iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32197

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Spark based Health Status Prediction on Real Time Streaming Data using Machine Learning

Mir Junaid Rasool[1], Amanpreet Singh Brar[2], Hardeep Singh Kang[3]
[1]*Department Of Computer Science and Engineering GNDEC Ludhiana*
[2]*Associate Professor Department of Computer Science and Engineering GNDEC Ludhiana*
[3]*Assistant Professor Department of Computer Science and Engineering GNDEC Ludhiana*

Abstract: ***To mine enormous and fast speed streams of data are amongst the major challenges nowadays in machine learning. The massive and huge chunks of data with changing velocity appears to require perspective change from conventional strategies in machine learning approaches. In the course of most recent decades cardiac disease is most well-known reason for worldwide deaths so early recognition can lessen the death rate. The amalgam of streaming big data and machine learning is advanced innovation that can demonstrate achievement in medical fields particularly early heart disease sensing. So, to find more affordable and powerful technology having both machine learning and big data streaming capabilities apache spark is the best choice. Apache spark is a strong data processing engine that can be used effectively for streaming continuous data having machine learning capabilities with its in_memory clustering and processing which is faster than Hadoop. This paper proposes building of machine learning models on health care heart disease data analyzing them and choose one showing best performance. The proposed work mainly aims on applying the chosen model on continuous streams of data when given certain input attribute variables. The data-set is used for training the model against real time streams of data to predict the disease.so having data of single patient or patients in predefined format the data can be sent into built system to predict presence of disease in real time.***

Keywords: ***Big data streaming, Spark, Heart disease, MLlib, Dataframes.***

Abbreviations: MLlib, machine learning library; LR, Logistic Regression; RF, Random Forest; ML, machine learning; GBT, Gradient Boosted trees; DT, Decision Tree;

I. INTRODUCTION

As of now we have been seeing large progressions in the size of data that is routinely produced and gathered so our capacity to analyze, store, process and mine these huge chunks of data is what is known to be big data science. The data is futile without a legitimate information gathered from it. The second-generation data handling and processing tools like spark can be utilized to process the large stable or streams of data and gather knowledge from it. Apache spark is a data processing system built to provide faster and easier to use analytics than Hadoop map reduce. The heart is that fundamental organ of the human physical body which assumes the primary role of blood circulation. In the off chance the heart will not work properly, genuine conditions of health like death will result at that moment [1][2].Coronary illness and stroke are world's biggest cause of death globally as per the World Health Organization (WHO) [3]. This paper represents the implementation of prediction of health status using real-time streaming data in the Spark framework using machine learning models. With the intention that person will send data attributes and can conclude whether or not to follow clinical consideration. The algorithms separately used to build models are LR, DT, RF and GBT. These algorithms are applied so as to choose which among the four shows high accuracy and takes less time concerning proposed work with the goal that we can pick best one. The application is built on affordable local machine although it can be developed and deployed on databricks and Amazon EC2 cloud. People in the modern world have exceptional worry for remaining strong and healthy. The progressions in the field of health services not only determining illness or cure but in addition chance of anticipating the possibility of attack its early detection is more valuable derived from already available healthcare datasets. In modern days people detect and know about various health related issues and information about them which are captured through wearable sensor devices and other equipment's available with them such as blood pressure, heart rate, sugar level etc. This data can be used against the trained model proposed in this work to get continuous monitoring of heart disease which will help them to take precautionary steps to reduce possibility of disease and improve the vitality rate [4][5].The paper is additionally sorted out as Section 1 further characterizes the inspiration of this work section 2 clarifies the literature regarding work done in this field. Section 3 characterizes the portrayal of proposed work, dataset and approach, section 4 involves Result and Discussions, the paper finally concludes in section V.

A. Inspiration for Health Status Prediction System

People in the modern world have exceptional worry for remaining strong and healthy. The progressions in the field of health services not only determining illness or cure but in addition chance of anticipating the possibility of attack its early detection is more valuable derived from already available healthcare datasets. In modern days people detect and know about various health related issues and information about them which are captured through wearable sensor devices and other equipment's available with them such as blood pressure, heart rate, sugar level etc. This data can be used against the trained model proposed in this work to get continuous monitoring of heart disease which will help them to take precautionary steps to reduce possibility of disease and improve the vitality rate.

