the accurate analysis of medical data becomes propitious for earlier detection of disease and patient care. However, the accuracy decreases when the medical data is partially missing.

[4] **Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil**, "An Improved Method for Disease Prediction using Fuzzy Approach" IEEE, DOI 10.1109/ICACCE.2015.67, pp. 569572, 2015.

Data mining is a process of extracting useful information from the huge amount of data. Data Mining has great scope in the field of medicine. This article deals with the working on PIMA and Liver-disorder datasets. Many researchers have proposed the use of K-nearest neighbor (KNN) algorithm for diabetes disease prediction. Some researchers have proposed a different approach by using K-means clustering for preprocessing and then using KNN for classification. These approaches resulted in poor classification accuracy or prediction. In our work we proposed and developed two different methods first one is fuzzy c-means clustering algorithm followed by a KNN classifier and second one is fuzzy c-means clustering algorithm followed by fuzzy KNN classifier to improve the accuracy of classification.

We are successful in obtaining the better results than the existing methods for the given datasets. Our second approach produced better result than the first one. Classification is carried out using ten folds cross-validation technique.

[5] **Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshra**m, "Designing Disease Prediction Model Using Machine Learning Approach" IEEE Xplore Part Number: CFP19K25-ART; ISBN: 9781-5386-7808-4, pp. 1211-1215, 2019.

Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor. The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Medical science has large amount of data growth per year.

[6] **Lambodar Jena and Ramakrushna Swain, "ChronicDisease** Risk Prediction using Distributed Machine Learning Classifiers" IEEE, 978- 1-5386-2924-6/17, pp. 170-173, 2017.

The prime use of the classification technique is to predict the target class accurately for each case in the dataset. The recent study is focused on the usage of classification techniques in the field of medical science and bioinformatics. The main focus of this paper is to predict Chronic- Kidney-Disease and its usage for classification in the field of medical bioinformatics. It firstly classifies dataset and then determines which algorithm performs better for diagnosis and prediction of Chronic- Kidney-Disease.

[7] **Dhomse Kanchan B. and Mahale Kishor M**., "Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis" IEEE, 978-1- 509004676/16, pp. 5-10, 2016.

The worldwide study on causes of death due to heart disease/syndrome has been observed that it is the major cause of death. If recent trends are allowed to continue, 23.6 million people will die from heart disease in coming 2030. The healthcare industry collects large amounts of heart disease data which unfortunately are not "mined" to discover hidden information for effective decision making. In this paper, study of PCA has been done which finds the minimum number of attributes required to enhance the precision of various supervised machine learning algorithms.

[8] **Ankita Dewan and Meghna Sharma**, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification" IEEE, 978-9- 3805-4416-8/15, pp. 704-706, 2015.

Heart disease prediction is treated as most complicated task in the field of medical sciences. Thus there arises a need to develop a decision support system for detecting heart disease of a patient. In this paper, we propose efficient genetic algorithm hybrid with the back propagation technique approach for heart disease prediction. Today medical field have come a long way to treat patients with various kind of diseases. Among the most threatening one is the Heart disease which cannot be observed with a naked eye and comes instantly when its limitations are reached. Bad clinical decisions would cause death of a patient which cannot be afforded by any hospital.

3-PROPOSED METHOD

In the context of Disease Prediction using Machine Learning Algorithms such as Decision Tree classifier, Random Forest classifier, and Naïve Bayes classifier., the proposed method typically involves a structured approach that covers various stages, from data preprocessing to model deployment. Here's an outline of a proposed method you can use for disease prediction:

The proposed method uses machine learning models to predict diseases based on input symptoms. The process involves data preprocessing followed by classification using multiple algorithms, with the final output being the predicted disease.

Steps:

1. **Input Symptoms**: The user inputs a set of symptoms related to a health condition.
2. **Data Preprocessing**: The raw symptom data undergoes preprocessing to ensure it is clean, standardized, and

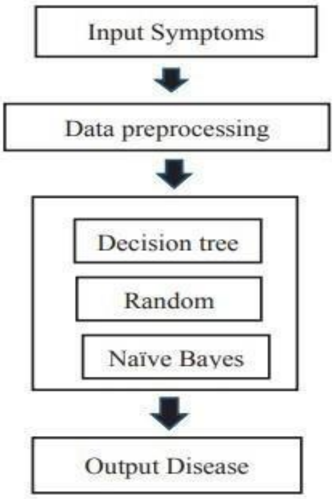
suitable for model training. This may include: o Handling missing values o Normalization or standardization of input data o Encoding categorical data if necessary

3. **Model Application:**

- **Decision Tree**: A decision tree model is applied, which splits the data based on feature values to make predictions.
- **Random (likely referring to Random Forest)**: A random forest algorithm, which is an ensemble of decision trees, is used to improve prediction accuracy. o**Naïve Bayes**: A Naïve Bayes classifier, which uses probabilistic reasoning based on Bayes' theorem, is applied for disease classification.

4. **Output Disease**: Based on the predictions from the applied models, the system provides the most probable disease as the output.

Block Diagram:



4-HARDWARE AND SOFTWARE REQUIREMENT

- Processor Type: Pentium -IV
- RAM: 512 MB RAM
- Hard disk: 20 GB

Software Requirement:

- Anaconda IDE
- Python Language

5-RESULT

We have utilized Symptoms and Diseases dataset to train different machine learning algorithms like Decision Tree, Random Forest and NB. Each algorithm performance is evaluated in terms of accuracy, precision, recall and FSCORE. After training user can input symptoms and then ML algorithm will predict Disease and recommended medicines related to predicted disease. To implement this project we have designed two modules

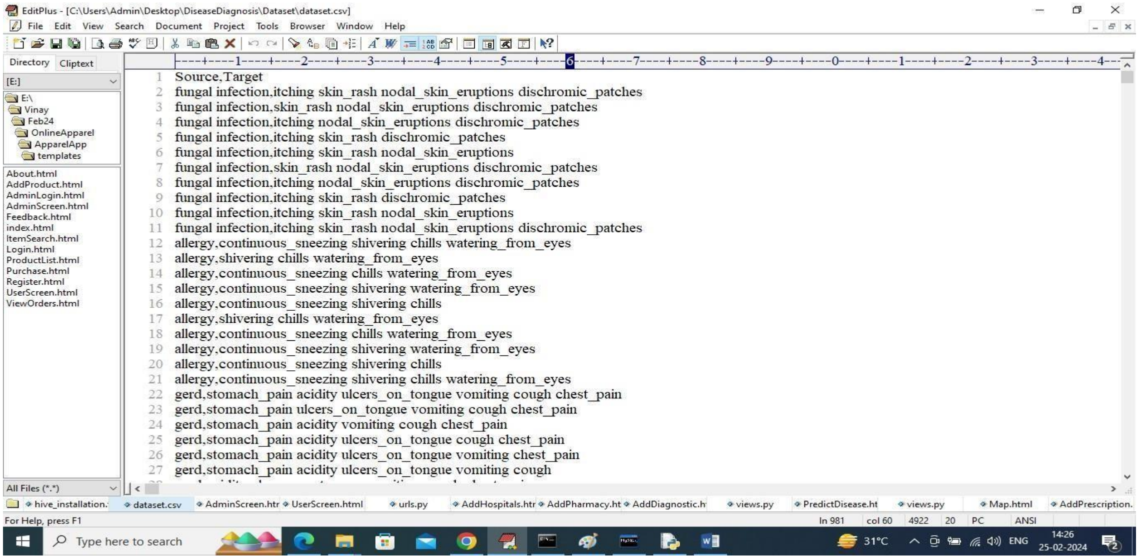
1) Admin: admin can login to system using username and password as 'admin' and then can add 'Hospitals, Pharmacy, Diagnostic centres'. Each location will be define with latitude and longitude so user can view hospital or pharmacy details on maps also. In web application there is no location available like MOBILE GPS so admin will add

latitude and longitude then user can see all hospitals with addresses and can see location in map also. Admin will add disease and suitable medicines so user can get those medicines as prescription upon disease predicted.

