

DISEASE PREDICTION USING MACHINE LEARNING

Velamala Naga Durga Gayathri (MCA Scholar), B V Raju College, Vishnupur, Bhimavaram,
West Godavari District, Andhra Pradesh, India, 534202.

S. K. Alisha, B V Raju College, Vishnupur, Bhimavaram, West Godavari District, Andhra
Pradesh, India, 534202.

Abstract

In this digital world, most of the people are prone to diseases, due to lack of healthy food, proper sleep and daily exercise. It is very crucial to know if we are suffering from a disease, at an early stage rather than discovering it at a later stage. Hence disease prediction system plays an important role as it predicts the diseases based on symptoms. This disease prediction system uses Machine Learning algorithm named Random Forest. This system also suggests drugs that are most commonly used to cure the disease.

Indexed Terms- Diseases, Drugs, Machine Learning, Prediction, And Random Forest.

1. INTRODUCTION

With the advancement in technology, Machine Learning is becoming more popular and commonly used technology by industry experts for solving problems faced in real life. Machine Learning is the scientific study of algorithms and statistical models that computer use to perform a specific task without using explicit instructions, relying on patterns and inference instead. Machine Learning is also used by the healthcare industry to bring advancement in their techniques so that they can provide better services to their patients. The disease prediction system predicts diseases based on patient's symptoms and also some commonly prescribed medicines for a particular disease.

1.1. EXISTING SYSTEM

There are many prevalent systems used for disease prediction. The existing systems

only predict the diseases. The various approaches used for predicting diseases is by using Machine Algorithms such as Naïve Bayes, Decision Tree, Random Forest, k-mean algorithm. Also, one of the approaches to build a disease prediction system is by using Big Data. Prediction using traditional disease risk model usually involves Machine Learning and supervised learning

algorithm which uses training data with the labels for the training of the models.

1.2. PROPOSED SYSTEM

In the proposed system, a disease prediction model is built using a Machine Learning algorithm that is Random Forest Algorithm. Based on the symptoms that are input by the user, the disease is predicted and the drug that is most commonly prescribed by the doctor is suggested.



2. MEDICAL DATA ANALYSIS USING ML

Since the arrival of advanced computing, the doctors' still requires the technology in various possible ways like surgical representation process and x-ray photography, but the technology perceptually stayed behind. The method still requires the doctor's information and experience due to alternative factors starting from medical records to weather conditions, atmosphere, blood pressure and numerous alternative factors. The huge numbers of variables are consider as entire variables that are required to understand the complete working process itself, however no model has analyzed successfully. To tackle this drawback, Medical decision support systems must be used. This system is able to assist the doctors to make the correct decision. Medical decision support system refers to both the process of attempting to determine or identify possible diseases or disorder and the opinion reached by this process. Thediagnostic opinion in the sense, it indicates either degree of abnormality on a continuum or a kind of abnormality in a classification. It's influenced by non medical factors such as power ethics and financial incentives for patient or doctor. It can be a brief summation or an extensive formulation, even taking the form of story or metaphor. It might be a means of communication such as computer code through which it triggers payment, prescription, notification, information or advice. Indication of medical diagnostic includes knowledge of what is normal and

measuring of patient's current condition. Automated decision support systems are rule based systems that are automatically providing solutions to repetitive management problems.

Medical decision could be extremely specialized and difficult job due to alternative factors or incase of rare diseases. The alternative factors include stress; tired misdiagnosis might vary from ignorance of doctors and incomplete information. Standard algorithm may go through the entire variables like prevailing conditions history of medical records, history of family records and various factors relating to the patient records, sheer magnitude of obtainable hidden factors. Differential diagnosis methods can be used to identify the presence of an entity where multiple alternatives are possible and also refers to include the candidate alternatives. This method is needs a process of elimination or obtaining information that shrinks the probability of candidate conditions to negligible levels. It contains four steps: 1) The doctor gather all information about the patients and create a symptoms list.2) The doctor should make a list of all possible causes of symptoms.3) The doctor should prioritize the list by which is the most dangerous possible cause of symptoms put in the top of the list.4) The doctor should rule out or treat the possible causes beginning with the most urgently dangerous conditions."Rule Out" in the sense to usethe test method or other scientific method. If there will be no such diagnosis means removing the diagnosis from the list and using tests that should have distinct results,



depends on which diagnosis is correct. This can be done based on the doctor's knowledge and experience. This method is very easy to implement.

To reduce the large number of variables and find the most probable diseases by using the K-Means algorithm. This algorithm is more suitable to cluster the more number of diseases. K-Mean is one of the unsupervised learning algorithms which are used to solve the clustering problem. The main idea is to determine the k centroids, one for each cluster. Different tests performed on the patients will served as a attributes for clustering. By using this algorithm it reduce the number of iterations, boundries of clusters are well define without overlapping, to produce the accurate result for each and every diagnosis. This system uses Service oriented architecture (SOA), anyone can access with internet connections and LAMSTAR Network can be used to calculate the weight, to increase the accuracy of algorithm, overall speed test and produce the better result.

3. CONCLUSION

In this paper, algorithm used to predict the disease based on symptoms is discussed. Various symptoms are provided in the dropdown menu, out of which user selects any five of them and using algorithm the disease is predicted. The drugs that are commonly prescribed for a particular disease can also be suggested in this system.

The main aim is to predict the disease at the early stage and lead to early diagnosis. This system can also be used by doctors to avoid

confusion while predicting the disease. This system can provide assistance to doctors.

4. REFERENCES

- [1] Arthur Samuel, Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. Artificial Intelligence in Design '96. Springer, Dordrecht.
- [2] A. C. Jamgade, Prof. S. D. Zade (May 2019). Disease prediction using Machine Learning. International Research Journal of Engineering and Technology (IRJET). [Online]. Available: <https://www.irjet.net/archives/V6/i5/IRJET-V6I5977.pdf>
- [3] A. Borad (October 2018). How to develop Machine Learning applications for Business. eInfochips- An Arrow Company [Online]. Available: <https://www.einfochips.com/blog/how-to-develop-machine-learning-applications-for-business/>
- [4] Classification Algorithms – Random Forest. Tutorialspoint [Online]. Available: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm