Journal of Theoretical and Applied Information Technology
15th June 2020. Vol.98. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIG DATA ANALYTICS IN HEART DISEASES PREDICTION

[1] AHMED ISMAIL, [1] SAMIR ABDLERAZEK, [3] I. M. EL-HENAWY

[1]Information Systems Dep., Faculty of Computers and Information, Mansoura University, Mansoura, Egypt.

[3] department of Computer science, faculty of Computers and Information, Zagazig University, Zagazig, Egypt.

E-mail: [1]a.ebada@outlook.com

ABSTRACT

The healthcare data can be employed to develop a health prediction system that can improve in heart disease prevention. Big data on health care, including patient records, clinical notes, diagnosis, parents and family past ailments, hospitals, and scan results can aid in the phase of disease identification and prediction. The emerging machine learning method offers an important framework for forecasting cardiac diseases. An advanced Support Vector Machine (SVM) classifier was used by the program to conduct parameter tuning to improve classification accuracy and performance. 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 real-time 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.

Keywords: *Data Analytics, Health Parameters, Medical Data, Prediction Science, Prevention, Big Data Analysis, Apache Spark, Smart Wearables.*

1. INTRODUCTION

There are currently several researchers who have developed machine learning approaches for the early detection of chronic diseases. The wearable devices deliver lightweight, reliable, low-cost, and light health monitoring systems in healthcare areas. Thanks to several medical awareness initiatives, the constant tracking of changes in the body using smart sensors became a lifestyle. Most health education initiatives require the prevention of the disease and early identification of diseases [1]. 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 [2]. It will help people get health hazards information sooner and get health problems warnings. It can also help doctors track patients in smartphone applications. It also facilitates the treatments based on an advanced diagnosis for human diseases by providing recommendation systems-based machine learning methods.

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[3].

As competition and the challenges facing today's organizations have become increasingly intense, the role of information in governance needs to be taken into consideration, prompting organizations to accelerate the development of management information systems and computer-based information systems. `The use of modern information technologies, which have made a remarkable change in management in various functions, is no longer the possession of an administrative information system is the decisive factor in the success of organizations and achieve the competitive advantage only, but there are many aspects to be taken into account the most important indicators of the success of the MIS. This aspect is reflected in the behavior and feelings of the beneficiary at a time when the beneficiary has a greater role in the success of the design and implementation of the information system. Information systems are generally concerned with providing information for use in the organization, and today it took a new organized dimension where it became treated as a resource of organizational resources.

Computerized health information systems have become a major issue of concern to all managers in health organizations. Health information systems and technology have brought distances far beyond human reach, and have enabled the storage of digital, text, audio, and image data. This helps to provide information that is very important to decision-makers. Health information systems play an important role in the practical life of health workers. They support them in identifying their training needs, continuing education and scientific research, Life, and the most important applications of health information systems are hospital information management systems, telemedicine, and medical and administrative decision support systems [4].

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. The General Directorate of Hospitals is responsible for organizing the work of hospitals and monitoring the performance and providing the necessary means while regulating the relationship with other departments in the Ministry of Health. Parallel Management aimed at integrating work among all hospitals, distributing competencies and manpower, and providing service at a decent level to all citizens by the most accessible means (www.moh.gov.ps).

The paper offers brief details on the major healthcare data in Sect. 2. Section 3 describes the conceptual framework for evaluating data on healthcare. Section 4 discusses the results suggested for the program. It addresses the findings and possible research in section 5.

2. 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 [3].

There are many different types of research on how to get the best out of health care results. The challenge nowadays is to evaluate healthcare issues to prevent diseases. In fig 1, there are many scientists and researchers in different areas who worked on the value of data availability nowadays [2]. 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.

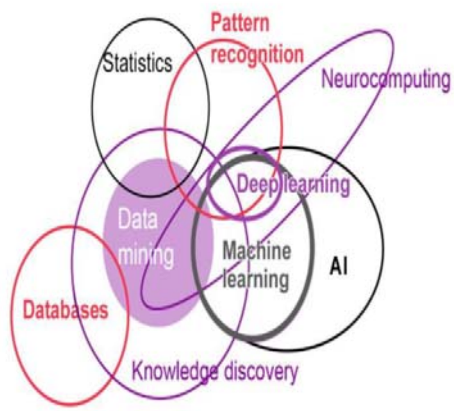

Fig 1: Sectors Of Analysing Big Data.

Apache Spark streaming to process the medical big data on the running time from the wearable devices that includes data from sensors and devices [4]. As shown in Fig 2 the power of applying Apache Spark is that it can process data from streams from different sources such as the Internet of Things(IoT) sources to keep analyzing data and provide insights of the data[5].

