

Human Disease Prediction using Machine Learning (CNN & RandomForest)

Raghul T¹, Prasanth P², Gowri V³

¹⁻²UG Scholar, Department of CSE, Kingston College, Vellore-59

³Asst.Professor, Department of CSE, Kingston College, Vellore-59

ABSTRACT

Disease Prediction system is based on predictive modeling predicts the disease of the user on the basis of the symptoms that user provides as an input to the system. The system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the Decision tree Classifier. Decision tree Classifier calculates the probability of the disease. With big data growth in biomedical and health care communities, accurate analysis of medical data benefits early disease detection, patient care.

The objective of this study is to investigate the use of pattern classification methods for distinguishing different types of brain tumors, such as primary gliomas from metastases, and also for grading of gliomas. A computer-assisted classification method combining conventional MRI and perfusion MRI is developed and used for differential diagnosis

Accurately predicting the survival rate of breast cancer patients is a major issue for cancer researchers. Machine learning (ML) has attracted much attention with the hope that it could provide accurate results, but its modeling methods and prediction performance remain controversial.

Adult-onset dementia disorders represent a challenge for modern medicine. Alzheimer's

disease (AD) represents the most diffused form of adult-onset dementias. Machine learning research applied to Magnetic Resonance Imaging (MRI) techniques can contribute to a faster diagnosis of AD and may contribute to predicting the evolution of the disease

Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International. Diabetes is a disease caused due to the increase level of blood glucose. This high blood glucose produces the symptoms of frequent urination, increased thirst, and increased hunger.

Pneumonia is a respiratory infection caused by bacteria or viruses; it affects many individuals, especially in developing and underdeveloped nations, where high levels of pollution, Early diagnosis of pneumonia is crucial to ensure curative treatment and increase survival rates. Chest X-ray imaging is the most frequently used method for diagnosing pneumonia.

Cardiovascular disease prediction aids practitioners in making more accurate health decisions for their patients. Early detection can aid people in making lifestyle changes and, if necessary, ensuring effective medical care. Machine learning (ML) is a plausible option for reducing and understanding heart symptoms of disease.

Key Words: Machine Learning, Symptoms based disease prediction, Python.

1. INTRODUCTION

Machine learning is programming computers to optimize a performance using example data or past data. Machine learning is the study of computer systems that learn from data and experience. Machine learning algorithm has two tracks: Training, Testing. Prediction of a disease by using patient's symptoms and history machine learning technology is striving from past decades. Machine Learning technology gives an immeasurable platform in the medical field so that healthcare issues can be resolved efficiently.

We are applying machine learning to maintained complete hospital data Machine learning technology which allows building models to get quickly analyze data and deliver results faster, with the use of machine learning technology doctors can make a good decision for patient diagnoses and treatment options, which leads to improvement of patient healthcare services. Healthcare is the most prime example of how machine learning is used in the medical field.

To improve the accuracy from massive data, the existing work will be done on unstructured and textual data. For the prediction of diseases, the existing will be done on linear, CNN, Decision Tree algorithm. The order of reference in the running text should match with the list of references at the end of the paper.

Brain tumor on clinical decisions regarding the treatment of brain neoplasms rely, in part, on MRI at various stages of the treatment process. Radiological diagnosis is based on the multi-parametric imaging profile (CT, conventional MRI, advanced MRI).

Breast cancer is the most common cancer among women in 154 countries and the main cause of cancer-related death in 103 countries. In 2018, there were approximately 2.1 million new cases of breast

cancer in women, accounting for 24.2% of the total cases, and the mortality rate was approximately 15.0% .

Alzheimer's disease (AD) is an adult-onset cognitive disorder (AOCD) which represents the sixth leading cause of mortality and the third most common disease after cardiovascular diseases and cancer.

Diabetes is the fast growing disease among the people even among the youngsters.

In understanding diabetes and how it develops, we need to understand what happens in the body without diabetes. Sugar (glucose) comes from the foods that we eat, specifically carbohydrate foods.

Pneumonia is an acute pulmonary infection that can be caused by bacteria, viruses, or fungi and infects the lungs, causing inflammation of the air sacs and pleural effusion, a condition in which the lung is filled with fluid. It accounts for more than 15% of deaths in children under the age of five years.

According to WHO data, heart disease is the leading cause of mortality globally, resulting in 17.9 million deaths annually [1]. The most behavioural risk factors for cardiovascular disease and stroke are unhealthy food, lack of physical activity, smoking, and alcohol drinking

2. OBJECTIVE

There is a need to study and make a system which will make it easy for an end-user to predict the permanent diseases without visiting a physician or doctor for a diagnosis. To detect the Various Diseases through the examining Symptoms of patient's using various methods of Machine Learning Models. To Manage Text data and Structured data is no Proper method. The Recommended system will examine both structure and unstructured data. The Predictions Accuracy will Improve using Machine Learning.