2) User Signup: user can sign up with the application

3) User Login: user can login to system and can view hospital, pharmacy, diagnostic details and can predict disease based on symptoms.

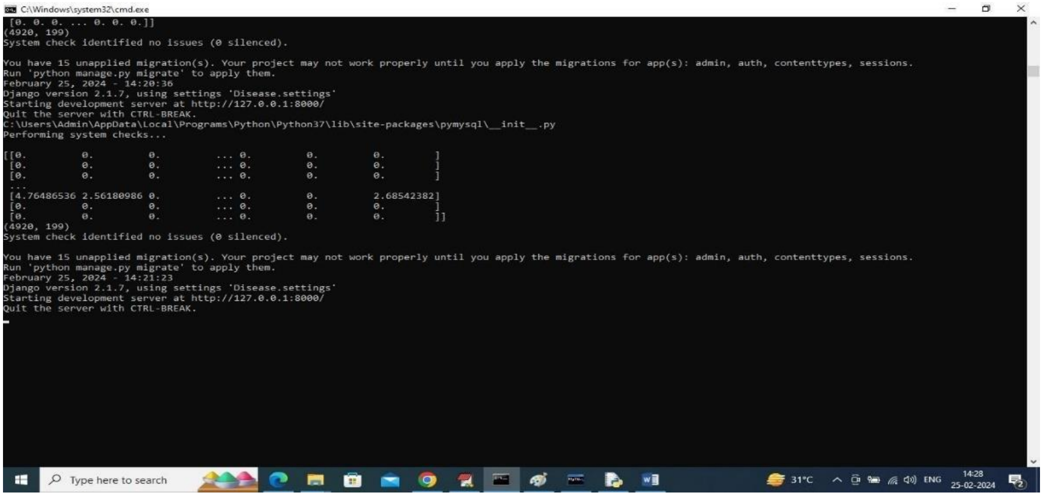
To train ML algorithm we are using below dataset



```
Source,Target
fungal infection,itching skin_rash nodal_skin_eruptions dischromic_patches
fungal infection,skin_rash nodal_skin_eruptions dischromic_patches
fungal infection,itching nodal_skin_eruptions dischromic_patches
fungal infection,itching skin_rash dischromic_patches
fungal infection,itching skin_rash nodal_skin_eruptions
fungal infection,skin_rash nodal_skin_eruptions dischromic_patches
fungal infection,itching nodal_skin_eruptions dischromic_patches
fungal infection,itching skin_rash dischromic_patches
fungal infection,itching skin_rash nodal_skin_eruptions
fungal infection,itching skin_rash nodal_skin_eruptions dischromic_patches
allergy,continuous_sneezing shivering chills watering_from_eyes
allergy,shivering chills watering_from_eyes
allergy,continuous_sneezing chills watering_from_eyes
allergy,continuous_sneezing shivering watering_from_eyes
allergy,continuous_sneezing shivering chills
allergy,shivering chills watering_from_eyes
allergy,continuous_sneezing chills watering_from_eyes
allergy,continuous_sneezing shivering watering_from_eyes
allergy,continuous_sneezing shivering chills
allergy,continuous_sneezing shivering chills watering_from_eyes
gerd,stomach_pain acidity ulcers_on_tongue vomiting cough chest_pain
gerd,stomach_pain ulcers_on_tongue vomiting cough chest_pain
gerd,stomach_pain acidity vomiting cough chest_pain
gerd,stomach_pain acidity ulcers_on_tongue cough chest_pain
gerd,stomach_pain acidity ulcers_on_tongue vomiting chest_pain
gerd,stomach_pain acidity ulcers_on_tongue vomiting cough
```

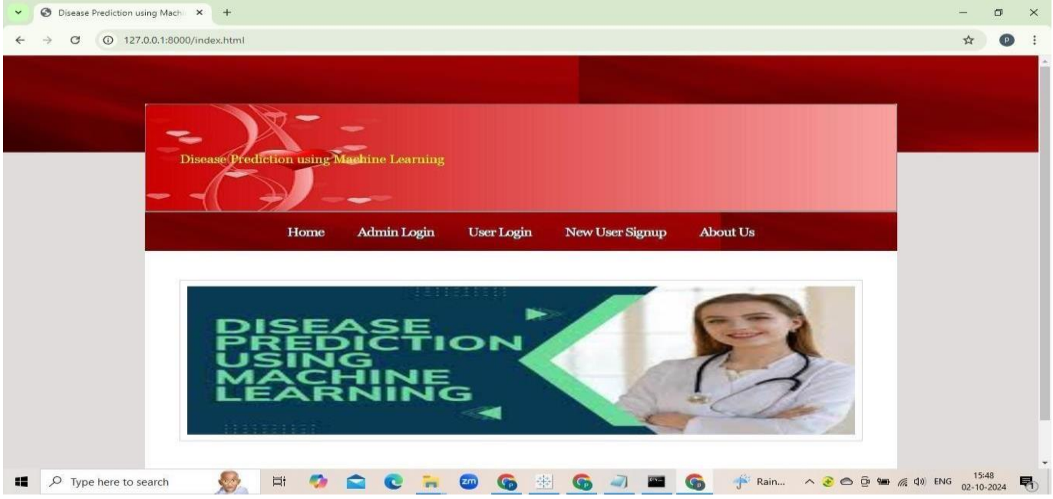
In above dataset screen source column contains Disease Name and Target column contains symptoms and then will train ML algorithm on above dataset and evaluate performance.

To run project install MYSQL and then copy content from DB.txt file and then paste in MYSQL console to create database. Install Python3.7.0 and then install packages given in requirements.txt file.

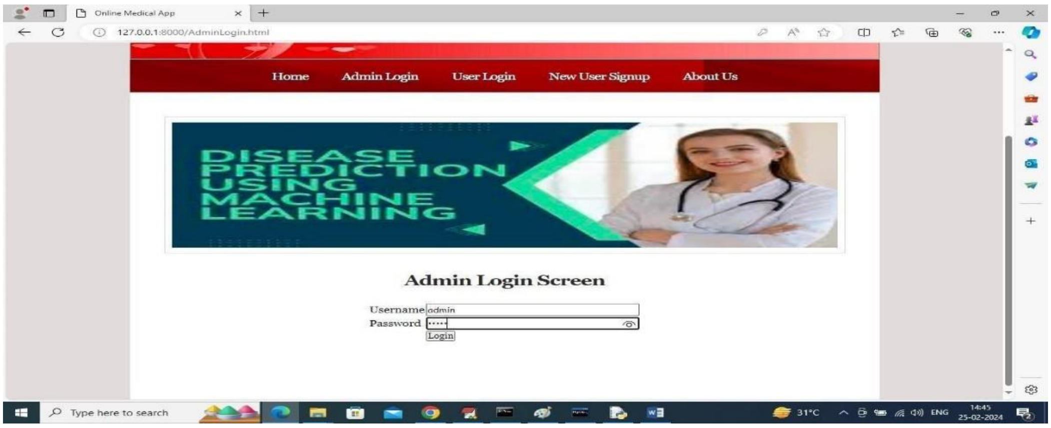


```
 [0. 0. 0. ... 0. 0. 0.]]
(4920, 199)
System check identified no issues (0 silenced).

You have 15 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.
February 25, 2024 - 14:20:36
Django version 2.1.7, using settings 'Disease.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\pymysql\__init__.py
Performing system checks...

[[0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]
 ...
 [4.76486536 2.56180986 0.         ... 0.         0.         2.68542382]
 [0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]]
(4920, 199)
System check identified no issues (0 silenced).

You have 15 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.
February 25, 2024 - 14:21:23
Django version 2.1.7, using settings 'Disease.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

In above screen python web server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and press enter key to get below page


Disease Prediction using Mach
127.0.0.1:8000/index.html
Disease Prediction using Machine Learning
Home
Admin Login
User Login
New User Signup
About Us
DISEASE PREDICTION USING MACHINE LEARNING
Type here to search
15:48
02-10-2024

In above screen click on 'Admin Login' link to get below login page.



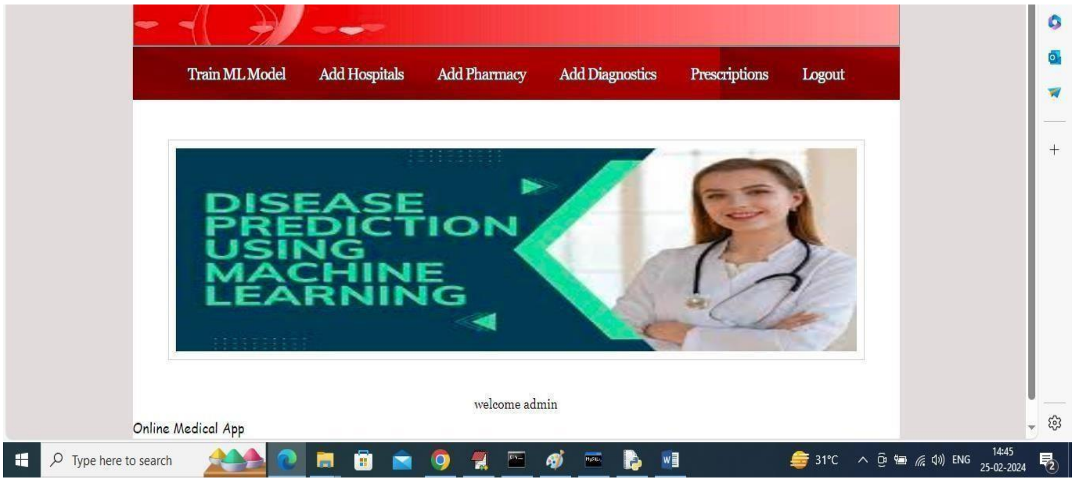
Admin Login Screen

Username admin

Password •••••

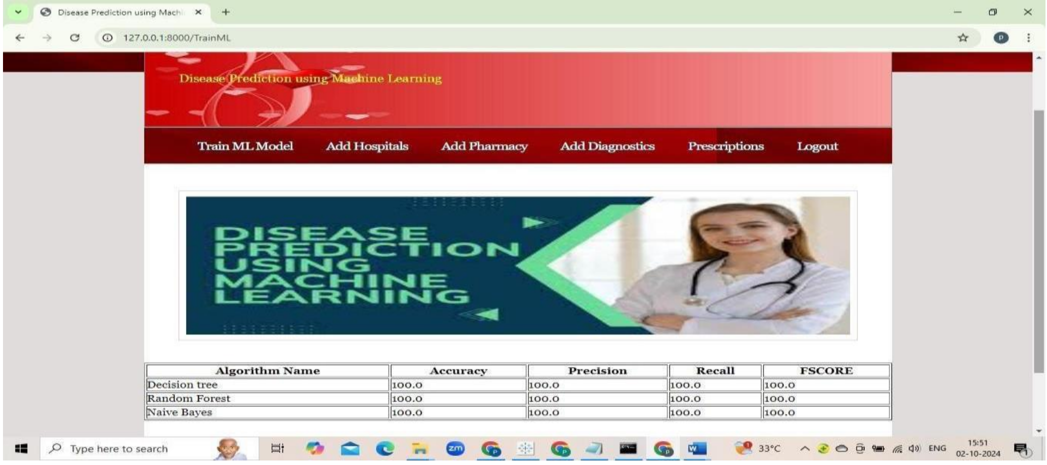
Login

In above screen admin is login and after login will get below page.



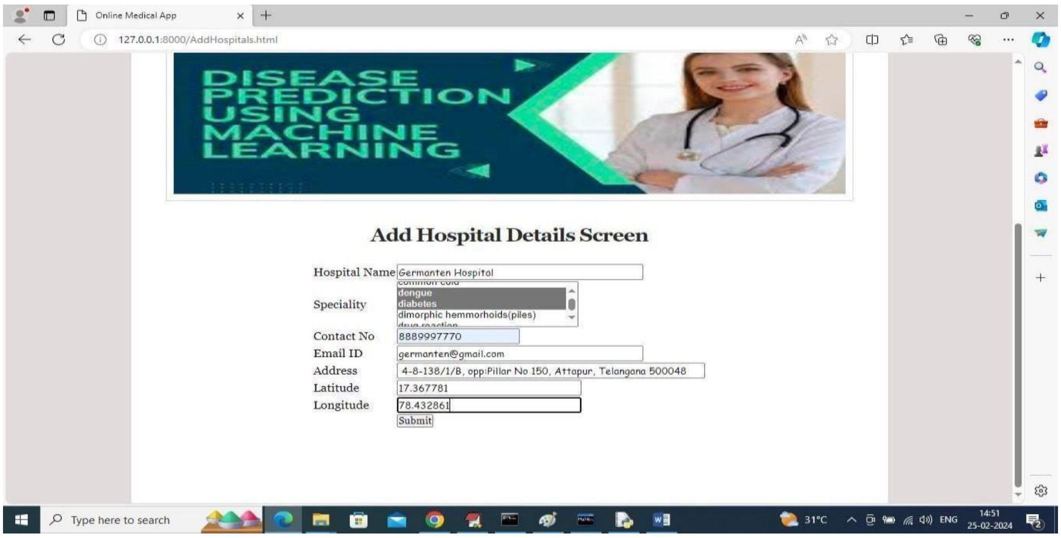
welcome admin

In above screen admin can click on 'Train ML Model' link to train ML algorithm and get below output.



Algorithm Name	Accuracy	Precision	Recall	FSCORE
Decision tree	100.0	100.0	100.0	100.0
Random Forest	100.0	100.0	100.0	100.0
Naive Bayes	100.0	100.0	100.0	100.0

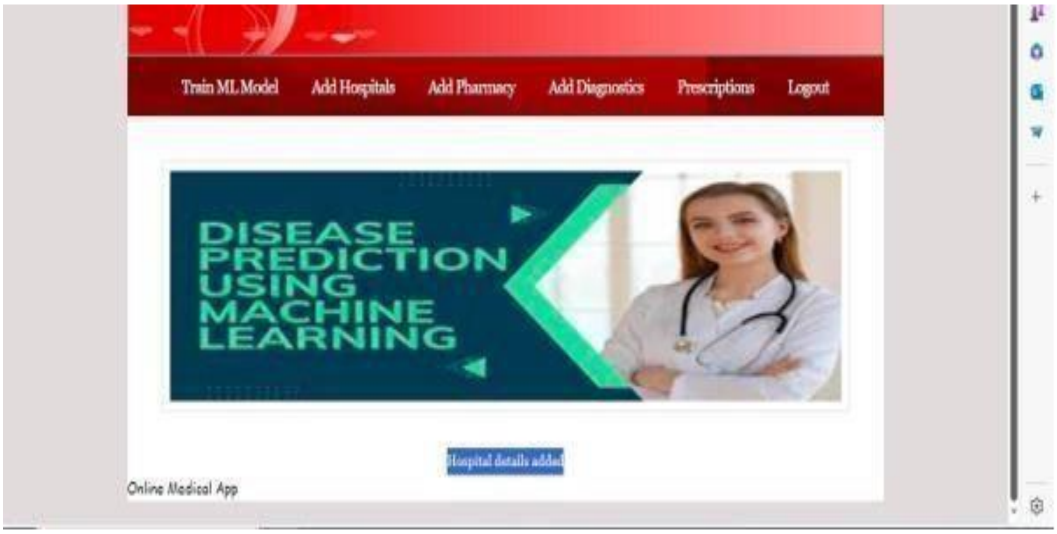