Journal of Theoretical and Applied Information Technology
15th June 2020. Vol.98. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

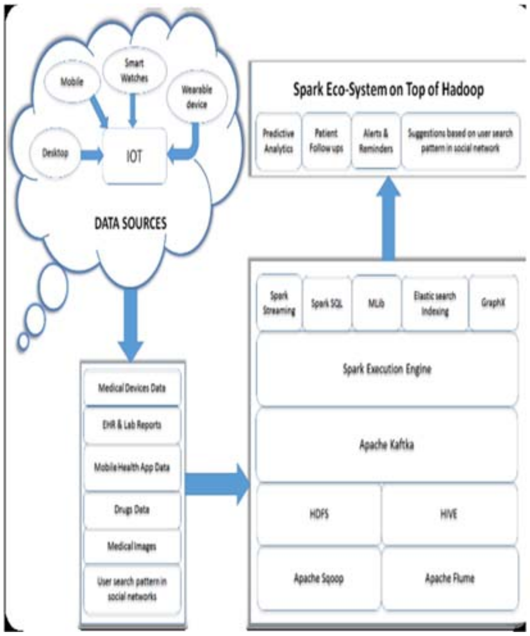

Fig 2: The Apache Spark Architecture.

Electronic medical records (EMRs) are a set of medical data to be processed to provide useful knowledge about health care [6]. Using new technology for saving healthcare data on the cloud-based Hadoop Distributed File System (HDFS) to use a distribution technique to store it in HDFS. HDFS is a reliable, fast-access technique [7]. The spark which is an analysis tool facilitates data processing streaming and batch processing [8]. Apache Spark offers machine learning techniques for different streaming data types to analyze streaming data as streams. For high velocity and efficiency, it can apply various algorithms. Apache Spark also facilitates data processing using the Hive.

The Major Medical Data must always be analyzed after they have been obtained and processed to obtain useful data. The method of analyzing data consists of four groups, as shown in fig 3 [9]. The first category is a concise study of what changes in data using the data history. Then analyzing the data to get the meaning of it to determine the issues. Another type is a statistical analysis which is a forward-looking forecast or prediction. Ultimately, prescriptive analytics is to use algorithms such as decision trees to respond to various circumstances based on probabilities [10].

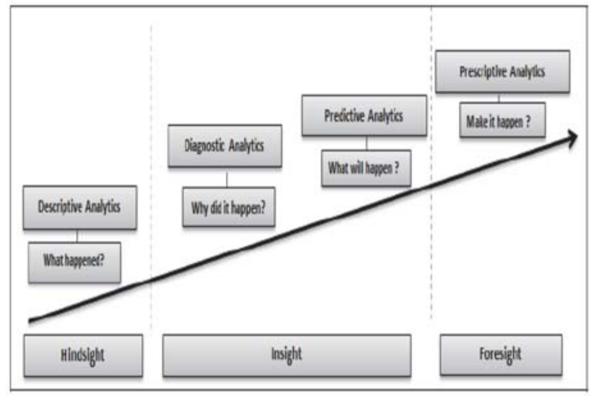

Fig 3: Analysing Big Data Processes.

Healthcare providers needed to use the big data from the health statuses of the users to analyze the data using parallel computing to create and validate personalized applications for medical data [11,12]. Apache streaming includes structures capable of handling the difficulty of parallelization. Apache Spark is a software that processes batch or processing which is a handy solution to follow the streaming of data. The key benefit of that framework is that at runtime it can process the MapReduce data in the memory and there is no need to spend more time than using local storage to read and write data from memory to local storage. Using Analysis of streaming data using Spark technologies can be used to predict diseases to produce characteristics from medical and physiological data.

Many research ventures have attempted to use emerging methods to monitor and predict diseases using medical wearables. Medical wearable devices are devices with certain physical sensors, such as heart rate sensors and sensors for blood pressure. These devices may provide sensor warnings to patients and the elderly or any user event to warn them of irregular reading or send them overall health information about the state of health [13]. There are several medical sensors available that can provide wearable device-based health monitoring applications.

The Features Reduction (FR) is a methodology of selection that selects features with a high-class correlation (output). FR is also known as Dimensionality Reduction since it often tries to lower the number dimensions. The number of functions in FR impacts the efficiency of the classification [15].

When a dataset is provided as data mining input with many classification features, the first target is cleaning the dataset and improving the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dataset [16] which has a big impact on machine learning methods performance[17]. In pattern classification, the use of the SVM [15] became popular because of the usability of many supervised machine learning solutions. SVM detects illnesses and can also identify high success rates datasets [18].

A ReliefF example of feature selection, which can operate well even on the missing data or issues of data competency [18]. There are much of literature work applied classifications using SVM to build diagnostic models for diseases [19], we present here a few.