3. EXISTING SYSTEM

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 knowledge 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 granted as entire variables that are required to understand the complete working process itself, nevertheless, no model has analyzed successfully. To tackle this drawback, Medical decision support systems must be used. This system can assist the doctors to make the correct decision.

We are applying machine learning to maintained complete hospital data Machine learning technology which allows building models to get quickly analyze data and deliver results faster, with the use of machine learning technology doctors can make a big decision for patient diagnoses and treatment choices, which leads to enhancement of patient healthcare services. Healthcare is the most prime example of how machine learning is used in the medical field.

4. PROPOSED SYSTEM

This system is used to predict disease according to symptoms. This system uses decision tree classifier for evaluating the model. This system is used by end-users. The system will predict disease based on symptoms. This system uses Machine Learning Technology. For predicting diseases, the decision tree classifier algorithm is used.

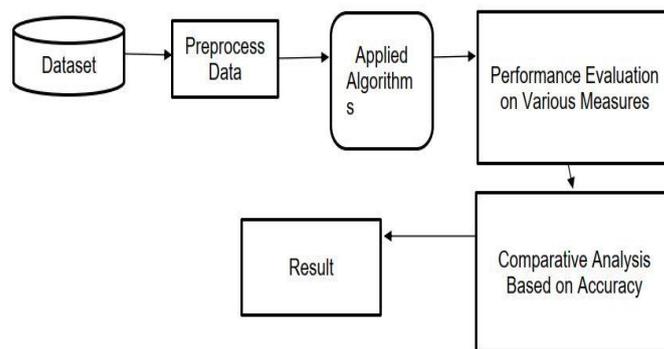
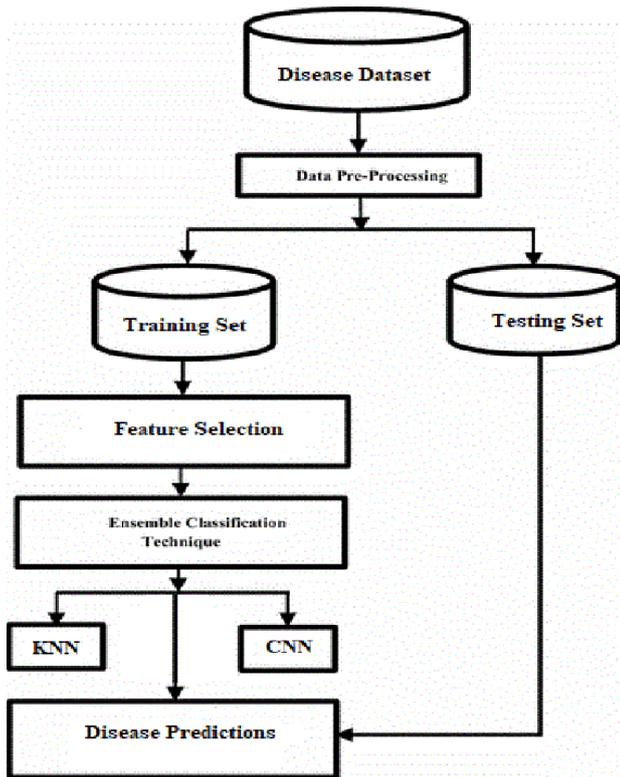


Fig.1 Proposed Model

We have named this system as 'AI THERAPIST'. This system is for those people who are always fretting about their health, for this reason, we provide some features which acknowledge them and enhance their mood too. So, there is a feature for the awareness of health 'Disease Predictor', which recognize disease according to symptoms.

5. DATASET AND MODEL DESCRIPTION

This dataset is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patients at New York Presbyterian Hospital admitted during 2004.



[Fig 2: DataSet Model]

The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of the disease, and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength of association. The method used the MedLEE natural language processing system to obtain UMLS codes for diseases and symptoms from the notes; then statistical methods based on frequencies and co-occurrences were used to obtain the associations.

