A SURVEY ON MACHINE LEARNING APPLICATIONS IN HEALTHCARE

LAKSHMI VIVEKA KESANAPALLI[1] and S. RAO CHINTALAPUDI[2]

[1]Assistant Professor

[2]Associate Professor

Department of CSE

Pragati Engineering College

India

E-mail: lakshmiviveka.k@gmail.com

      srao.chintalapudi@gmail.com

Abstract

Now-a-days, the technology has become a part of our lives and we are living with the technology. Machine Learning technology has become a part to understand and know what is happening with the person's health and we need to analyze one's own health. Healthcare is the field where we can apply machine learning and big data applications on it. Prediction and detection of one's health has become a difficult for the health care practitioners. So, predicting heart disease, cancer, neuro and cardiac at the initial stages will be helpful to the people so that they can take the necessary steps/action before the situation gets severe. Over the years the machine learning algorithms are showing better observations in making decisions and prediction from a huge set of information provided by the healthcare industries. ML cannot be executed for all the problems in reality. A number of the supervised machine learning algorithms utilized in the predictions are Artificial Neural Networks, Decision Trees, Random Forest, SVM, Navie Bayes, K Nearest Neighbor algorithms.

I. Introduction

Healthcare is the largest domain which produces a great deal of information for the researchers to implement something new from data. In this article we will identify the machine learning methodology which helps the people to know the different things they have to follow to take care of a far better health mechanism. Here we will be following only the quality and

traditional techniques to spot the human problems.

Machine learning as a part of AI, helps the computers to think like humans and take their own decisions without the intervention of the people. Moreover the algorithms give better prediction and performance on a specific problem. Machine learning is broadly divided into different types as shown in the following Figure 1:
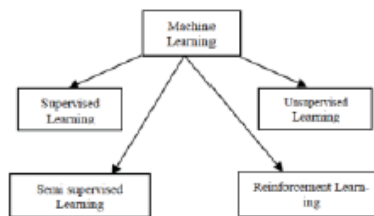

Figure 1. Different type of Machine learning.

Algorithms in Machine Learning:

Artificial Neural Networks (ANN): ANN is mostly implemented for computational purposes; the major principle of this model is to take care of the responsibility better than the conventional model. This is comparable to the shape of neurons in the brain. A single layer neural community is recognized as a perceptron which offers us a solitary output.

Support Vector Machine (SVM): It is a machine learning algorithm. In supervised support vector machine algorithms, if we consider any labeled data for training, which produce a classifier that divides the labeled information into various classes.

Decision tree (DT): It is one of the supervised algorithms which can be implemented for both regression and classification techniques. In this, the information will be split based on the parameters. A tree contains leaves and nodes. The nodes will split the data and at leaves we get the outcome. The decision trees can be implemented in two ways they are the classification and regression trees.

Random forest (RF): It is one of the supervised learning algorithms used for regression and classification. However, primarily it's for classification functions. The name itself indicates that it's a group of trees equally in an exceedingly random forest algorithmic rule square measure

going to have trees and the trees are known as decision trees.

Navie Bayes (NB): It is one of the supervised machine learning classification algorithm. Initially it is used for classifying the text data. It handles the datasets with high spatiality.

The rest of the paper was organized as follows: In section 2, related works are discussed. Section 3 deals with data collection. Section 4 is about research challenges and section5 is about conclusion and future scope.

II. Related Work

Awais Nimat et al. [1] proposed a specialist framework supported by two support vector machines (SVM) to anticipate the heart condition efficiently. These two SVM's have their motivation; initially, one is utilized to get rid of the unnecessary features, and hence the other is utilized for prediction. Besides, they need the HGSA to optimize the two strategies. By utilizing this model, they accomplished around 3.3% preferably accurate than the traditional model.

Mehtaj Banu H [3] the researcher studied distinct techniques in machine learning like Supervised, unsupervised and reinforcement and also furthermore investigation on UCI dataset database and finalize that *K*-Nearest Neighbour and Support Vector Machine algorithms have proved execution and precision for the prediction of a specific disease.

Parthiban and Srivatsa [4] the author analyzed the condition of heart in diabetic patients by victimisation strategies of machine learning. Algorithms of Naive Bayes and Support Vector. Datasets of around 490 patients is utilized that are collected from Research Institute of Chennai. Patients that have the sickness are 141 and malady is absent in around 348 patients. By utilizing the Naive Bayes Algorithm 74% of precision is acquired. SVM provide the absolute accuracy of about 93.60.

Sarwar and Sharma [5] have recommended the work on Naive Bayes to anticipate diabetes Type-2. Diabetes is categorized into 3 types they are Type-1 diabetes, Type-2 diabetes and gestational diabetes. Type-2 diabetes originates from the extension of Insulin obstruction. The information comprises of 411 cases and for the motivation behind assortment; information

are gathered from different areas of society in India. Around 94% of right forecasting is accomplished by Naive Bayes.