Zhang et al. [20] developed a model using SVM for a diabetes diagnosis using standardized tongue images. To boost the classification of SVM. They used the SVM definition for a diabetes diagnosis using generic images of the tongue. To boost the system.

In [21], the authors proposed a system based on the optimization of parameters of SVM and added weight to boost SVM output features. They introduced a model that was a hybrid system that optimized SVM by using SVM, genetic algorithms (GAs) [22,23].

The adaptive SVM classification system for the treatment of different diseases was introduced in [24]. The authors have presented a diagnostic method for disease diagnosis. They applied for their work on disease datasets such as diabetic and breast cancer. The program used principles of 'if-then' for decision making. The proposed system results provided 100 percent classification accuracy for datasets on and diabetes and breast cancer.

In [25], a genetic algorithm-based diagnostic method with feature selection for lung diseases were presented. The proposed method was a hybrid system of the k-nearest neighbors KNN algorithms with SVM and back-propagation classification measures. The method achieved 99.5 percent with high predictive performance.

The proposed system is a bit different than the discussed solutions that the solution is based on a wearable smart ring that monitors different health parameters based on multi-sensor. The proposed system is a healthcare system based on the Microsoft Azure cloud system to get streaming data from wearable devices to predict breast cancer disease. The system is a multisensory system that monitors HR, blood pressure, brain activities, muscle activities. The monitoring data is sent to the smartphones wirelessly using the ZigBee [18] radio standard. The smartphone sends the data to the cloud user profile which has all medical and historical data of the patient. Finally, the machine learning model is running on the cloud system to recommend the healthcare provider about heart disease predictions.

3: The proposed healthcare prediction system

Many healthcare companies employ medical big data to make suggestions and recommendations in the healthcare area. Apache Spark can process data quickly in dealing with big data of unstructured and structured healthcare data sets because it provides memory computations to work on analyzing the data at Memory which is very fast than reading and writing data from Hard disks.

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. The proposed methodology is using cloud structure as shown in fig 4 to process medical data and to try to support doctors to decide in the diagnosis of heart diseases using SVM. This methodology is an efficient system because the proposed system applies a selection method based on the main features of the given dataset to classify the heart disease from the user profile in the cloud.

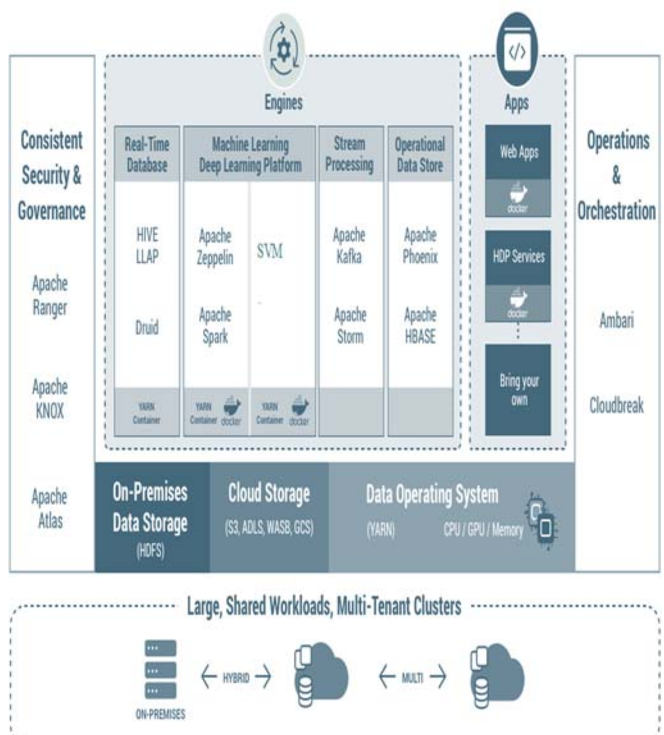

Fig 4: Big Data Analysis On The Cloud.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The proposed framework used clusters from MapReduce. The system introduced a platform for storing heterogeneous and large data in IoT. The wearable devices send data as time series to IoT hub on Microsoft Azure as seen in fig 5. The streaming data is analyzed then can be stored on the cosmos DB.

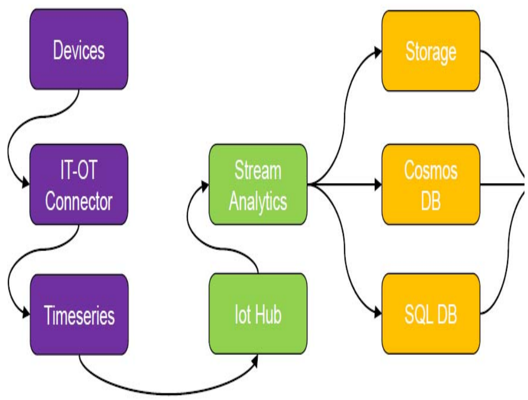

Fig 5: The Streaming Data From Wearables.