In above screen can see performance of each ML algorithm and in all algorithm Decision Tree, NB and Random Forest performing best with 100% accuracy and now click on 'Add Hospital' link to add hospital details and get below output.


Online Medical App
127.0.0.1:8000/AddHospitals.html
DISEASE PREDICTION USING MACHINE LEARNING

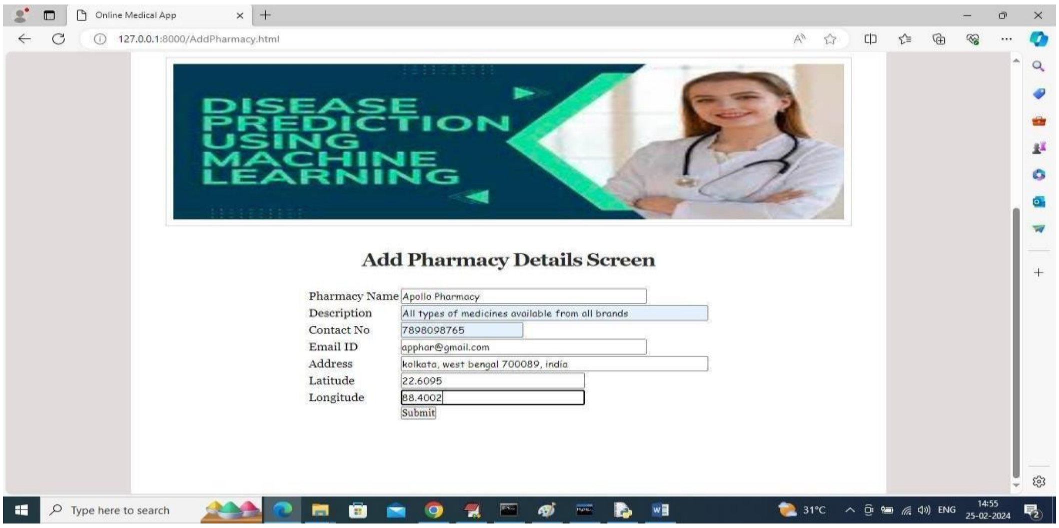
Add Hospital Details Screen

Hospital Name	Germanten Hospital
Speciality	common cold dengue diabetes dimorphic hemmorhoids(piles) drug reaction
Contact No	8889997770
Email ID	germanten@gmail.com
Address	4-8-138/1/B, opp:Pillar No 150, Attapur, Telangana 500048
Latitude	17.367781
Longitude	78.432861
	Submit

In above screen admin will add hospital details and in speciality we can select all disease names by holding CTRL key which hospital is treating and then rom Google can collect latitude and longitude and then press button to get below output.



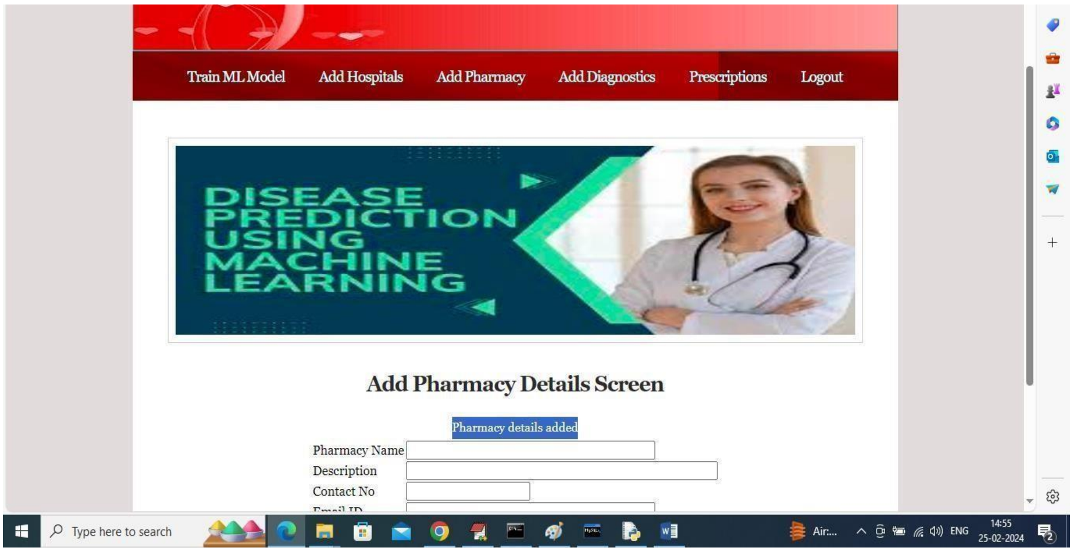
In above screen hospital details added and now click on 'Add Pharmacy' link to add pharmacy details.


Online Medical App
127.0.0.1:8000/AddPharmacy.html
DISEASE PREDICTION USING MACHINE LEARNING
Type here to search

Add Pharmacy Details Screen

Pharmacy Name Apollo Pharmacy
Description All types of medicines available from all brands
Contact No 7898098765
Email ID apphar@gmail.com
Address kolkata, west bengal 700089, india
Latitude 22.6095
Longitude 88.4002
Submit

In above screen adding pharmacy details and then press button to get below page.

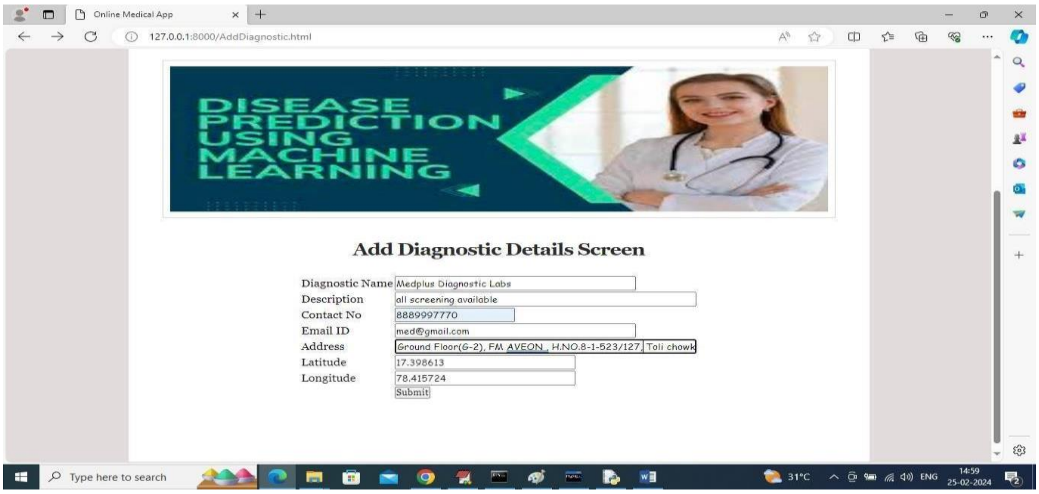


Add Pharmacy Details Screen

Pharmacy details added

Pharmacy Name
Description
Contact No
Email ID

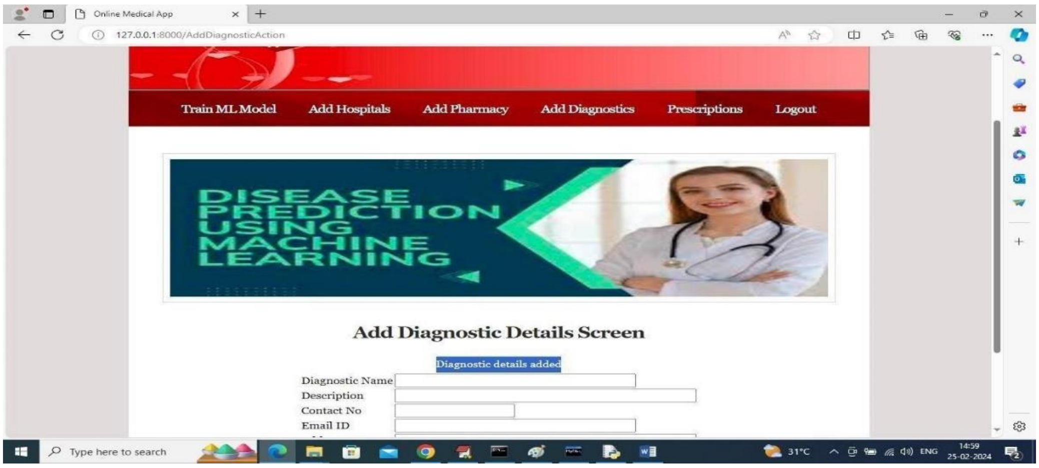
In above screen pharmacy details added and now click on 'Add Diagnostic' link to add diagnostic centres details.