6. SYSTEM ARCHITECTURE

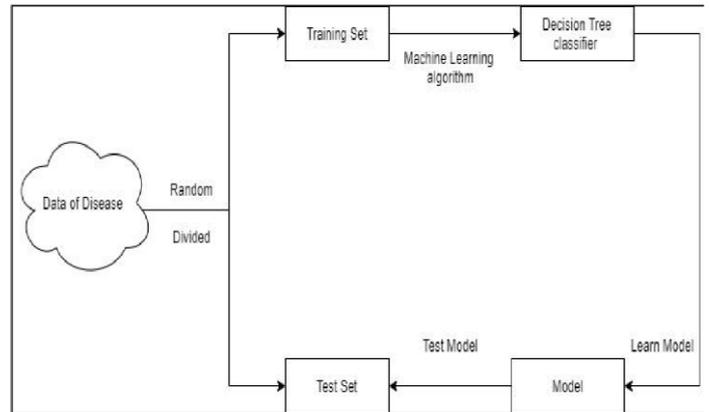


Fig -3: System Architecture

7. ALGORITHM

7.1 CONVOLUTION NEURAL NETWORKS

- A [Convolutional Neural Network \(ConvNet/CNN\)](#) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other.
- The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.
- The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex.
- A ConvNet is able to **successfully capture the Spatial and Temporal dependencies** in an image through the application of relevant filters.

- The role of the ConvNet is to reduce the images into a form that is easier to process, without losing features that are critical for getting a good prediction.

[Table 3: : Accuracy comparison table of the six classifiers]

9. CONCLUSIONS

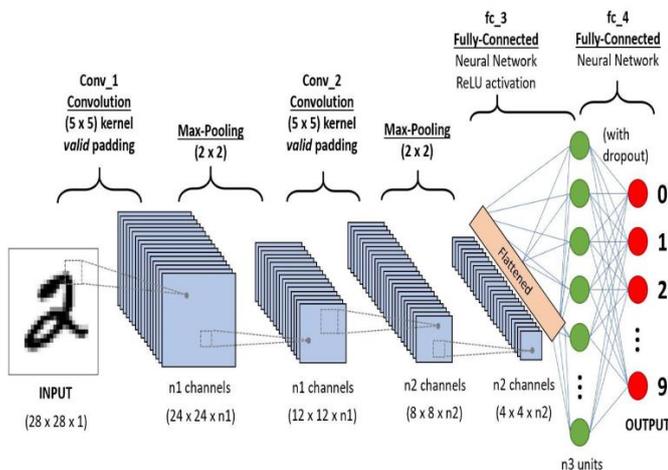
To conclude, our system is helpful to those people who are always worrying about their health and they need to know what happens with their body. Our main motto to develop this system is to know them for their health. Especially, people who are suffering from mental illness like depression, anxiety. They can come out of these problems and can live their daily lives easily.

Besides, our system provides better accuracy of disease prediction according to symptoms of the user, and also it will provide motivational thoughts and images. In the end, we can say that our system has no boundary of the user because everyone can use this system.

In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of analysis including some other machine learning algorithms.

REFERENCES

[1] Pingale, Kedar, et al. "Disease Prediction using Machine Learning." (2019). Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.



[Fig 3: :Convolution Neural Network Algorithm]

8. EVALUATING THE MODEL& RESULTS

It can be observed that the preprocessing technique, discretization has improved the performance of all the three algorithms. Though Naïve Bayes saw significant increase in accuracy due to discretization, Random Forest gave the highest accuracy for our dataset.

The results obtained from our model are summarized in the following table:

pneumonia	84.17%
heart disease	96.96%
Diabetes	66.8%
Alzheimer	76.8%
breast cancer	87.8%
brain Tumor	91.8%

- [2] Pingale, K., Surwase, S., Kulkarni, V., Sarage, S., & Karve, A. (2019). Disease Prediction using Machine Learning. *Classifiers for Diagnosis of Alzheimer Diseases*. IEEE Access, 7, 73373–73383.
- [3] Aiyasha Sadiya, Differential Diagnosis of Tuberculosis and Pneumonia using Machine Learning(2019)
- [4] S. Patel and H. Patel, "Survey of data mining techniques used in healthcare domain," *Int. J. of Inform. Sci. and Tech.*, Vol. 6, pp. 53-60, March, 2016.
- [5] Balasubramanian, Satyabhama, and Balaji Subramanian. "Symptom based disease prediction in medical system by using Kmeans algorithm." *International Journal of Advances in Computer Science and Technology* 3.
- [6] Dhenakaran, K. Rajalakshmi Dr SS. "Analysis of Data mining Prediction Techniques in Healthcare Management System." *International Journal of Advanced Research in Computer Science and Software Engineering* 5.4 (2015).
- [7] Cinarer, G., & Emiroglu, B. G. (2019). Classification of Brain Tumors by Machine Learning Algorithms. 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 .
- [8] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, 415–416, 190–198.
- [9] Ahmed, S., Choi, K. Y., Lee, J. J., Kim, B. C., Kwon, G. R., Lee, K. H., & Jung, H. Y. (2019). Ensembles of Patch-Based
- [10] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, no. 6, pp. 395–405, 2012.