3.1 The supervised machine learning system

We used a hybrid system using adaptive SVM with feature selection to improve SVM performance. The system, in the beginning, extracts features from the records depending on disease parameters, and then we classify all potential features that relate to the diseases in the next phase depending on the threshold. The SVM machine learning is then used to identify and estimate the probability that the diseases may arise based on the data.

The parameter tuning is performed using the matrix analysis method to determine the best features to be regarded as parameters. Removing the redundant attributes from the dataset merely focuses on essential disease-related parameters.

The proposed method suggested a precision device algorithm as seen in Algorithm 1. It begins with data cleaning. It uses a threshold to perform the weighting process on the values. The aim or the main target is to neglect the features in case these features are below the disease's threshold. The lambda feature is used in Spark to pick the required data only. After that, the selection function is employed, then the parametrization of the SVM is implemented.

Algorithm 1
Starts Loading the dataset of Heart disease pre-processing with cleaning data providing the model script Setting the weights threshold Preparing the list of the feature (LF) For number 1 to length(weights) If weights >= threshold Applying (lambda Value: Value> Threshold of the dataset). Add a new entry in (LF) as a new column Implement the partitioning using Cross-validation Implement b starting to train and test the system using Cross-validation on the dataset Evaluation. Prediction Visualization END

3.2 The proposed health user profile on the cloud

Healthcare prediction or as in the proposed system is called a healthcare recommender system is any healthcare system that can help patients to preventive care through diagnosis and treatment to monitoring and follow-up. The EHRs contain all the data needed by clinics or healthcare provides to give recommendations or needed medicines, drug recommendations. The proposed System Provides a smart solution to manipulate the big data coming from different resources, it depends on the information of the data itself. The proposed system depends on the information which supposed to be provided by special people such as doctors. The system does not collect just information about the disease for the patient, the information is about the disease with one choice answer "yes" or "No" and the number that represents the disease level or state. The user also provides the system information about the normal activities with the number of occurrences.

To reduce the costs of medical visits or medical check-ups, the proposed system is providing the telemedical system that helps the patients to share their profiles on the cloud with healthcare providers to provide recommendations. This series is called the Medical Internet of Things (MIoT). The proposed system employed MapReduce from Hadoop to break down the data from different medical sources to smaller datasets, to be processed separately. The results of these smaller datasets are sum-reduced and mapped back to provide the outcome of processed data. The map-reduce framework can handle a large data set to be analyzed by Apache Spark.

Apache Spark offers an effective method for analyzing the health care data produced in Spark by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

applying machine learning to predict the outputs using memory processing. The framework offers a predictive healthcare method to provide a prompt analysis of the health situation using the environment of the individual, physical wellbeing indicators, and all available HER data. The cloud app media medical profile includes all personal medical background details from the customer and families and viewing details from mobile apps that can be used.

In the proposed system, Apache Spark(Fig 6) is used for medical data streaming and Cosmos DB as a NoSQL database. The Apache spark used to get data from Cosmos DB in real-time. The Spark gets the real-time data every 30 seconds as a default interval time which can be changed by the user based on the healthcare issues. Spark streaming receives data from wearable devices to analyze abnormal healthcare issues based on the user's thresholds. Then, it sends alerts to Cosmos DB which is used to notify doctors about emergencies. spark streaming is used to analyze the health data continuously and to use the data attributes to predict health risks. The output from the data is presented using Microsoft Power Bi to present the health issue as visualizations that can help healthcare providers to act as early as possible.

Spark streaming components such as MLib helps to perform predictive analytics on healthcare data using a machine learning algorithm. It helps to perform real-time analytics on data generated by wearable health devices. It generates data such as weight, BP, respiratory rate ECG, and blood glucose levels. Analysis can be performed on these data using k-clustering algorithms. It will intimate any critical health condition before it could happen. Apache Spark's RDD based computations are extremely fast in processing a large amount of data. Real-time streaming data from social networking sites can be processed effectively. Milb –Spark's in-built library supports machine learning which is essential for designing health recommender systems. The prediction and Recommendation component is built using a machine learning algorithm.