Online Medical App
127.0.0.1:8000/AddDiagnostic.html
DISEASE PREDICTION USING MACHINE LEARNING
Submit

Add Diagnostic Details Screen

Diagnostic Name: Medplus Diagnostic Labs
Description: all screening available
Contact No: 8889997770
Email ID: med@gmail.com
Address: Ground Floor(G-2), FM AVEON , H.NO.8-1-523/127, Toli chowk
Latitude: 17.398613
Longitude: 78.415724

In above screen adding diagnostic centres details and then press button to get below page.

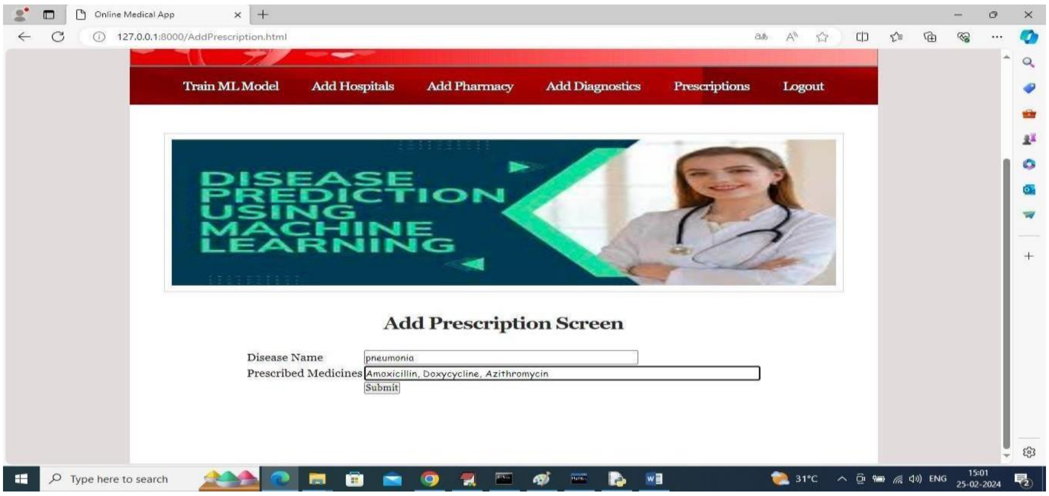


Add Diagnostic Details Screen

Diagnostic details added

Diagnostic Name
Description
Contact No
Email ID

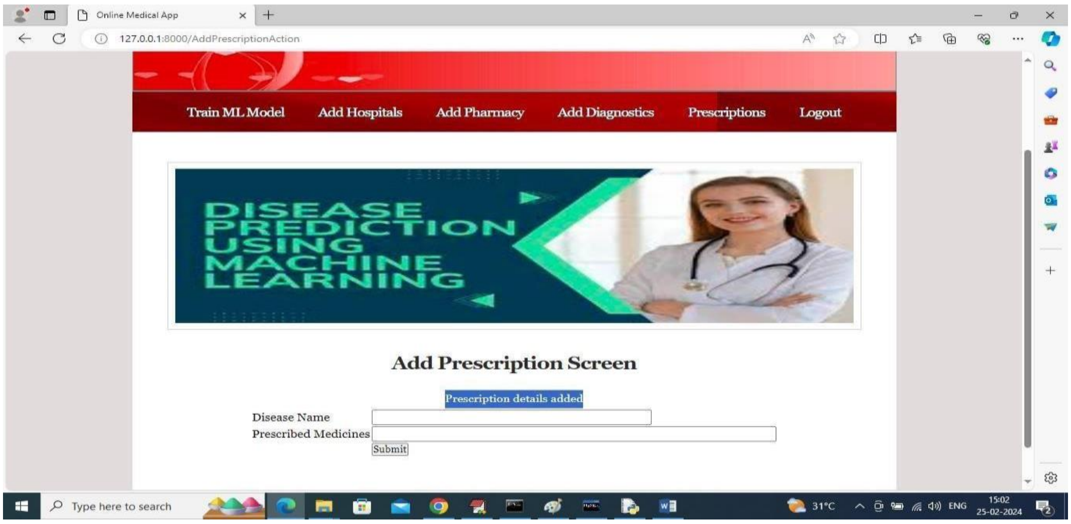
In above screen diagnostic centres details added and now click on 'Prescription' link to add disease and medicines details.



Add Prescription Screen

Disease Name: pneumonia
Prescribed Medicines: Amoxicillin, Doxycycline, Azithromycin

In above screen admin will add 'Disease name' and its medicines so application can suggest medicine upon disease predicted and now click on button to get below page.

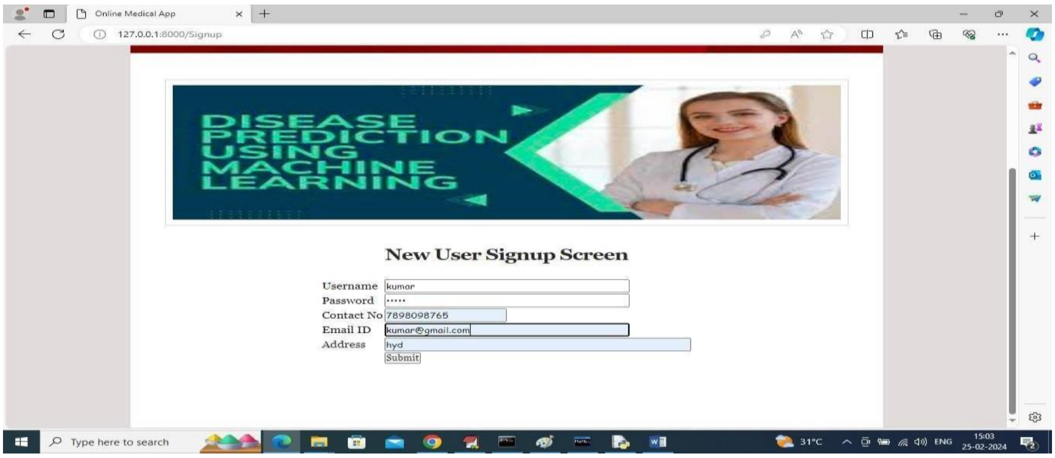


Add Prescription Screen

Prescription details added

Disease Name
Prescribed Medicines
Submit

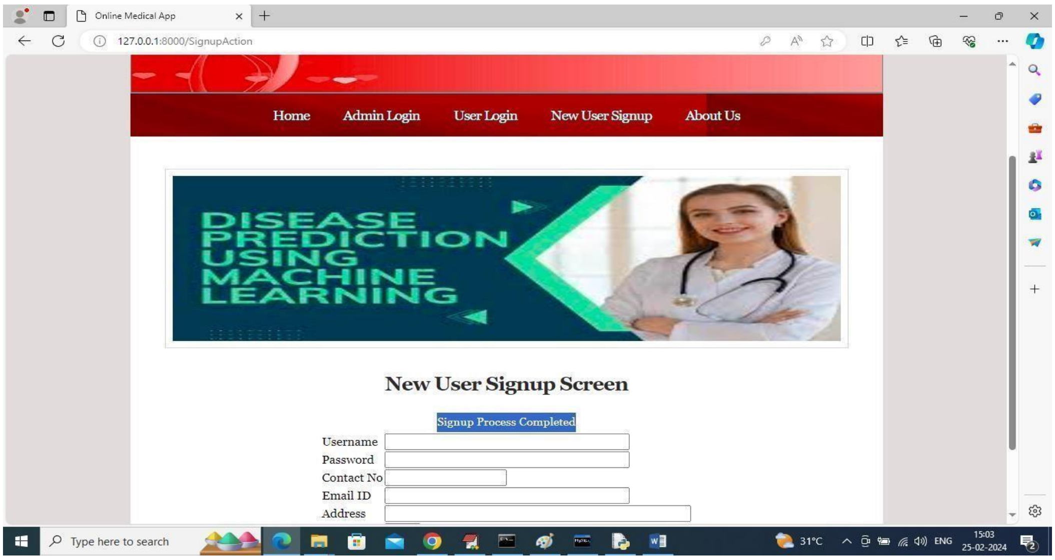
In above screen medicine details added and now logout and signup new user.



New User Signup Screen

Username kumar
Password •••••
Contact No 7898098765
Email ID kumar@gmail.com
Address hyd
Submit

In above screen user will enter sign up details and then press button to register user.

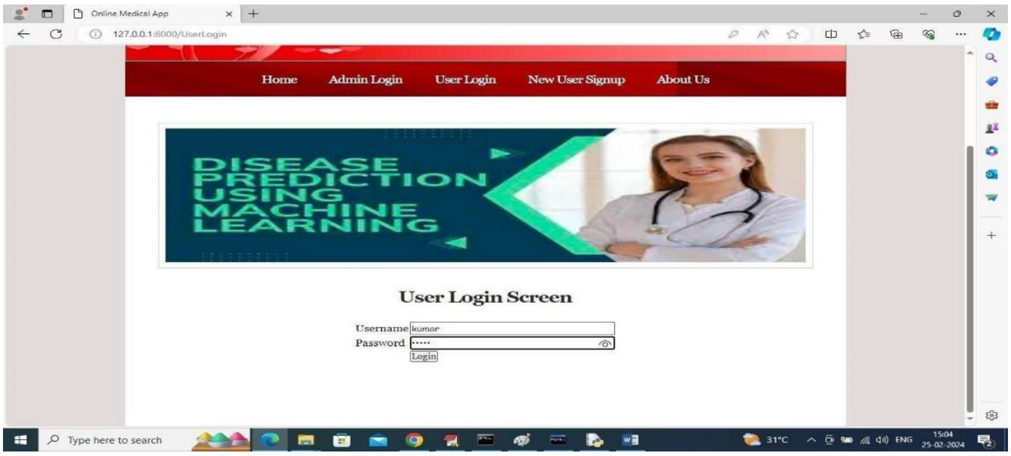


New User Signup Screen

Signup Process Completed

Username
Password
Contact No
Email ID
Address

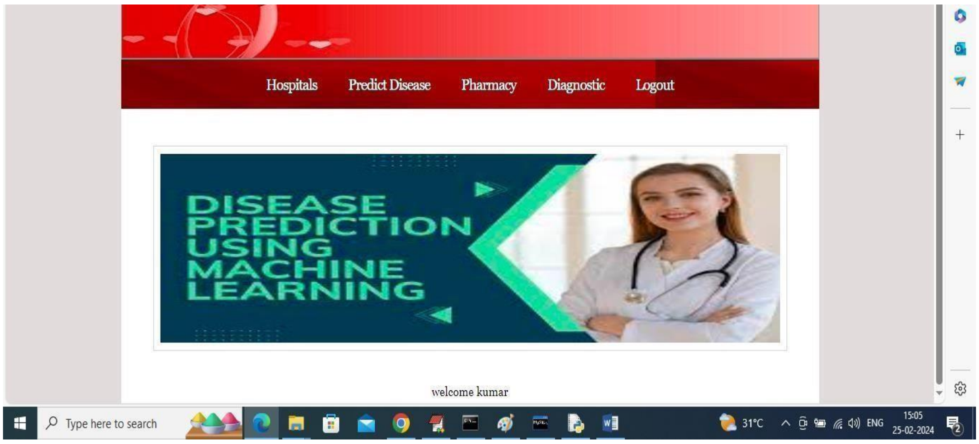
In above screen user sign up completed and now click on ‘User Login’ link to get below page.



User Login Screen

