VOICE BASED HEALTH DIAGNOSIS USING MACHINE LEARNING PREDICTION

Ayushi Garg[1], Sadiya Zainab[2], Shivani S Avadhani[3], Sushmitha M[4], Suhas S[5]

[1,2,3,4]*BE, Department of CSE, NIE Mysore, Karnataka, India*
[5]*Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India*

---***---

Abstract - *There are lot of treatments that are available for various diseases. No human can possibly know about all the medicines and the diseases. So, the problem is that there isn't any place where anyone can have the details of the diseases or the medicines. What if there is a place where you can find your health problem just by entering symptoms then it will help us to deduce the problem and to verify the solution. Our project "Voice Based Health Diagnosis Using Machine Leaning Prediction" using Machine Learning is developed so that users can get to know the disease and its treatment. The perspectives in this project are the diseases and their respective treatments. Our vision is to build a real time application which can be used in remote areas and in case of emergencies. The information of the members and the database that includes the diseases and treatments can updated regularly by the administrator who has the access to the whole information. The member will be asked to enter an individual's details for successful registration. Once registered, the person must input the information like gender, blood pressure, height, weight and the symptoms through voice or text. Our Application uses K- nearest neighbour (KNN) which is an algorithm under Machine Learning. KNN is the most efficient and accurate algorithm for disease prediction. Since the access to doctors in remote areas at the time of emergencies is not instant, our application will help in saving lives.*

Key Words: **Machine Learning, K- nearest neighbour (KNN), Database.**

1. INTRODUCTION

People care deeply about their health and want to be more than ever in charge of their health and healthcare. Life is more hectic than it has ever been. The medicine that is practiced today is not only based on years of practice but on the latest discoveries as well. Tools that can help us manage and keep better track of our health, help make people more powerful when it comes to healthcare knowledge and management. The traditional healthcare system is also becoming one that embraces the Internet and the electronic world. People want fast access to reliable information and in a manner that is suitable to their habits and workflow. Medical care related information is a source of power for both healthcare providers and people. So, we are developing an application that predicts the diseases and their treatments based on the symptoms that are input. With this application one can have faster access to heath diagnosis in case of emergencies where one requires immediate medical attention.

1.1 Prediction System

In this paper the focus on machine learning technique to predict diseases-based symptoms provided through voice or text input. Online JavaScript ready library is used for voice input. Once the input is given, it uses Keyword Extraction Technique (KET) which comes under Natural language Processing. The KET takes the input and extracts the required keyword for processing the data. The disease prediction is done by machine learning using K-nearest neighbor (KNN) algorithm. It considers the parameters such as age, gender, blood pressure, height, weight and symptoms.

1.2 K-nearest neighbor (KNN)

KNN is a simple algorithm which uses entire dataset during its training phase whenever prediction is required for unseen data. It searches through entire training dataset for k most similar instances and data with most similar instance is returned.

1.2.1   Features of KNN

1) KNN stores the entire training dataset which it uses as its representation.
2) It makes predictions just-in-time by calculating the similarity between an input sample and each training instance.
3) It works on similarity measures.

2. RELATED WORK

World needs better, faster, and more reliable access to information. In the medical domain, the richest and most used source of information is Medline database of extensive life science published articles. All research discoveries come and enter the repository at high rate, making the process of identifying and disseminating reliable information a very difficult task. One task is automatically identifying sentences published in "Automatic Medical Disease Treatment System Using Data Mining" by M. Thagamani, of medical abstracts (Medline) containing or not information about diseases and

Treatments and automatically identifying semantic relations that exist between diseases and treatments [1].

M. Craven examined the problem of distinguishing among seven relation types that can occur between the entity's "treatment" and "disease" in bioscience text, and the problem of identifying such entities. The scheme was the correct management of word position information, which may be critical in identifying certain relationships. In this approach that facilitates the automatic recognition of relationships defined between two different concepts in text [2].

Based on the above research papers, we are proposing a system which predicts general disease and its respective treatments based on the given voice or textual input by the user. Since there are no web-based applications that are specialized in general diseases and are disease specific, we are proposing a web-based application that predicts general diseases. In this system, the users can provide the input of their symptoms. The application uses KNN that gives the output to the user in the form of the disease and its related treatment. It is an algorithm used in supervised learning which stores all available cases and classifies new cases based on similarity measures.

3. METHODOLOGY

K-nearest neighbors classifier (KNN)

K-Nearest neighbors (KNN) [3] is a simple, lazy and nonparametric classifier. KNN is preferred when all the features are continuous. KNN is also called as case-based reasoning and has been used in many applications like pattern recognition, statistical estimation. Classification is obtained by identifying the nearest neighbors to determine the class of an unknown sample. KNN is preferred over other classification algorithms due to its high convergence speed and simplicity. Figure1 shows nearest neighbors classification. KNN classification has two stage
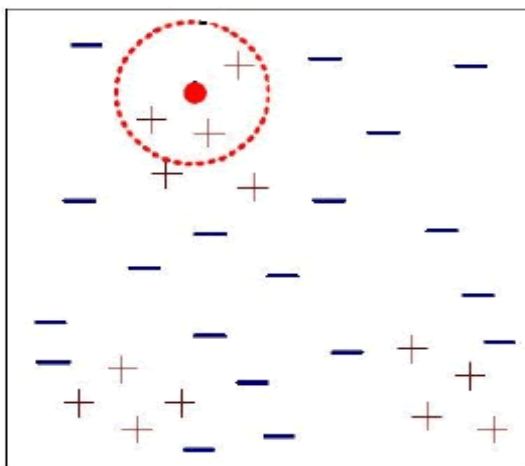

Figure 1

1)To find the k number of instances in the dataset that is closest to instance S

2) These k number of instances then vote to determine the class of instance S

The accuracy of KNN depends on distance metric and K value. Various ways of measuring the distance between two instances are cosine, Euclidian distance. To evaluate the new unknown sample, KNN computes its K nearest neighbors and assigns a class by majority voting.

Natural Language Processing (NLP)

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken. The development of NLP applications is challenging because computers traditionally require humans to "speak" to them in a programming language that is precise, unambiguous and highly structured, or through a limited number of clearly enunciated voice commands. Human speech, however, is not always precise -it is often ambiguous, and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context [4] The NLP uses Keyword extraction technique to process the data. Keyword Extraction is tasked with the automatic identification of terms that best describe the subject in a context [5]

Proposed Approach

Our proposed method aims to enhance the performance of KNN classifier for disease prediction. Algorithm for our proposed method is shown below

Step 1: Input: Symptoms through voice or text.

Step 2: Output: Disease predicted, and treatment recommended.

Step 3: Input the data set.

Step 4: Apply KNN algorithm for input parameters.

Step 5: Euclidean distance between existing parameters and newly entered is calculated.

Step 6: Based on similarity measures output is predicted.

The nearest neighbor algorithm measures the distance $d_e(X_i, X_j)$ between query points X_i and a set of training samples X_j to classify a new object based on majority of K-nearest neighbor category of Y attributes of training samples.

Query point $X_i = x_1, x_2, x_3, \ldots\ldots, x_n$

Training Sample $X_j = x_1, x_2, x_3, \ldots\ldots, x_n$

$$Dist(c_1, c_2) = \sqrt{\sum_{i=1}^{N}\left(attr_i(c_1) - attr_i(c_2)\right)^2}$$

$$k - NearestNeighbors = \left\{k - MIN(Dist(c_i, c_{test}))\right\}$$

$$prediction_{test} = \frac{1}{k}\sum_{i=1}^{k}class_i \; (or \; \frac{1}{k}\sum_{i=1}^{k}value_i)$$

4. FUTURE ENHANCEMENTS

The future enhancements of the project are to focus on extending it as a mobile based application and to use advanced sophisticated algorithm for improvising the voice input. We use KNN for prediction. Other algorithms like ID3 or C4.5 or Naive Bayes Algorithms can also be used. Future, we may add more and different disease and related treatments to the dataset which makes it more approachable to the people.

5. CONCLUSION

This application is efficient in predicting diseases and providing medication details. Our system provides a dedicated web-based application over mobile based application. Our project "Voice Based Health Diagnosis Using Machine Leaning Prediction" system is integrated with trained dataset. There are many Web Based Applications which focus on specific kind of disease prediction by providing the input only through text. As the paper suggest us to implement web-based application for generic diseases prediction by providing the input through either text or voice making it more approachable.

ACKNOWLEDGEMENT

REFERENCES

[1] "Summarization and sentiment analysis from user health post", MAEER'S MIT, Pune, India, 2018 IEEE

[2] "Automatic Labelling of Important Treatment System using Data Mining", Conference on Information and Communication Technology, 2017 IEEE

[3] http://www.alliedacademies.org/articles/prediction-of-heart-disease-using-knearest-neighbor-and-particle-swarm-optimization.html

[4] https://searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP

[5] https://en.wikipedia.org/wiki/Keyword_extraction