International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

Diagnosis of Disease Using Machine Learning

Rajeshwari Lokhande[1], Rajesh Nasare[2]
[1]*Student, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering & Research, Nagpur, India*
[2]*Professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering & Research, Nagpur, India*

Abstract*: Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. In the field of healthcare, there is a huge amount of data present as day by day, many new diseases introduced. Dealing with huge datasets and analyzing it is beyond the scope of human capability. Hence Machine learning came into picture who has ability to process huge datasets beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care. With the help of Machine Learning, it will helpful for Doctors and Patients as well. As the cause and solution over the disease problem will be detected by Machine with help of Artificial Intelligence and algorithms, there will less chance of fraud increasing day by day in field of Healthcare. Healthcare is an important industry which offers value-based care to millions of people, while at the same time becoming top revenue earners for many countries. From playing a critical role in patient care, billing, and medical records, today technology is allowing healthcare specialists develop alternate staffing models, IP capitalization, provide smart healthcare, and reducing administrative and supply costs. Machine learning in healthcare is one such area which is seeing gradual acceptance in the healthcare industry. This paper mainly focus on the development of a system or we could say an immediate medical provision which would incorporate the symptoms collected from multisensory devices and other medical data and store them into a healthcare dataset.

***Keywords*: Disease, Machine Learning.**

1. Introduction

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect. Machine Learning (ML) is already lending a hand in diverse situations in healthcare. ML in healthcare helps to analyze thousands of different data points and suggest outcomes, provide timely risk scores, precise resource allocation, and has many other applications. Google recently developed a machine-learning algorithm to identify cancerous tumors in mammograms, and researchers in Stanford University are using deep learning to identify skin cancer.

Medical facilities need to be advanced so that better decisions for patient diagnosis and treatment options can be made. Instead of diagnosis, when a disease prediction is implemented using certain machine learning predictive algorithms then healthcare can be made smart. Some cases can occur when early diagnosis of a disease is not within reach. Hence disease prediction can be effectively implemented. As widely said “Prevention is better than cure”, prediction of diseases and epidemic outbreak would lead to an early prevention of an occurrence of a disease.

Algorithms can provide immediate benefit to disciplines with processes that are reproducible or standardized. Also, those with large image datasets, such as radiology, cardiology, and pathology, are strong candidates. Machine learning can be trained to look at images, identify abnormalities, and point to areas that need attention, thus improving the accuracy of all these processes. Long term, machine learning will benefit the family practitioner or internist at the bedside. Machine learning can offer an objective opinion to improve efficiency, reliability, and accuracy.

The intended purpose of medical algorithms is to improve and standardize decisions made in the delivery of medical care. Medical algorithms assist in standardizing selection and application of treatment regimens, with algorithm automation intended to reduce potential introduction of errors.

Broadly, there are 3 types of Machine Learning Algorithms. They are as; 1) Supervised Learning, 2) Unsupervised Learning, 3) Reinforcement Learning. With the help of these algorithms, there are many sub algorithms which has been mostly used in Healthcare.

1) Supervised Learning

This algorithm consists of a target/outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The

training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

2) Unsupervised Learning

In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

3) Reinforcement Learning

Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

2. Literature survey

Predictive algorithms and machine learning can give us a better predictive model of mortality that doctors can use to educate patients. Predictive analytics can support population health management, financial success, and better outcomes across the value-based care continuum. Instead of simply presenting information about past events to a user, predictive analytics estimate the likelihood of a future outcome based on patterns in the historical data. Predictive analytics is the process of learning from historical data in order to make predictions about the future (or any unknown). For health care, predictive analytics will enable the best decisions to be made, allowing for care to be personalized to each individual.

In the case of unstructured text data, we select the features automatically with the help of k-mean algorithm. We propose a k-mean algorithm for both structured and unstructured data.

1) K-means Algorithm

k-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster.