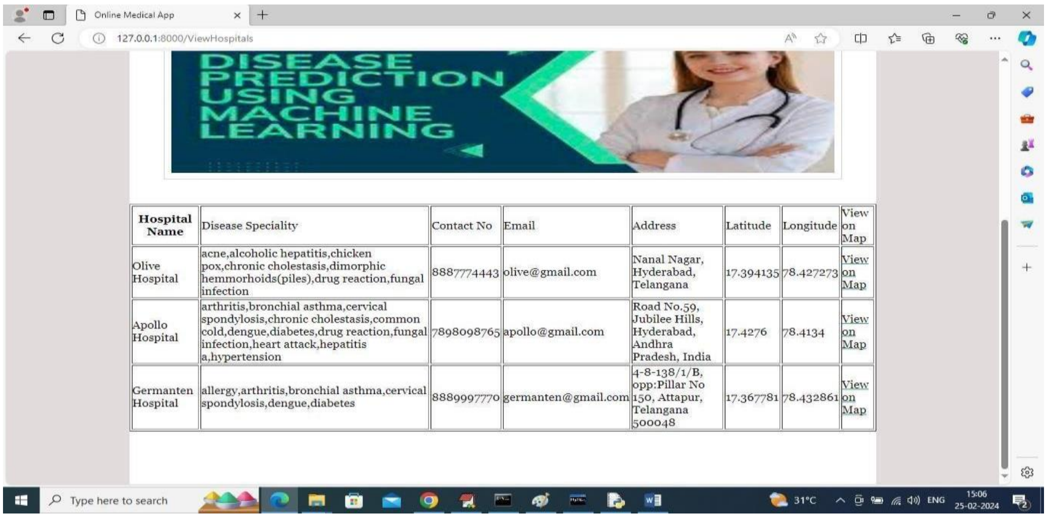
Username kumar
Password •••••

In above screen user is login and after login will get below page.



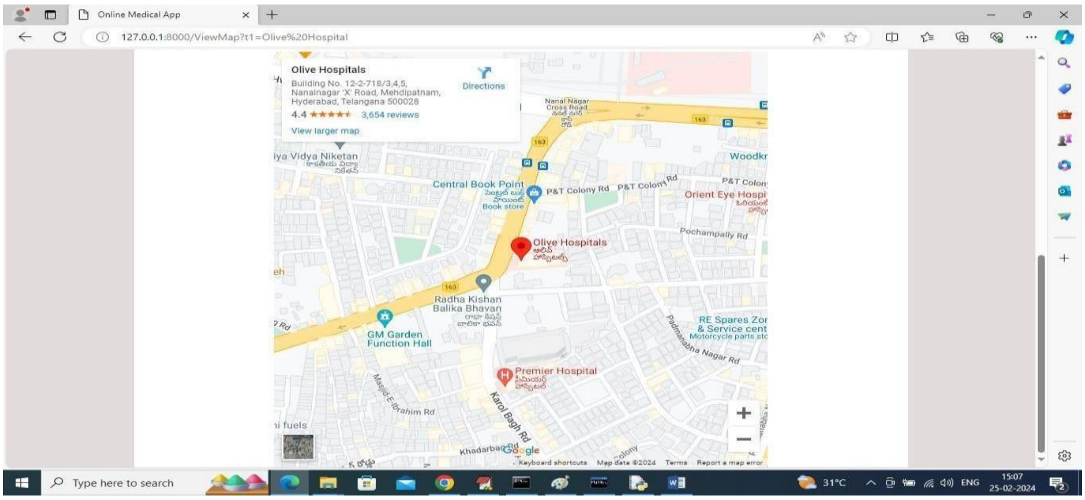
welcome kumar

In above screen user can click on ‘Hospitals’ link to get list of hospitals with addresses so he can understand his location and hospital address.

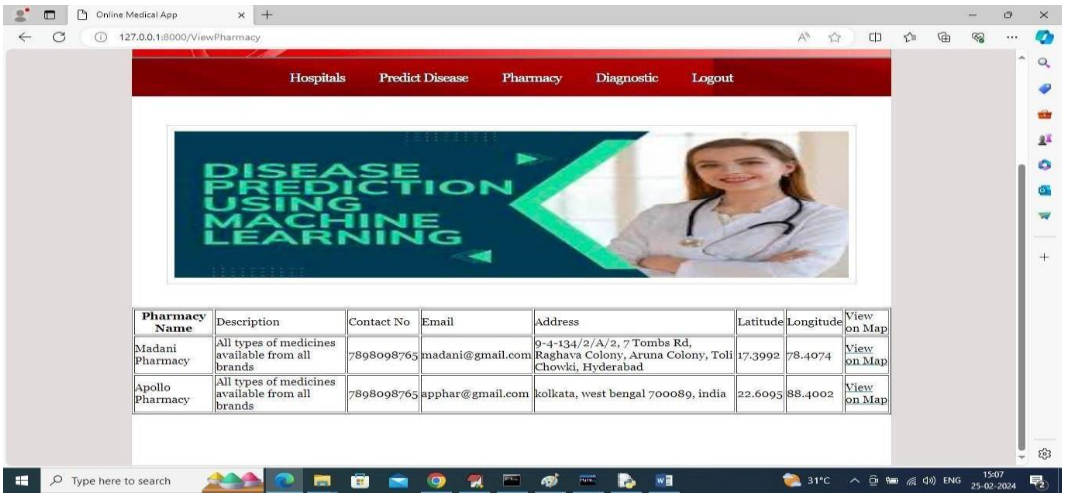

Online Medical App
127.0.0.1:8000/ViewHospitals
DISEASE PREDICTION USING MACHINE LEARNING
Type here to search
31°C
ENG
15:06
25-02-2024

Hospital Name	Disease Speciality	Contact No	Email	Address	Latitude	Longitude	View on Map
Olive Hospital	acne,alcoholic hepatitis,chicken pox,chronic cholestasis,dimorphic hemmorhoids(piles),drug reaction,fungal infection	8887774443	olive@gmail.com	Nanal Nagar, Hyderabad, Telangana	17.394135	78.427273	View on Map
Apollo Hospital	arthritis,bronchial asthma,cervical spondylosis,chronic cholestasis,common cold,dengue,diabetes,drug reaction,fungal infection,heart attack,hepatitis a,hypertension	7898098765	apollo@gmail.com	Road No.59, Jubilee Hills, Hyderabad, Andhra Pradesh, India	17.4276	78.4134	View on Map
Germanten Hospital	allergy,arthritis,bronchial asthma,cervical spondylosis,dengue,diabetes	8889997770	germanten@gmail.com	4-8-138/1/B, opp:Pillar No 150, Attapur, Telangana 500048	17.367781	78.432861	View on Map

In above screen user will see list of available hospitals and can click on ‘View on Map’ link to view hospital on map and get below page.

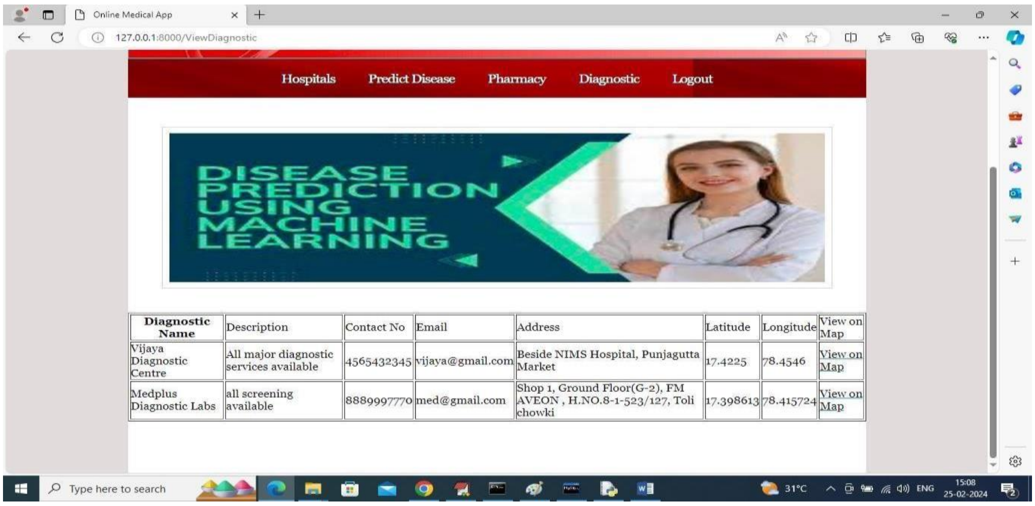


In above screen user will see hospital details on maps and similarly user can see all pharmacy and diagnostic details.



Pharmacy Name	Description	Contact No	Email	Address	Latitude	Longitude	View on Map
Madani Pharmacy	All types of medicines available from all brands	7898098765	madani@gmail.com	9-4-134/2/A/2, 7 Tombs Rd, Raghava Colony, Aruna Colony, Toli Chowki, Hyderabad	17.3992	78.4074	View on Map