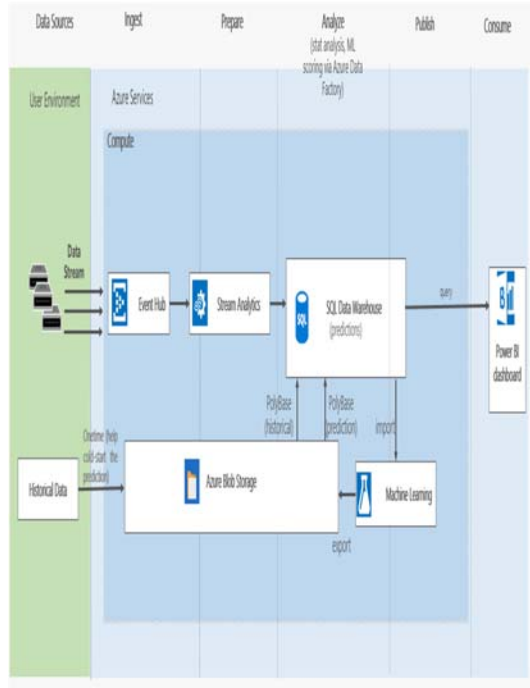

Fig 6: The Proposed System Based On Azure Data Analysis.

Spark Streaming can handle the streaming data from different sources in real-time. The data streams are sent as small batches called discretized streams (DStreams) to be processed in the Spark engine. Apache provides a machine learning library (Mlib) that contains a variety of algorithms such as clustering, classification, and collaborative filtering to filter attributes in the data.

The proposed presents as shown in fig 7 an architecture for dealing with the streaming data applications for the healthcare system. The input data could be from different input sources such as Kafka, Flume. The proposed system is a cloud-based system that employed Microsoft Azure Databricks to focus on streaming medical data to use the power of Spark streaming to analyze the streaming data on the cloud. Then the streaming data is batched to micro-batches for processing through Spark core. After the data processed, it is sent to the Cosmos DB which is connected to Microsoft Power BI to present the data on

Journal of Theoretical and Applied Information Technology
15th June 2020. Vol.98. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dashboards and help in healthcare decision making.

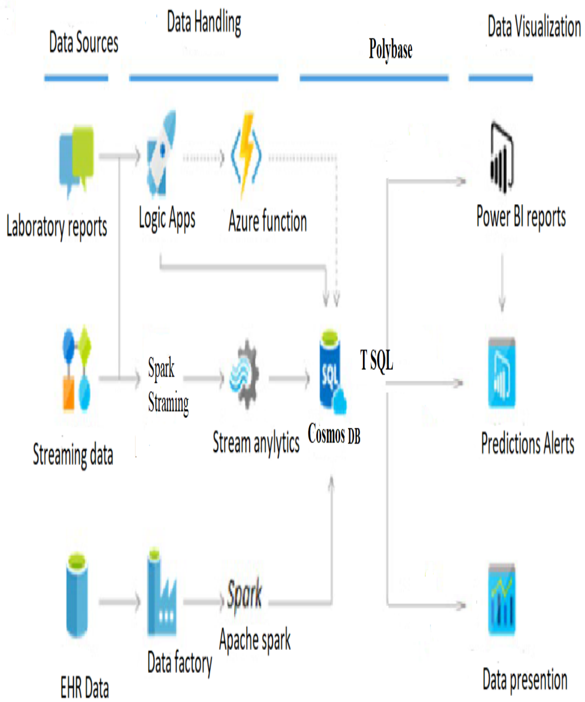

Fig 7: The Structure Of Streaming Medical Data On The Cloud.

4. 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. Emphasis on detecting heart disease involvement (value 1,2,3 and 4) and absence (value 0).

The data needed to be pre-processed and distinct for cleaning up the data after checking the dataset. The method of preparing or pre-processing data is to eliminate duplicate data and incomplete data to address anomalies in the data. The next step is data transformation, reducing features, and selection of attributes. The collection of features is then added. Then on the data set are implemented using SVM classification with tuning for parameters. The proposed system started the implementation (fig 8) with importing the heart disease dataset, pre-processing the data to clean the data, and remove redundant records. The next step is tuning the parameters based on the selected features, then the SVM classification has applied the training, and finally, the system is evaluated.

