Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.10 : 2024
ISSN :**1906-9685**

MULTIPLE DISEASE PREDICTION USING VOICE ASSISTANT AND REGIONAL LANGUAGES

[1]I.Deepthi Priya, [1]L.Lokesh, [1]G.Jayanth Roy, [1]G.Viswas,[1]D.Raditya
[2]Nekkanti Renu
[1]UG Students,Computer Science and Engineering,Vignan'sInstitute Of Information Technology, Duvvada, Visakhapatnam
[2]Assistant Professor, Computer Science and Engineering,Vignan's Institute Of Information Technology, Duvvada, Visakhapatnam

ABSTRACT

Machine learning techniques have become increasingly important in healthcare. They can help predict illnesses and provide decision support. When someone is sick, they often need to see a doctor, which can be time-consuming and costly. This can also be challenging if the person is far from medical facilities and as a result the illness cannot be detected. Rather than making an appointment and visiting a hospital, now you can get accurate solutions online from the comfort of your own home. Machine Learning lets a system anticipate illnesses from user data. We created a multiple Disease Diagnosis System using Voice Assistant through Machine Learning. We developed by keeping this in mind to help both literate and illiterate- people. This system serves as a suggestion system to doctors and the users for the treatment and any subsequent actions regarding the treatment is completely the responsibility of the doctors. The customer will be able to figure out how likely it is to use the signs and symptoms of a disease and enjoy the comfort of knowing their output through voice in their preferred language. AI in medicine analyzes healthcare information with computer programs.It detects patterns in lots of data and these sophisticated algorithms use statistical mode-ls to make predictions based on ide-ntified trends in various datasets.

KeywordsMachine Learning,VoiceAssistant,healthcare,sophisticated,datasets

---

1.INTRODUCTION

In terms of data collection and processing, the healthcare industry can be pretty overwhelming. With the digital age and all the tech around, there's just tons of information about patients being generated like illness diagnoses, medical records, and even data from medical equipment. All this data needs to be sorted, analyzed, and made sense, so that doctors and other healthcare professionals can make informed decisions. Now, medical data mining is one way to make sense of all this information. By looking for patterns and connections between different pieces of data, we can learn a lot more about how certain things work in the medical world. For instance, we might find that certain symptoms tend to show up together in certain illnesses. One really cool thing about medical data mining is that it can help us predict diseases using machine learning (ML). By analyzing huge databases of medical data, we can train ML models to spot patterns and make predictions about who might be at risk for certain illnesses. One really cool thing about medical data mining is that it can help us predict diseases using machine learning (ML) This could be really helpful in helping healthcare professionals provide better care for their patients and maybe even prevent some diseases from happening in the first place. Consequently, there is considerable interest among medical practitioners in automating diagnosis through the fusion of machine learning methods and physician knowledge. Data mining and machine learning techniques are actively working to convert readily available data

into useful insights, thereby enhancing the efficiency of the diagnostic process. Studies have revealed that machine learning algorithms achieve a higher accuracy rate of 91.1% compared to the 79.97% accuracy rate of the most seasoned physician in diagnosing medical conditions. Furthermore, aside from its accuracy, the inclusion of a voice assistant aims to offer convenience to both literate and illiterate individuals via the Google voice assistant. This feature employs artificial intelligence (AI) and natural language processing (NLP) techniques to furnish users with voice-activated support and language translation capabilities.

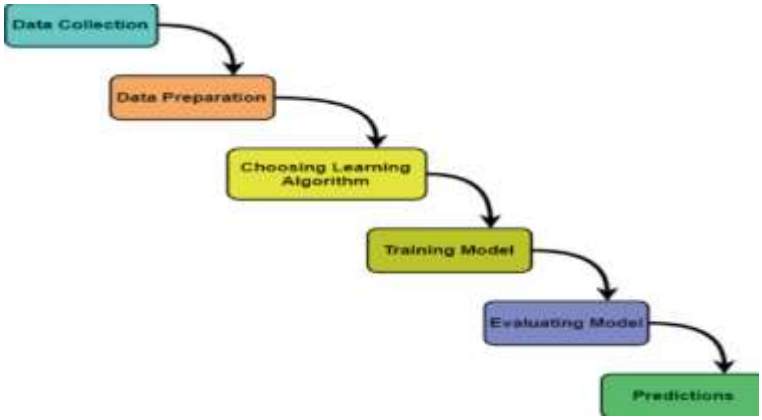

Fig 1. Workflow of Machine Learning

The depicted process in Fig 1 for multi-disease prediction encompasses several stages, including data gathering, data preprocessing, algorithm selection, model training, model evaluation, and ultimately, making predictions.