Apollo Pharmacy	All types of medicines available from all brands	7898098765	apphar@gmail.com	kolkata, west bengal 700089, india	22.6095	88.4002	View on Map

In above screen user can see list of available pharmacy details.



Diagnostic Name	Description	Contact No	Email	Address	Latitude	Longitude	View on Map
Vijaya Diagnostic Centre	All major diagnostic services available	4565432345	vijaya@gmail.com	Beside NIMS Hospital, Punjagutta Market	17.4225	78.4546	View on Map
Medplus Diagnostic Labs	all screening available	8889997770	med@gmail.com	Shop 1, Ground Floor(G-2), FM AVEON , H.NO.8-1-523/127, Toli chowki	17.398613	78.415724	View on Map

In above screen user can see list of available diagnostic details and now click on ‘Predict Disease’ link to get below page.

6-CONCLUSION

In conclusion, this research successfully demonstrates the development of a disease prediction system utilizing machine learning algorithms, including Decision Tree, Random Forest, and Naïve Bayes classifiers. By analyzing a sample dataset of 4920 patient records across 41 diseases, and optimizing 95 symptoms as independent variables, the study highlights the potential of machine learning in early disease prediction and diagnosis. The comparative analysis of the classifiers showcases their effectiveness in aiding healthcare professionals by providing accurate predictions, ultimately contributing to improved patient care and early intervention strategies in healthcare systems.

Despite the success of the proposed models, certain challenges remain, such as addressing the complexity of medical data, managing imbalanced datasets, and ensuring model interpretability for clinicians. Future work should focus on refining models, incorporating explainability frameworks like SHAP or LIME, and integrating new types of medical data, including genetic and imaging data, to expand the scope of disease prediction.