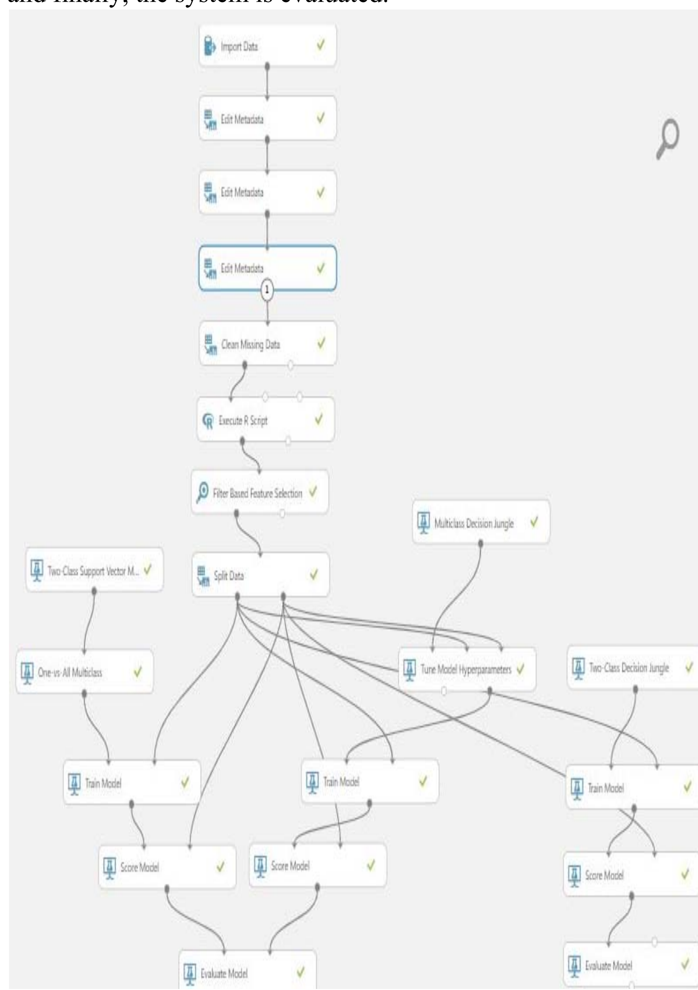

Fig 8: The Implementation Of The Proposed System On Azure.

The uncertainty matrix or a confusion matrix as it is called as well is used to demonstrate how the proposed model is distributed. As shown in figure 9, normal as false positive (FP) and false-negative (FN), True Positive (TP), and True Negative (TN).TP and TN are presented in white cells while FN and FP are presented in blue cells in the uncertainty matrix. The system proved a high accuracy and low false alarms.

Fig 9: The Uncertainty Matrix.

An assessment of the method is used to test the proposed model. Cross-validation then-fold splits the entire set of data into n folds. Education is performed with the n-1 folds. The process of testing is performed for each fold from n. Table 1 displays the average performance of those algorithms. 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. The system performance is calculated using F1, recall, and precision as seen in A, B, and C calculations.

$$F1 = \frac{2(\text{Precision} * \text{Recall})}{\text{Recall} + \text{Precision}} \quad (A)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (B)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (C)$$

Table 1: The System Assessment Based On The Dataset.

Positive Label	Negative Label
1	0

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	13	1	0.154	0.714	0.500	0.929	0.342	0.675	0.981	0.000
(0.800,0.900]	5	1	0.220	0.758	0.621	0.900	0.474	0.718	0.962	0.008
(0.700,0.800]	4	0	0.264	0.802	0.710	0.917	0.579	0.761	0.962	0.008
(0.600,0.700]	2	1	0.297	0.813	0.738	0.889	0.632	0.781	0.943	0.019
(0.500,0.600]	3	5	0.385	0.791	0.740	0.771	0.711	0.804	0.849	0.084
(0.400,0.500]	1	1	0.407	0.791	0.747	0.757	0.737	0.815	0.830	0.098
(0.300,0.400]	4	3	0.484	0.802	0.780	0.727	0.842	0.872	0.774	0.145
(0.200,0.300]	3	10	0.626	0.725	0.737	0.614	0.921	0.912	0.585	0.315
(0.100,0.200]	1	9	0.736	0.637	0.686	0.537	0.947	0.917	0.415	0.473
(0.000,0.100]	2	22	1.000	0.418	0.589	0.418	1.000	1.000	0.000	0.871

In fig 10, To give more picture of the false alarms, we used the Area under the curve (AUC) as shown in fig 10 to calculate the efficiency of the machine learning models. AUC indicates the increase in positive levels in the proposed method which are rated based on the qualified model dataset.

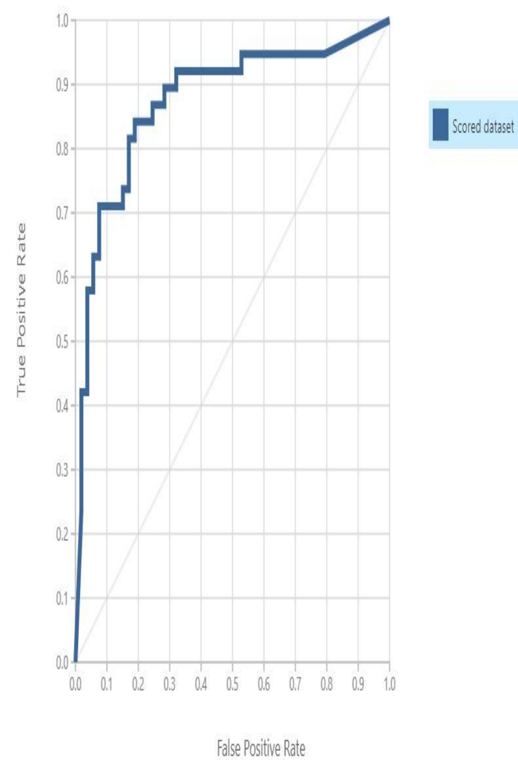

Fig 10: The System Evaluation Using AUC.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The age of participants in the dataset is shown in Figure 11, it shows that most participants are between 38 and 68 pages but usually, the participants in the dataset are between 34 and 75 years of age. The distribution of peak heart rates reading bases on the depression of the exercise in figure 12.