Fathima and Manimeglai [6] attempted to anticipate the disease Arbovirus-Dengue. Information processing is utilized by the analysts are SVM. Dataset for the examination is derived from King Institute of Preventive Medicine and studies of numerous emergency clinics and research centres in and around Chennai and Tirunelveli from India. It includes of 29 attributes and 5000 sample data and this was analyzed using the R version 2.12.2. Accuracy obtained using SVM is around 0.9041.

III. Data Collection

Machine learning is currently deployed widely across the various health sectors owing to its ability to form the real time predictions and draw the unnoticed insights from the given voluminous and unstructured datasets. Here are few repositories listed where we can get the data sets related to health care.

• WHO (World Health Organization): Its open source data contains classes which include child nutrition, neglected diseases, risk factors pertaining to certain diseases among others.

• OGD Platform India: This web site consists of all the information collected from the Indian health agencies and different entities.

• Open fMRI: It is a project dedicated for sharing the free and open source datasets related to imaging.

• Data.gov: This site consists of all the information collected from the various primary healthcare centres, community health centres of various district hospitals and mobile medical units from various states and union territories.

There are different other sites where we can get the information related to the particular research works being handled by various exploration research scholars. Depending on the area of work all the data that is available is the data collected from various sources.

IV. Research Challenges

The Indian healthcare scenario witnesses the rapid pace of change currently happening in the glorious tradition of the public health. There are many challenges being faced in the healthcare sector in India and various parts of the nation. We have many challenges but listing only few in this paper:

• Awareness: How individuals know about the significant issues with respect to their own wellbeing? Studies on mindfulness are numerous and differing, yet satisfactory information and mindfulness seem to cut over the life expectancy in our nation.

• Access or the lack of it: "The right or opportunity to use or benefit from healthcare" is the definition of access given by the Oxford dictionary in terms of healthcare. Physical reach is one of the fundamental determinants of access, [7].

• Absence or the human power crisis in healthcare: According to the study India has approximately 20 health examiners for every 10,000 population, with medical aid comprising 31% of the manpower, medical attendants and 11% drug specialists. [8].

• Affordability or the cost of healthcare: It is quite simple, how costly the medical services are in India. Practically 75% of the healthcare consumption comes from the households units. [9].

Conclusion and Future Scope

This article gives us the essential plan of the previously published paper of identification and assurance of different infections/diseases based on various learning algorithms. With this review and study it is clearly found and discovered that some machine learning algorithms like Decision tree, Random forest, Navie Bayes and ANN give the higher accuracy in detecting and predicting various diseases. And additionally the paper gives a review on various types of machine learning techniques utilized by totally different authors and every machine learning techniques has some smart and unhealthy outcome based upon the datasets and feature selection etc. With the review we have a tendency to know that the absoluteness and execution

will be improved by victimisation of totally various combination or hybrid machine learning algorithms and in future we will additionally implement on a lot of variables that facilitate to urge higher execution than the current technique.

References

[1]   Archana Singh and Rakesh Kumar, Heart Disease Prediction Using Machine Learning Algorithms, International Conference on Electrical and Electronics Engineering (ICE3-2020).

[2]   L. Ali, An Optimized Stacked Support Vector Machines Based Expert System for the Effective Predict ion of Heart Failure, IEEE Access 7, pp. 54007-54014, 2019,doi: 10.1109 / ACCESS. 2019. 2909969.

[3]   H. Mehtaj Banu, Liver Disease Prediction using Machine-Learning Algorithms, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 895, Volume-8 Issue-6, August (2019).

[4]   G. Parthiban and S. K. Srivatsa, Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients. International Journal of Applied Information Systems (IJAIS) 3 (2012), 25-30.

[5]   A. Sarwar and V. Sharma, Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2. Special Issue of International Journal of Computer Applications (0975-8887) on Issues and Challenges in Networking, Intelligence and Computing Technologies-ICNICT 3 (2012), 14-16.

[6]   A. S. Fathima and D. Manimeglai, Predictive Analysis for the Arbovirus Dengue using SVM Classification, International Journal of Engineering and Technology 2 (2012), 521-527.

[7]   M. Rao, K. D. Rao, A. K. Shiva Kumar, M. Chatterjee and T. Sundararaman, Human resources for health in India, The Lancet. 377 (2011), 587-598.

[8]   Y. Balarajan, S. Selvaraj and S. V. Subramanian, Health care and equity in India, Lancet. 377 (2011), 505-515. [PMC free article] [PubMed]

[9]   Ayeldeen, Heba, Olfat Shaker, Ghada Ayeldeen and Khaled M. Anwar, Prediction of liver fibrosis stages by machine learning model: A decision tree approach, Third World Conference on Complex Systems (WCCS), IEEE (2015), pp. 1-6.

[10]   Data Mining Techniques for optimization of liver disease classifi-cation, SadiyahNoor Novita Alfisahrin, Teddy Mantoro Electron-ic ISBN: 978-1-4799-2758-6 DOI: 10.1109/ACSAT.2013.81-IEEE

[11]   N. H. Farhat, Photonit neural networks and learning machines the role of electron-trapping materials, IEEE Expert. Syst. Their Appl., 7(5) (1992), 63-72 doi: 10.1109/64.163674.