Human Disease Prediction based on Symptoms

There are new diseases discovered in the 21st century which have identical and additional symptoms to the previous diseases but these new diseases are far more dangerous than the previous ones and have some additional symptoms. S o, In this work the new diseases can be identified by taking the input from the user and identifying the disease and displaying it on the users smart screen display. This work aim is to reduce the deaths of people by identifying the disease and cure it in advance. The Random Forest and Naive Bayes algorithms are used for predicting the disease. This prediction is done by considering and comparing the accuracies of both the algorithms and gives the predicted disease as output which has best accuracy. Its implementation is completed via python programming language and tkinter library..

**Description of Existing Systems**

This literature review normally concentrates on the sources that support the Machine learning and hospital related approaches analysis. The analysis articles outlined fully however numerous algorithms and frameworks are perhaps accustomed to classify facts.

Authors used four machine learning procedures: Decision Tree, Naive Bayes, Random Forest and KNN. Their result is computed using in a way that when an end handler enters all the signs and simply clicks on the Random Forest button. This paper tells that they have the same accuracy from all the four algorithms. The accuracy mentioned was 0.976. The drawback of this system is the user was not aware of algorithms to choose them wisely [1].

Authors made a disease prediction system using multiple ML algorithms. More than 230 diseases were present in the data set that was processed. The breakdown system identifies the disease as the output based on a person’s symptoms, gender and age. The KNN algorithm has given better results when compared with all the other algorithms. The accuracy mentioned by the weighted KNN algorithm was 93.5%. Drawback was the number of diseases [2]. In this Classification Decision tree, Random Forest, Naïve Bayes, SVM , KNN Algorithms are used to forecast diseases. System accuracy reaches 98.3%. Comparison between all the algorithms performance was done based on the accuracy of the algorithm. Decision Tree gives 84.5%, Random Forest performs much better than all the algorithms 98.95%,SVM 96.49%, KNN 71.28%,Naïve Bayes 89.4%. Since it compares the number of algorithms it requires more t ime complexity for the execution of the systemit is the major drawback of this system[3].

This paper discusses an algorithm that predicts a

disease from its symptoms. On the user interface symptoms are provided in a drop down box from which the end user can select the signs of his disease. In addition to the disease the research can suggest the medication for the user based on his symptoms [4].

Authors made a prediction machine for the diseases like heart, kidney, liver, diabetes, breast cancer by using classification. A prediction accuracy probability of 95% is obtained on average. Since the risk level was higher due to the diseases being more effective compared to other diseases [5].

Authors have taken into consideration a database of 106 attributes as symptoms, from it they're predicting the forty-three illnesses. After detecting the ability of predicting

the ailment. They used KNN algorithm and Naïve bayes for their technique. This research embeds the ML predictive technique that abstracts information from existing information units to see and expect destiny consequences [6].

Authors are using machine learning algorithms like SVM Support Vector Machine and MLR Multilinear Regression that try to accurately predict possible diseases. They analyzed the accuracy of this system for 5 different diseases. The results generated by the proposed system have an accuracy of up to 87%. The diseases were limited since it can predict only 5 diseases

**Advantages of the Existing Systems:**

High Accuracy: Several systems report accuracy levels exceeding 90%, indicating their potential for effective disease prediction.

Multi-algorithm Support: Some systems utilize multiple algorithms like Decision Trees, Random Forests, and KNN, offering flexibility and robustness.

Consideration of Multiple Factors: Some systems take into account factors like gender, age, and symptoms for a more comprehensive analysis.

Disease Prediction and Treatment Recommendations: Some systems suggest medications based on the predicted disease, adding a valuable feature.

User-friendly Interfaces: Drop-down menus and symptom selection boxes simplify user interaction with the system.

**Disadvantages of the Existing Systems:**

Limited Disease Coverage: Some systems focus on a small number of diseases, limiting their applicability.

Black Box Problem: Some systems lack transparency in their decision-making process, raising concerns about trust and explainability.

Data Privacy Concerns: Privacy and security measures need to be robust to protect user information.

Overreliance on Accuracy: High accuracy numbers may not translate perfectly to real-world scenarios, requiring further validation and testing.

Limited User Awareness of Algorithms: Some systems lack guidance for users to choose the most relevant algorithm for their needs.

Potential for Misdiagnosis: Incorrect symptom input or system errors can lead to misdiagnosis, necessitating careful user education and system validation.