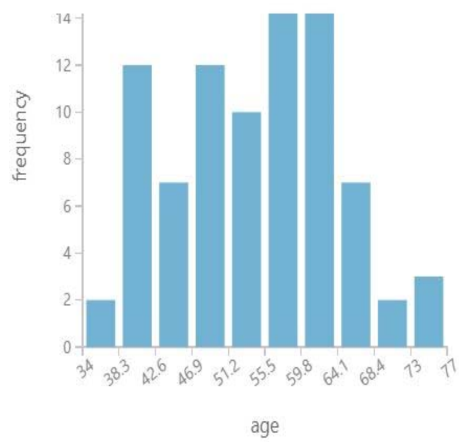

Fig 11: The Age Of Participants.

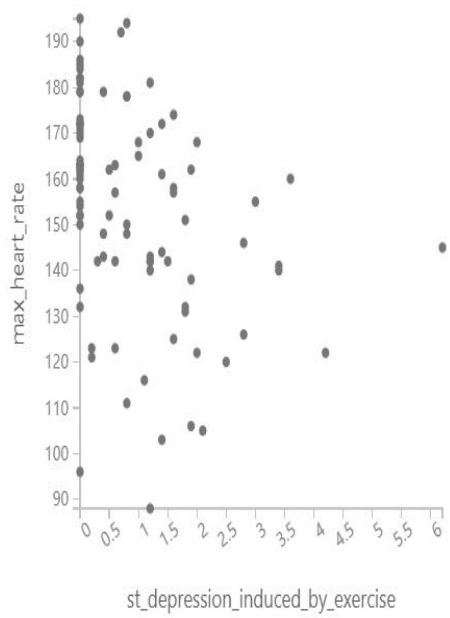

Fig 12: The Depression And Heart Rate Relationship.

Table 2 contrasts the proposed system and other methods applied to the UCI heart disease dataset and the proposed system showed a 90.6 percent improvement in accuracy with the highest accuracy compared to other classifiers achieving our study target. In comparison to the previous work from Su[27], the proposed system proved a higher and better accuracy as shown in fig 13.

Table 2: The Evaluation Of The Proposed System

Method	Dataset	Performance
Logistic Regression (SVM) [27]	UCI	86%
Naïve Bayes	UCI	84.2%
Random Forest	UCI	84.1%
Proposed System	UCI	90.6%

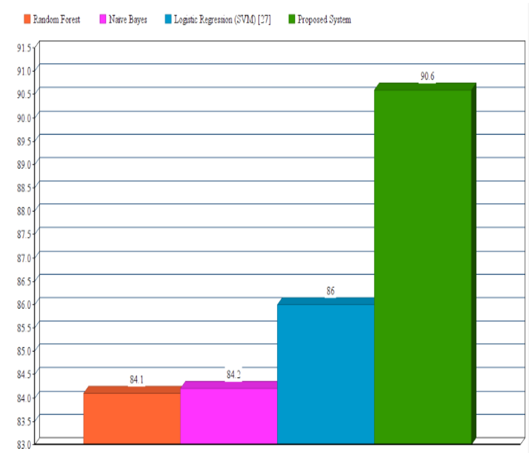

Fig 13: The Comparison Of The Proposed System To Previous Work.

5. 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.

6. CONCLUSION AND FUTURE WORK

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. The proposed method relies on the main parameters for the SVM based on tanning for these parameters.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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.

The implementation of the system proved a high accuracy of 90.6%. The suggested methodology was of better precision than the traditional SVM of parameter tuning. The efficiency of the SVM classifiers was assessed in terms of consistency, recall, precision, and goodness. The paper provides an efficient approach for heart disease prediction using a supervised machine learning algorithm to analyse the healthcare data to give insights and predictions to improve the quality of life of the people.

- The authors of the article confirm that there is no conflict with any published articles.
- Informed consent: Informed consent was obtained from all individual participants included in the study.
- Data availability: the data can be requested from the authors.

REFERENCES:

[1] J. Archenaa, A.Mary, "Health recommender system using big data analytics.", J. Manage. Sci. Bus. Intell. Pp. 17-24, 2017.

[2] A. Ismail, A. Shehab, I. El-Henawy, "Healthcare Analysis in Smart Big Data Analytics: Reviews, Challenges, and Recommendations", In Security in Smart Cities: Models, Applications, and Challenges, Springer, Cham, pp. 27-45, 2019.

