Chapter 21
A Study on the Behavior of Machine Learning Algorithms in Healthcare Datasets

Richa Singh
Bennett University, India

Shivani Gupta
Vellore Institute of Technology, Chennai, India

Ninad Pravin Sabnis
Vellore Institute of Technology, Chennai, India

ABSTRACT

In this work, the authors have provided a study of various machine learning algorithms that makes use of datasets for diseases related to three different organs. Apart from the three separate organs, namely heart, kidney, and liver, this work also monitors the signs of two specific diseases: diabetes and breast cancer. It is possible that the number of patients that require constant attention may be far greater than the number of available doctors, and hence, the requirement of a system may arise wherein a machine may predict the severity of the person's condition based on his vitals and summon the required medical expertise. In this case, the prediction system might prove to be useful. Machine learning approaches have proven their capability in the prediction of disease diagnosis. The machine learning models are evaluated on datasets that include diabetes and breast cancer dataset and kidney, liver, and heart datasets to compare the accuracy of the models.

INTRODUCTION

Describe the general perspective of the chapter. End by specifically stating the objectives of the chapter. Today prioritization of patients based on their present health condition is vastly required by the doctors to provide their services to the person who is at high risk. Nowadays, we see around the world except

DOI: 10.4018/978-1-6684-6697-1.ch021

for a few countries, most of the country's health infrastructure is very poor. Today where countries like the USA, Japan and Germany spend more than 10% of the GDP on health, India spends only 2.5% to 3% of the GDP on health. In fact, there are very few nurses and doctors available per thousand people in India (Reinhardt et al., 2004). Thus the development of a system having automated Artificial Intelligence systems will not only be of great help to our healthcare system but will add the much desperately needed automated health systems for our medical professionals, we have already seen the impact of these medical professionals being overloaded in the recently passed COVID-19 pandemic resulting in the death of thousands without any medical care being afforded to them and hence the development of AI technologies in the field of healthcare is absolutely necessary (Guo & Li, 2018). Diabetes and breast cancer risks can also be checked as earlier a disease is diagnosed more are the chances of it being treated successfully leading to complete healing especially in the case of cancer which can turn out to be very painful and even lethal. Patients are in a dangerous situation of early death owing to the unexpected emergence of numerous illnesses and seizures because of a lack of good patient nutrition and quick corrective modifications.

The health checkup system is a very critical examining as well as monitoring system that is used to detect the chances of a patient having diseases based on the various inputs entered by the user. In the case of heart for example, this system is utilized for observing physiological signs including Electrocardiograph (ECG), Respiration, Invasive and Non-Invasive blood pressure, oxygen saturation in Human Blood (SpO2), body temperature, heart rate and so forth during treatment (Suh & Evangelista, 2010). This pattern of recognition similarly can be extended to include liver and kidneys. A few examples related to the liver are hepatitis, jaundice, liver cirrhosis and various other alcohol related diseases and a few related to the kidney are kidney stones, urinary tract infections and other chronic kidney diseases (Lee et al., 2018; Suh et al., 2010; Yasodhara et al., 2021).

There are few challenges that come across the way of handling the dataset of the diseases. First is the validation which is fundamental for elements working in various areas to build up trust for doing computerized wellbeing exchanges. Another perspective that requires consideration as a predominant advanced medical problem for countries is the absence of interoperability between countries aiming to coordinate on computerized wellbeing. This inadequacy is because of not just the monetary restrictions or shortage of IT abilities, yet additionally the absence of strategy for worldwide collaboration among countries on the trading of touchy clinical information, which would work with telemedicine and provisioning of top-notch clinical consideration distantly. This would generate reliable datasets for producing greater accuracies. The second challenge is related to authenticity and operability issues. The current system will have to increase scaling and load balancing capabilities once multiple users start logging in and accessing the application at the same time. Moreover, the accounts of these users will have to be validated and our servers protected from attacks. The third challenge relates to the collection and management of data. Data can be collected and stored in our systems through various means. It can be manually entered by the diseased patients or their relatives; it can also be automatically recorded through sensors if they are connected to the patient. Other metadata related to the patient can be stored in databases and be retrieved whenever required (Sai et al., 2023; Tyagi et al., 2021).

Overview of Various Machine Learning Algorithms

As per Andrew Ng (Tyagi et al., 2021), famous writer of Machine Learning, the concept of Deep Learning and Artificial Intelligence has been around for many more years than one can think of. The theories