Data Collection

It includes collection of data related to healthcare from different sources and ensures data quality, consistency, and compliance with privacy regulations.

Data Preparation

It includes cleaning of the collected data and pre-processing of data to handle missing values, outliers, and inconsistency.

Choosing Learning Algorithm

This step includes data exploration to understand its characteristics depending upon which appropriate machine learning algorithms are chosen and taken into consideration.

Training Model

In this step dataset is split into testing and training dataset, for model training. Selected machine learning models are trained using training data.

Evaluating Model

Evaluation of the models using proper evaluation metrics is done on the validation sets.

Predictions

Lastly, model validation is performed using an independent test dataset for accessing the model's generalization of new data.

2.LITERATURE SURVEY:

Gresha Bhatia et al, implemented deep learning methods for disease prediction. They mostly worked on diseases that occur due to change in weather conditions like mosquito-borne diseases.[1].Dataset is collected from health institute in USA. Deep Recurrent Neural Network (RNN) along with LSTM to remember previous data. Mean absolute error of 13.05 is calculated from visual representations but the model is not perfectly accurate. More hypothesis testing should be done to get more accurate predictions.

C K Gomathy et al, worked on machine learning models for the prediction of diseases. A Supervised learning algorithm called Naıve Bayes Classifier is used to pre the probability of the disease[2]. KNN and linear regression are also used in predicting diseases. They obtained a system accuracy of 98.3% by using measures like recall, precision, accuracy and f1-measure.

Rinkal Keniya et al, took the advent of machine learning to predict diseases based on various symptoms[3]. They used multiple ML algorithms to create a medical diagnosis system that predicts the disease by using a dataset that consists of 230 plus processed diseases. They implemented weighted KNN algorithm which gave an accuracy of 93.5%.

Rayan Alanazi et al, different approaches of machine learning are used for identification and prediction of chronic diseases.[4] They used Convolutional Neural Network (CNN) and KNN for disease prognosis. The dataset is prepared along with living habits. The proposed system obtained an accuracy of 95% more than the existing algorithms.

Palle Pramod Reddy et al, implemented machine learning algorithms for disease prediction[5]. They worked on random forest classifier to accomplish the task of predicting the diseases. The classifier is trained using flask framework. The data is taken from medical records and latent vector is used for filling missed values. A prediction probability of 95% is obtained. They used grails system for disease prediction.

Marouane Ferjani et al, worked on prediction of diseases using machine learning algorithms[6]. The aim of the research is to see how machine learning algorithms can be used to improve the health care. Algorithms like KNN, linear regression, random forest, support vector machine etc are used for prediction of different diseases and their respective accuracies are calculated. Complex algorithms can enhance the performance of disease prediction.

K Arumugam et al, implemented multiple disease prediction using algorithms of machine learning[7]. A finetuned decision tree is used for better performance in forecasting. The framework used algorithms like decision tree, support vector machine, naıve bayes. The traditional check up by physician the accuracy is 79.97% whereas machine learning algorithms give a correctness of 91.1%.

Aftab Attar et al, they used techniques of machine learning and deep learning for the prediction of diseases[8]. They used sequential CNN of 8 layers in which 5 are convolutional layers and 3 are fully connected layers. They used a real dataset from medical records which consists of nearly 5842 images for training the model. They used certain modules like TensorFlow and Keras to build the system in python language. They achieved an accuracy of 75% over 30 epochs.