[3] L. Nair, S. Shetty, "Applying spark-based machine learning model on streaming big data for health status prediction", Computers & Electrical Engineering, vol 65, pp. 393-399, 2018.

[4] M. Kamran1, A. Javed, "A Survey of Recommender Systems and Their Application in Healthcare", Technical Journal, University of Engineering and Technology (UET) Taxila, 2015

[5] M. Babar, F. Arif, M. Jan, Z. Tan, F. Khan, "Urban data management system: Towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop", Future Generation Computer Systems, vol 96, pp. 398-409, 2019.

[6] M. Sarowar, M. Kamal, N. Dey, "Internet of Things and Its Impacts in Computing Intelligence: A Comprehensive Review–IoT Application for Big Data", In Big Data Analytics for Smart and Connected Cities, pp. 103-136, 2019.

[7] V. Jeyabalaraja, P. Abirami, K. Janapriya, N. Raghavi, "Analysis of Disease from Multiple Healthcare Data using Hl7 Message on HDFS", International Journal of Computer & Mathematical Sciences (IJCMS), PP. 2347 – 8527, vol 7, 2018.

[8] M. Zaharia, R. Xin, P.Wendell, "Apache spark: A unified engine for big data processing", Communications of the ACM, vol. 59(11), pp. 56–65, 2016.

[9] L. Benhlima, "Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation", Advances in bioinformatics, 2018.

[10] M. Shahbaz, C. Gao, L. Zhai, F. Shahzad, Y. Hu, "Investigating the adoption of big data analytics in healthcare: the moderating role of resistance to change", Journal of Big Data, vol 6(1), 2019.

[11] A. Dumka, A. Sah, "Smart ambulance system using the concept of big data and internet of things. In Healthcare Data Analytics and Management", pp. 155-176, 2019.

[12] J. McPadden, T. Durant, D. Bunch, A. Coppi, N. Price, K. Rodgerson, W. Schulz, "Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform", Journal of medical Internet research, vol 21(4), 2019.

[13] A. Ismail, A. Shehab, I. El-Henawy, L. Osman, M. Elhoseny, "Quantified self-using IoT wearable devices", In International Conference on Advanced Intelligent Systems and Informatics", Springer, pp. 820-831, 2017.

[14] G. Manogaran, D. Lopez, "Health data analytics using scalable logistic regression

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with stochastic gradient descent", International Journal of Advanced Intelligence Paradigms, vol 10(1-2), pp. 118-132, 2018.

[15] Ismail, A., Abdlerazek, S., & El-Henawy, I. M. (2020). Development of Smart Healthcare System Based on Speech Recognition Using Support Vector Machine and Dynamic Time Warping. Sustainability, 12(6), 2403.

[16] J. Ang, A. Mirza, H. Haron, "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection", IEEE/ACM Trans Comput Biol Bioinf, vol 1-13(5), pp. 971–989, 2016.

[17] W. Gao, L. Hu, P. Zhang, "Feature selection by integrating two groups of feature evaluation criteria", Expert System Appl, pp.11–19, 2018.

[18] S. Bashir, F. Khan, Z. Khan, A. Anjum, "Improving Heart Disease Prediction Using Feature Selection Approaches", IEEE, pp. 619-623, 2019.

[19] H. Polat, H. MehrH, A. Cetina, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods", J Med Syst. pp. 41-55, 2017.

[20] J. Zhang, J. Xu J, X. Hu, "Diagnostic method of diabetes based on support vector machine and tongue images", Biomed Res Int, pp. 1–10, 2017.

[21] A. Phan, M. Nguyen, L. Bui, "Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems", Appl Intell, vol 46, pp.455–469, 2017.

[22] Y. Ren, F. Hu, H. Miao, "The optimization of kernel function and its parameters for SVM in well-logging, IEEE, 2016.

[23] Y. Xiao, H. Wang, W. Xu, "Parameter selection of Gaussian kernel for one-class SVM", IEEE Trans Cybern, vol 45(5), pp. 941–953, 2015.

[24] E. Gürbüz, E. Kılıç, "A new adaptive support vector machine for diagnosis of diseases", Expert System, vol 31(5), pp. 389–397, 2014.

[25] C. Lu, Z. Zhu, X. Gu, "An intelligent system for lung cancer diagnosis using a new genetic algorithm-based feature selection method", J Med Syst, 2014.

[26] A. Janosi, W. Steinbrunn, M. Pfisterer, R. Detrano, "UCI Machine Learning Repository", [cited 2019 August 1], https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[27] Wu CS, Badshah M, Bhagwat V. Heart Disease Prediction Using Data Mining Techniques. InProceedings of the 2019 2nd International Conference on Data Science and Information Technology 2019 Jul 19 (pp. 7-11).