Chapter 11
Demystifying Disease Identification and Diagnosis Using Machine Learning Classification Algorithms: Demystify Medical Data With Machine Learning Algorithms

Ashok Suragala
https://orcid.org/0000-0002-8431-2479
JNTU Vizianagaram, India

PapaRao A. V.
JNTU Vizianagaram, India

ABSTRACT

The exponential surge in healthcare data is providing new opportunities to discover meaningful data-driven characteristics and patterns of diseases. Machine learning and deep learning models have been employed for many computational phenotyping and healthcare prediction tasks. However, machine learning models are crucial for wide adaption in medical research and clinical decision making. In this chapter, the authors introduce demystifying diseases identification and diagnosis of various disease using machine learning algorithms like logistic regression, naive Bayes, decision tree, MLP classifier, random forest in order to cure liver disease, hepatitis disease, and diabetes mellitus. This work leverages the initial discoveries made through data exploration, statistical analysis of data to reduce the number of explanatory variables, which led to identifying the statistically significant attributes, mapping those significant attributes to the response, and building classification techniques to predict the outcome and demystify clinical diseases.

DOI: 10.4018/978-1-5225-9643-1.ch011
INTRODUCTION

There is a growing need for accurate classification of disease as if the disease is detected early, then it becomes easy to cure rather than in the future stages.

The purpose of a medical diagnostic machine learning model research is to find better ways to demystifying diseases identification and diagnosis in patients. The result of diagnostic model research is a diagnostic machine learning model that allows telling with reasonable certainty whether or not the patient has a certain disease, based on a number of variables.

Medical diagnostic model research improves diagnosis by avoiding the need for surgery, thereby making it easier to diagnose disease earlier which substantially improves survival rates. Now how are the diagnostic models calculated? They are calculated using machine learning technology called machine learning classification models.

The idea with this technique is to collect a lot of data from patients with and without the diseases. Clinicians collect a lot of other information about the patients; information that can be obtained using non-invasive techniques. This can be demographic data, information obtained from ultrasound or other images or any other information that medical researchers suspect might be linked to the true diagnosis. The information obtained non-invasively is the inputs, while the gold standard diagnosis is the output. Both are fed to a classifier which will calculate a model to predict the output from the inputs. This, in the case of diagnostic mobile research, means predicting a patient’s diagnosis based on information that was obtained from the patient by non-invasive means. Now, diagnostic model research isn’t just a case of running patient data through a classifier. It requires stepping through an entire workflow.

Evaluation of clinical data from different patient cohorts and decision making are significant factors in diagnosis and making prediction of diseases using machine learning classifiers. For this purpose, we first organize this chapter by defining the patient population from different cohorts, identifying phenotypes, perform statistical analysis, apply different feature selection techniques to identify statistically significant features, applying machine learning classification algorithms and validate the model.

In this chapter, we organize our work by defining a study, continue with data collection & data analysis, identifying computational phenotypes, apply classifiers (machine learning models) with significant features (phenotypes) and conclude with validation and model dissemination to predict the outcome and demystify clinical diseases and a tool for valuable clinical decision making.

BACKGROUND

Machine Learning Algorithms are widely used in diagnosis of different clinical diseases. In recent years researchers adopted different machine learning classification algorithms for disease diagnosis and prediction.

Machine Learning Algorithms classified as supervised, Un-supervised and ensembles algorithms aims to focus on making predictions using computers in social media, financial, Medical, Entertainment and building product, movie and song recommendation engine domains. (Jafar Alzubi, Anand Nayyar & Akshi Kumar, 2018). In this chapter, we use various supervised learning algorithms like classification, regression and Ensemble techniques (Jafar Alzubi, Anand Nayyar & Akshi Kumar, 2018) such as Logistic Regression, decision tree, Naïve bayes and Random Forest Algorithms.