3MACHINE LEARNING

Machine learning (ML) is a subset of artificial intelligence (AI) concerned with the creation of algorithms and statistical models. These tools empower computers to acquire knowledge from data and utilize it to make predictions or decisions.The fundamental concept of machine learning revolves around identifying patterns within data and employing these patterns to make knowledgeable decisions or predictions without the need for explicit programming for every conceivable scenario. ML algorithms undergo iterative learning from data, enhancing their effectiveness gradually as they encounter additional data. Various methodologies exist within machine learning, such as supervised learning, unsupervised learning, and reinforcement learning.In supervised learning, algorithms undergo training using labeled data, where each data point corresponds to a specific target label or output. The objective is for the model to grasp the relationship between inputs and outputs, enabling it to predict outputs accurately for new, unseen data. Conversely, unsupervised learning operates with unlabeled data, seeking to uncover concealed patterns or structures within the dataset. Reinforcement learning entails teaching an agent to make decisions by engaging with an environment and receiving feedback, either through rewards or penalties.Machine Learning is utilized across diverse fields, including natural language processing (NLP), computer vision, predictive analytics, recommendation systems, and beyond. In NLP, ML algorithms are employed for tasks like sentiment analysis, language translation, and chatbots.

Within computer vision, ML models are utilized for tasks such as object detection, image classification, and facial recognition. Predictive analytics leverages ML to forecast future events or outcomes based on past data, while recommendation systems use ML to furnish users with tailored suggestions according to their preferences and activities.Overall, machine learning continues to drive innovation and automation across industries, unlocking new possibilities for data-driven decision-making and intelligent systems.

4 PROPOSED MODEL

The proposed system aims to develop a voice-based assistant capable of predicting multiple diseases based on user symptoms, while also supporting regional languages. The system will provide an intuitive interface for users to interact with and receive predictions in their preferred language.
We've employed algorithms to construct a system that utilizes voice assistance to predict a patient's disease based on their symptoms. By comparing these symptoms with the system's pre-existing dataset, the machine can determine the presence or absence of a disease and convey the results through voice using text-to-speech (TTS) technology and a language translator.

- **Text Input and Processing** It would start with the interaction of system with user through voice assistant,such as Google Assistant, in their regional language to read the attributes in the dataset to take input. The voice assistant uses speech recognition technology to convert the text into preferred language of the user to take input for healthcare-related symptoms.
- **Regional Language Translation** The text data in the regional language is then passed through a regional language translation module. This module translates the text into a standardized medical language or English, which is commonly used in healthcare datasets and systems.
- **Disease Prediction Model** The translated text data, now in a standardized medical format, is fed into a disease prediction model. This model can be based on machine learning algorithms, such as SVM, logistic regression model to predict multiple diseases based on symptoms or risk factors.
- **Prediction Output and Feedback**The model generates assessments for various diseases based on the input symptoms or information. The results are translated back into the regional language by the language translation module to provide feedback to the user through the voice assistant.

4.1 VOICE ASSISTANT

Google's voice assistant, a virtual assistant powered by artificial intelligence (AI) and developed by Google, allows users to interact with devices using natural language instructions. It utilizes sophisticated natural language processing (NLP) and machine learning techniques to comprehend and address a broad spectrum of queries and tasks. These tasks include setting reminders, answering questions, providing weather updates, playing music, controlling smart home devices, and more. Integrated into various devices like smartphones, smart speakers, and wearables, Google's voice assistant provides users with a seamless and hands-free means to access information and execute tasks through voice commands, thus augmenting convenience and productivity in daily activities.

4.2 TEXT-TO-SPEECH

Voice assistants, such as Google's voice assistant, offer functionality to convert text to speech, providing a natural and accessible way to interact with content. This text-to-speech (TTS) capability utilizes advanced speech synthesis technologies to generate human-like speech from written text. Users have the option to input text either through voice commands or text input interfaces. The voice

assistant then analyzes the text using linguistic rules, natural language processing (NLP) methods, and machine learning algorithms to generate spoken responses.

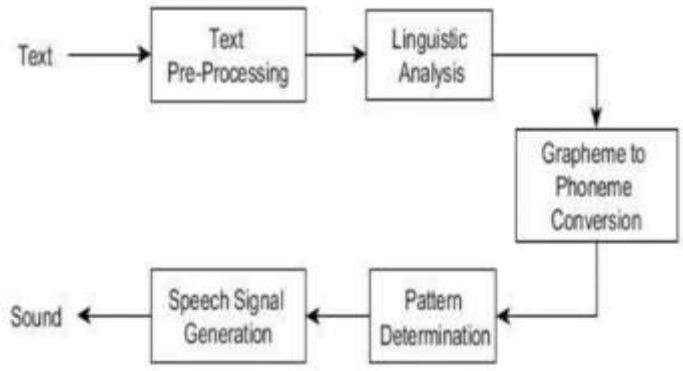

Fig 2. Text-To-Speech

From fig 2 the TTS feature of voice assistants is particularly beneficial for users with visual impairments, language learners, individuals with reading difficulties, or anyone seeking hands-free access to information. Additionally, users can often customize the voice preferences, including choosing different languages, accents, and speech speeds, to suit their preferences and needs. Overall, the text-to-speech functionality in voice assistants contributes to a more inclusive and versatile communication environment, empowering users to consume and interact with digital content effortlessly through spoken language.