The k-means algorithm is a simple iterative method to partition a given dataset into a specified number of clusters, k. This algorithm has been discovered by several researchers across different disciplines. The algorithm operates on a set of d-dimensional vectors, $D = \{xi \mid i = 1. . . N\}$, where $xi \in Rd$ denotes the i^{th} data point. The algorithm is initialized by picking k points in Rd as the initial k cluster. Techniques for selecting these initial seeds include sampling at random from the dataset, setting them as the solution of clustering a small subset of the data or perturbing the global mean of the data k times.

3. Existing system

Prediction using traditional disease risk model usually involves a machine learning and supervised learning algorithm which uses training data with the labels for the training of the models. High-risk and Low-risk patient classification is done in groups test sets. But these models are only valuable in clinical situations and are widely studied. A system for sustainable health monitoring using smart clothing by Chen et.al. He thoroughly studied heterogeneous systems and was able to achieve the best results for cost minimization on the tree and simple path cases for heterogeneous systems.

The information of patient's statistics, test results, and disease history is recorded in EHR which enables to identify potential data-centric solutions which reduce the cost of medical case studies. Bates et al. propose six applications of big data in the healthcare field. Existing systems can predict the diseases but not the subtype of diseases. It fails to predict the condition of people. The predictions of diseases have been non-specific and indefinite.

4. Proposed system

In this paper, we have combined the structure and unstructured data in healthcare fields that let us assess the risk of disease. The approach of the latent factor model for reconstructing the missing data in medical records which are collected from the hospital. And by using statistical knowledge, we could determine the major chronic diseases in a particular region and in particular community. To handle structured data, we consult hospital experts to know useful features. Below are the algorithms we can use to analyze the data and process the data for Healthcare system. As more data is available, we have better information to provide patients.

5. System architecture

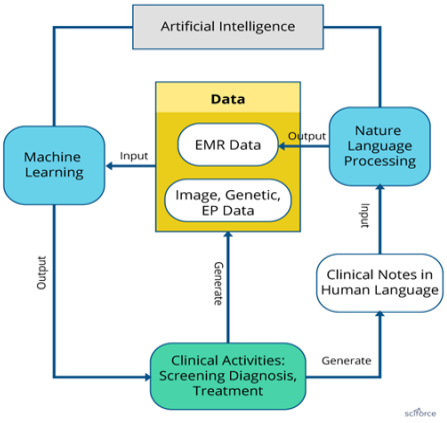

Fig. 1. Machine learning and natural language processing in healthcare

6. Conclusion

With the proposed system, higher accuracy can be achieved. We not only use structured data, but also the text data of the patient based on the proposed k-mean algorithm. To find that

out, we combine both data, and the accuracy rate can be reached up to 95%. None of the existing system and work is focused on using both the data types in the field of medical big data analytics. We propose a K-Mean clustering algorithm for both structured and unstructured data. The disease risk model is obtained by combining both structured and unstructured features.

Acknowledgement

I express my sincere gratitude towards my guide of Prof. Rajesh Nasare for their constant help, encouragement and inspiration throughout the project work. Also I would like to thank the Head of Computer Science and Engineering Department Prof. Tirpude for him valuable guidance, ability to motive me and even willingness to solve difficulty made it possible to make my project unique and made task easier. My sincere thanks to Principal, Dr. Kshirsagar for providing me necessary facility to carry out the work.

References

[1] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014.
[2] K. R. Lakshmi, Y. Nagesh and M. VeeraKrishna, "Performance comparison of three data mining techniques for predicting kidney disease survivability", International Journal of Advances in Engineering & Technology, March 2014.
[3] Mr. Chala Beyene, Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.
[4] Boshra Brahmi, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015, pp.164-168.
[5] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, "Incorporating temporal ehr data in predictive models for risk stratification of renal function deterioration," Journal of biomedical informatics, vol. 53, pp. 220–228, 2015.
[6] 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.