**PRO PO SED METHO D :**

Design Methodology

The Disease Prediction based on signs is developed to overcome standard sickness in advance degrees as all of us realize in the serious stage of the problem and to avoid the death rate in society due to lack of awareness among the people about the deadly diseases. The project "Disease Prediction based on signs and symptoms'' is implemented using python completely. Even the interface of this assignment is carried out by the usage of python's library interface called T-kinter. This prediction is basically finished with the assistance of the Random Forest algorithm and Naive Bayes algorithm with a set of rules. Right here first the user needs to enter his name, age, gender and blood group then choose the symptoms from the given drop-down menu. After entering the symptoms user can click the predict button to predict the disease which is related to given signs. The user needs to enter the signs to expect the disease, for better accurate results the user wishes to enter all of the given symptoms, then the machine will offer the correct end result along with its accuracy between 0 and 1. To reset the values there is a clear button and to exit the navigation or user interface an exit button is provided.

3.2 System Architecture Diagram

Disease prediction using machine learning predicts the presence of the disease for the person based on various signs and symptoms. The structure of the Disease prediction Architecture by using machine learning encompasses various datasets which can exa mine the symptoms of the person and predict it. In between preprocessing step is done in which all the noisy data can be cleared, missing values and incomplete data while collecting the data from the various hospitals and different persons with different symptoms.

Preprocessing can be done by various techniques such as tuple can be ignored if it has more missing values or automatically filling manually, which is a difficult task so it can be done by automatically filling the values by mode or it can also be done by binning, regression and clustering. Then datasets are converted into the smaller units by column wise and from there it gets categorized based on the classification algorithms in a while the labelled statistics is then processed into the thru which the records receive processed and goes in to the disease prediction model using all the inputs from the person that is stated above. After taking the inputs it will convert the input data into binary format. It will consider the 0s tuple of each disease and then it will replace the given symptoms values with 1s. Then it compares with all the records in the dataset and it will find the similar tuple or nearly matchable tuple. Then after the user gets into the above facts and overall processed information combines and compares inside the prediction version of the system and subsequently predicts the disorder.

A structure diagram is a graphical illustration of a set of principles, which might be part of an architecture, consisting of their principles, elements, and components. The diagram e xp lains about the system software in the belief of review of the device. An ensemble of machine learning techniques is used for class, regression and other tasks that operate to build many selection trees at training time.

**Advantages of the Proposed System:**

Early Disease Detection: The system aims to predict diseases in early stages, potentially improving treatment outcomes and reducing mortality rates.

Accessibility: The Python-based implementation and Tkinter interface make the system potentially accessible to a wider audience.

User-friendly Interface: The drop-down menu for symptom selection and buttons for prediction and reset offer a simple and intuitive user experience.

Multiple Algorithms: Utilizing both Random Forest and Naive Bayes algorithms provides a degree of redundancy and potentially improves accuracy.

Accuracy Transparency: The system provides accuracy scores between 0 and 1, allowing users to assess the confidence level of the predictions.

**Disadvantages of the Proposed System:**

Limited Disease Coverage: The system's accuracy depends on the training data, and it may not be effective for less common or newly emerging diseases.

Black Box Problem: The system's decision-making process remains hidden, potentially raising concerns about transparency and trust.

Data Quality Dependence: The system's accuracy relies heavily on the quality and completeness of the training data, requiring careful data acquisition and preprocessing.

Overreliance on Accuracy Scores: High accuracy scores in controlled settings may not translate directly to real-world scenarios, requiring further testing and validation.

Potential for Misdiagnosis: Incorrect symptom input or system errors could lead to misdiagnosis, highlighting the need for careful user education and system validation.

Ethical Considerations: Privacy and security of user data must be addressed to ensure responsible use and prevent misuse of sensitive information.

**SYSTEM SPECIFICATION:**

**HARDWARE REQUIREMENTS:**

* **System :** Intel i7
* **Hard Disk :** 1 TB.
* **Monitor** : 14’ Colour Monitor.
* **Mouse :** Optical Mouse.
* **Ram :** 8GB.

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 10.
* **Coding Language :** Python.
* **Front-End :** Html. CSS
* **Designing :** Html,css,javascript.
* **Data Base :** SQLite.

**REFERENCES**

[1] M Vamshi Krishna Reddy, G V P Sai Abhijith, K Sai Nath, Mangali Sathyanarayana(2022) Disease Predictor Based on Symptoms Using Machine Learning.

[2] Rinkal Keniya, Aman Khakharia, Vruddhi Shah,Vrushabh Gada, Ruchi Manjalkar, Tirth Thaker, Mahesh Warang, Ninad Mehendale.(2020) Disease Prediction from various Symptoms using Machine Learning .

[3] Dr C K Gomathy, Mr. A. Rohith Naidu (2021) The prediction of disease using machine learning

[4] Priyanka J. Panchal, Shaeezah A. Mhaskar, Tejal S. Ziman, "Disease Prediction using Machine Learning", 2020.