5 RESULTS AND ANALYSIS

5.1 DATASET

The dataset used for multiple disease prediction typically comprises structured and/or unstructured healthcare data collected from various sources. The selection of dataset relies on the particular research or application scenario, the nature of the diseases under consideration for prediction, and the accessibility of data resources.The dataset utilized for predicting multiple diseases usually comprises electronic health records (EHRs), which encompass various patient details like demographics, medical background, lab test outcomes, prescribed medications, and diagnoses. These datasets are frequently enriched with medical imaging data, such as X-rays, MRI scans, and histopathology images, especially for diseases necessitating visual assessment. The amalgamation of these varied data sources facilitates the creation of robust machine learning models for precise and early forecasting of multiple diseases.

5.2 DATASET AUGUMENTATION

Dataset augmentation is a crucial technique in enhancing the performance and generalization capabilities of machine learning models for multiple disease prediction tasks. In the context of healthcare, where datasets may be limited in size and imbalanced across disease classes, augmentation strategies are essential. Techniques such as synthetic data generation, feature engineering, and image augmentation can be applied. Feature engineering involves creating informative features related to diseases, symptoms, demographics, and temporal patterns, enriching the dataset's information content. However, it is critical to ensure data privacy, compliance with

regulations, and ethical considerations while augmenting healthcare datasets to maintain patient confidentiality and trust in predictive models. Overall, dataset augmentation plays a pivotal role in improving the accuracy, robustness, and fairness of multiple disease prediction models, benefiting healthcare decision-making and patient outcomes.

5.3 METRICS

Metrics play a crucial role in evaluating the performance of multiple disease prediction models. Here are some common metrics used in the context of multiple disease prediction:

- **Accuracy**
  Accuracy evaluates the ratio of accurately predicted instances to the total instances within the dataset. While it serves as a primary metric for evaluating overall model effectiveness, it might not adequately address imbalanced datasets characterized by uneven distribution of classes.
- **Precision and Recall**
  Precision quantifies the proportion of accurately predicted positive instances among all instances predicted as positive, while recall (also known as sensitivity) gauges the ratio of correctly predicted positive instances to all actual positive instances in the dataset. In healthcare contexts, precision and recall play crucial roles in managing false positives and false negatives, respectively, thereby ensuring precise disease predictions while mitigating misclassifications.
- **Specificity**
  Specificity quantifies the proportion of accurately predicted negative instances among all actual negative instances within the dataset.It is particularly relevant in disease prediction scenarios where correctly identifying non-disease cases is crucial to avoid false alarms or unnecessary interventions.

