ABSTRACT

The proposed work aims to develop a real-time prediction system for health issues based on big medical data processing on the cloud. In the proposed scalable system, the medical parameters are sent to Apache Spark to extract the attributes from the data and to apply the proposed machine learning algorithm aiming to predict the healthcare risks and send them as alerts and recommendations to the users and the healthcare providers as well. The purpose of this paper is to evaluate the impact of applying machine learning algorithms using electronic health records. The proposed work aimed to provide an effective recommendation system using streaming medical data, historical data on the user profile, and knowledge database to provide users with the best recommendations and alerts in realtime according to the sensors measurements. The proposed system of prediction could offer high accuracy in comparison with literature work with the predictability of 90.6 for heart disease. The methodology of this research is applying parameterization for parameters on SVM to make the possibility of prediction is higher using the most effective features.

INTRODUCTION

- There are currently several researchers who have developed machine learning approaches for the early detection of chronic diseases.
- Most health education initiatives require the prevention of the disease and early identification of diseases.

- Big data analysis in healthcare is very convenient and useful to use technology to produce medical data with spark and machine learning algorithms to predict health problems.
- The world is currently experiencing the era of information technology, which relies on modern satellite communication systems and information systems related to electronic computers, and there is no doubt that many of the rapid changes that have produced many challenges for these organizations so that they must move towards the application of information technology to enhance its position.
- Computerized health information systems have become a major issue of concern to all managers in health organizations.
- The hospitals and specialized centers and health centers of the General Administration of Primary Care are the main providers of health services to the public of patients and citizens.

Related Work

Many researchers are developing new approaches and techniques for using Big Data machine learning algorithms to prevent health issues at the very early stages of diseases. Spark processes large quantities of data very easily because it uses the structure of parallel computing to handle the data streams from big data sources.

Researchers are using the power of analytics, pattern recognition, neurocomputing, data processing, machine learning, deep learning, artificial intelligence (AI), databases, knowledge discovery, and exploration of information to achieve the meaning of the data and make it understandable.

Healthcare prediction system
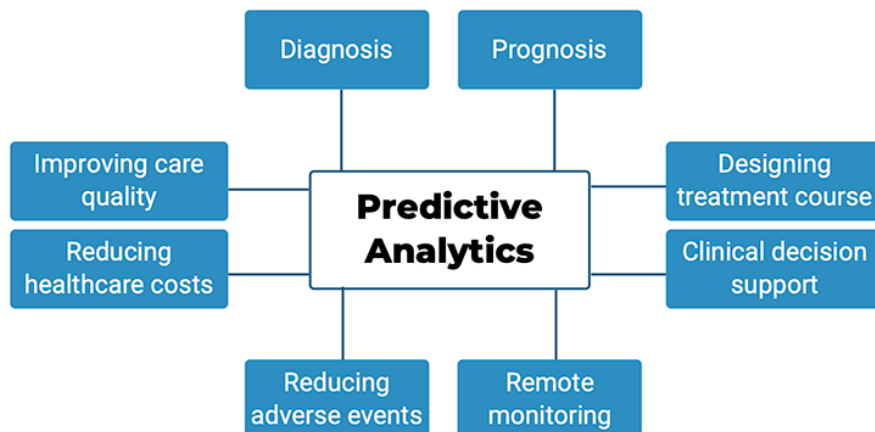
The research work provides a big data analytics architecture that is based on healthcare data from healthcare providers and wearable devices. The health tracking devices can provide interesting health data over time. The main target of the system is analyzing streaming data from different sources such as smart clothes, wearables, smart rings, smartwatches, and other advanced medical devices.

Implementation and experimental testing

The target of the testing is to show the efficiency of the system using Microsoft Azure. The system is being trained using a Microsoft cloud framework to make use of the power of the cloud architecture in processing the streaming data. The system was evaluated by using the UCI dataset which is an online repository of heart disease, and it contains a wide database network. The proposed approach to completing this work is initiated by loading the dataset collection of UCI data [26]. The system used a heart disease dataset from the UCI Machine Learning repository to train a model for predicting heart disease using SVM. We developed a heart disease classification model using the following characteristics as input features: sex, type of chest pain, age, blood pressure resting, serum cholesterol, max heart rate, EEG resting, exercise-induced depression angina, blood sugar rate, training style, and some of the major vessels.

The evaluation module provides metrics of accuracy, precision, recall, and F1 which collectively show how good the model is based on the evaluation module test. The distribution of the likelihood of the correct prediction placed in 10 bins. For example, there are thirteen appearances of TP and one appearance of TN.

Discussion

The need for applying machine learning algorithms on available large medical data rather than traditional methods became one of the most driving research topics in healthcare. The article presented mainly an efficient architecture to work with streaming data from different wearable devices in the healthcare systems. The proposed healthcare system employs Microsoft Azure instead of using a standalone server to work with streaming data.

Conclusion

The program approach offered a broad computer structure that manages EHR computer focused on streaming data from connected medical devices and patient history for patients. This paper provides an important approach for the health prediction method utilizing the EHR platform and consumer social health profile information exploration methods. The proposed framework also offers an efficient way to link Spark and Microsoft Azure-based streaming data from wearable devices to forecast diseases. Future research is to apply this theoretical program to its data utilizing deep learning techniques. The next phase of our proposed research is to include a cloud framework that supports the dataset of records of the disease.