REFERENCES

[1] Ankita Dewan and Meghna Sharma, “Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification” IEEE, 978-9- 3805-4416-8/15, pp. 704-706, 2015.

[2] Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, “Designing Disease Prediction Model Using Machine Learning Approach” IEEE Xplore Part Number: CFP19K25-ART; ISBN: 9781-5386-7808-4, pp. 1211-1215, 2019.

[3] Dhomse Kanchan B. and Mahale Kishor M., “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis” IEEE, 978-1- 509004676/16, pp. 5-10, 2016.

[4] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang,“Disease prediction by machine learning over big data from healthcare communities” IEEE Access, vol. 5, no.1, pp.8869– 8879, 2017.

[5] M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu, “A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries,” J. Am Med Inform Assoc, vol. 18, no. 5, pp. 601–606, 2011.

[6] Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, “An Improved Method for Disease Predictionusing Fuzzy Approach” IEEE, DOI 10.1109/ICACCE.2015.67, pp. 569572, 2015.

[7] Lambodar Jena and Ramakrushna Swain, “ChronicDisease Risk Prediction using Distributed Machine Learning Classifiers” IEEE, 978- 1-5386-2924-6/17, pp. 170-173, 2017.

[8] Sayali Ambekar, Rashmi Phalnikar, “Disease RiskPrediction by Using Convolutional Neural Network” IEEE, 978-1-5386-5257-2/18, 2018.

[9] Turing A. Computing machinery and intelligence. *Mind.* 1950;**LIX**(236):433–460. doi: 10.1093/mind/LIX.236.433.

[10] Wati D.A.R., Abadianto D. Design of face detection and recognition system for smart home security application.; 2017 2nd International conferences on Information Technology.