5.4 RESULT

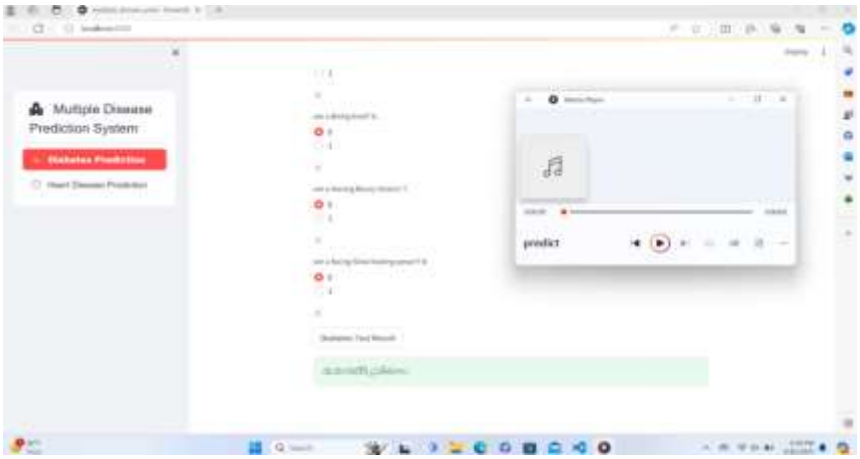

Fig 3. Diabetes test result

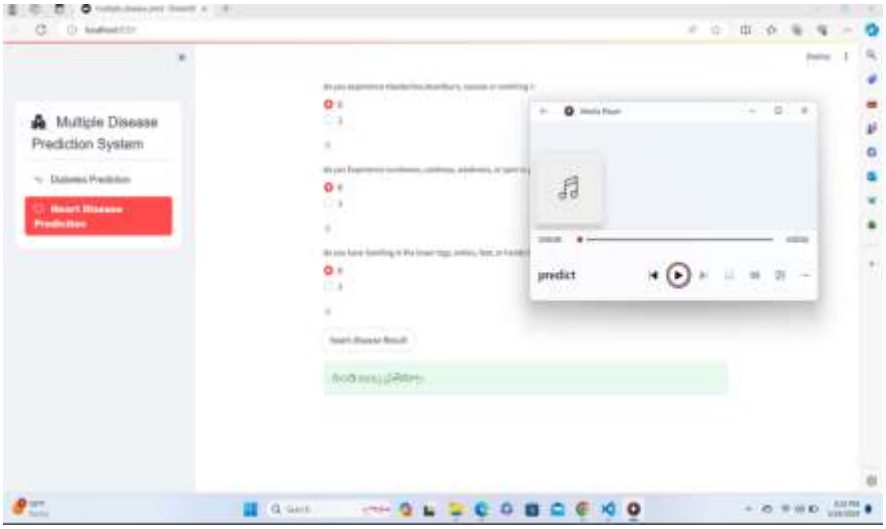

Fig 4. Heart Disease Test Result

6 .CONCLUSIONS

The integration of voice assistance and regional language capabilities into multiple disease prediction systems represents a transformative advancement in healthcare accessibility and accuracy. The problem of linguistic barriers and limited access to healthcare information for non-English speakers is addressed by enabling natural language interactions in regional languages. The existing system often struggles with prediction of one disease by the system and language-specific problems lacking completeness, hindering effective disease prediction for diverse populations. The proposed system leverages voice commands in regional languages, coupled with advanced natural language processing and machine learning techniques, to enhance disease prediction accuracy and user experience. Results showcase improved model performance and user engagement. Future scope lies in refining language models, integrating real-time data sources, and incorporating feedback mechanisms for continuous learning and personalized healthcare recommendations, ultimately advancing equitable healthcare delivery globally.

7 REFERENCES

[1] Gresha Bhatia et al, implemented Disease Prediction using Deep learning to predict the mosquito-borne diseaseswhich represent numerous public health challenges May 21-23, 2021

[2] C K Gomathy et al, worked on prediction of disease using machine learning from the symptoms given by the users ,December 2021

[3] Rinkal Keniya et al, worked on Disease prediction from various symptoms using machine learning The accuracy of the weighted KNN algorithm for the prediction was 93.5 %. Our diagnosis model can act as a doctor for the early diagnosis of a disease to ensure the treatment can take place on time and lives can be saved , 8 October 2020

[4] Rayan Alanazi et al, Identification and Prediction of chronic diseases using machine learning approach data mining plays a critical role in disease prediction. The system offers a broad disease prognosis based on patient's symptoms by using the machine learning algorithms such as convolutional neural network (CNN) , 25 Feb 2022

[5] Palle Pramod Reddy et al, implemented Disease Prediction using machine learning.The method examines the user's symptoms as input and returns the disease's likelihood as an output. Disease prediction is accomplished using the random forest classifier, 5 May 2021

[6] Marouane Ferjani et al,supervised machine learning (ML) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases, December 2020

[7] K Arumugam et al, Multiple Disease Prediction using machine Learning algorithms. Diagnosis of chronic illnesses is a vital issue in the medical industry since it is based on many symptoms, volume 80 part 3 2020

[8] Aftab Attar et al, Disease prediction using machine learning and deep learning , The Health Industry plays a major role in curing the diseases of the patients and our system in turn aids the health industry through disease identification and prediction, July 2022