II. RELATED WORK

Immense analysis of data these days has been a significant issue for countless scientists, particularly in investigating health care issues. Large Data mining including machine learning is used widely and numerous analysts have been utilizing sparks MLlib to predict health status. In [2] a machine learning model using Random forest classifier is built using spark and Casandra which can monitor and detect heart disease using streaming big data approach. In[4]Throughout this research, a flexible health monitoring prediction framework was developed with open source data processing tool, Spark, validated and implemented on cloud, where machine learning model is tested to streams of data. In [5]Spark with apache Casandra is utilized to predict breast cancer by applying machine learning using decision trees on real data event. This approach focuses on incorporating the breast cancer characteristics in the real-time classification model for patient care monitoring. Within this paper[6] , a machine learning approach based on Spark is applied dealing with the issue of the seizure spotting using classification on EEG data streams. In[7]Parallel Random forest classifier is presented for big data taking complementary hybrid approach to PRF Combining data parallel optimisation with task parallel optimisation performed on Apache Spark. In [8]an experiment was performed on the application of various data mining algorithms to predict and compare best predictive process the research results do not pose a dramatic difference in the prediction when using various classification algorithms in data mining. In [9] Researchers use structured and unstructured hospital data in this, and suggest a new algorithm for risk prediction ("CNN-MDRP") which is based on the covolutionary neural network. In[3] This uses Logistic Regression and SVM to detect lung cancer as the majority of deaths are caused by ignorance and delay in disease spotting. Logistic regression is highly accurate compared with SVM. In [10]Researchers provided a real-time heart disease predictive model which had been built using Apache Spark and Apache Kafka the built architecture contains three modules offline model, pipeline for processing data streams and prediction of real time data. In [11] classification is done to fast-speed big data streams using 'nearest neighbor' in spark environment. They have designed a solution called Ds_Rnge to address the problems of huge scale data streams. In[12] gives SVM based feature enhancement to get the heart disease they proved that SVM classifier performed well with ROC evaluation. In [13]the Architecture is designed to collect data store in S3 buckets of amazon then prediction model is trained using dataset. Performance like accuracy precision is computed as well. [14] provided RDD based fault tolerant interface for cluster apps to share data. RDDs can support wide variety of parallel requests and processes them in spark with iterative computing and being fault-tolerant. In[15] The hybrid method is pursued to boost the efficiency of predicting cardiovascular disease using ANN(artificial neural network) and the decision tree together. These hybridized decision trees were seen to have demonstrated better than existing separate algorithms of machine learning. In[16] it is seen by assessing the logical qualities, the technique of J48 tree has end up being the better model for expectation of coronary illness as it contains more accuracy and less all out ideal opportunity to develop.

III. DESCRIPTION OF PROPOSED WORK

The primary objectives of this research are analyzing the already available data-set of health care heart disease and to build machine learning models using various classifiers to predict the heart disease of new users whose streaming data is passed through this built model. The structure or architecture of this research work is described in Figure 1. In the proposed framework heart disease related text data is prepared in predefined manner, collected through net-cat server through TCP connection. The input Dstreams are processed through spark streaming which transforms these Dstreams to ML pipeline and are passed through machine learning Model built and trained using historical heart disease dataset. The Model analyses the input streaming data perform prediction which is stored as data frame and showed as output.

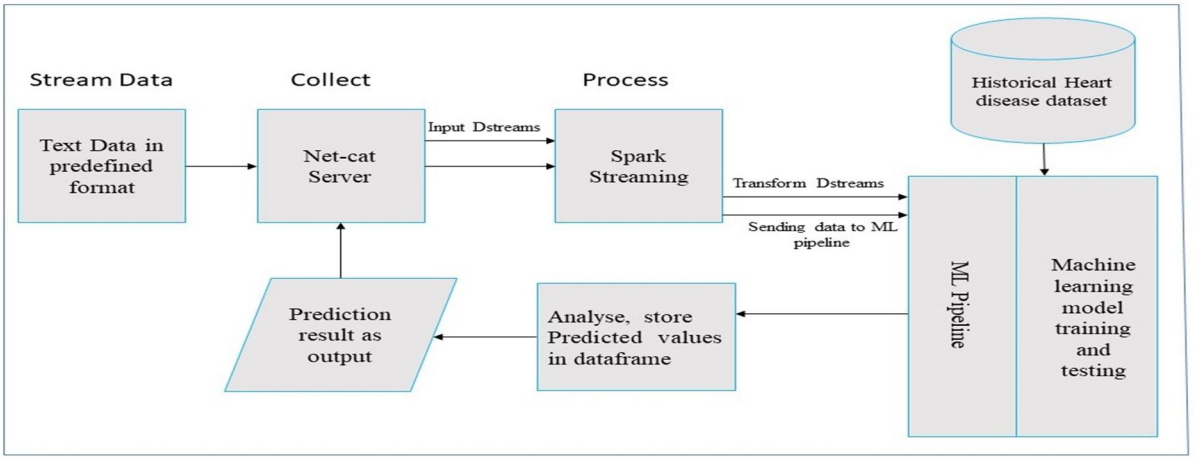


Figure 1 Architecture of prediction system using MLlib

A. Dataset

In our study "Healthcare data-set" "Cleveland.csv" related to heart disease which is available from machine learning UCI Repository is examined and used. For each record the data includes 303 records and 14 signs or attribute names. Dataset description is in the table I above. The field "target" indicates the patient's presence or existence of heart disease. It is indexed in integer from 0 (no presence) to 4(presence). Studies with the Cleveland database focused on simply trying to distinguish between the existence (values 1,2,3,4) and lack (value 0). The data-set is just used for training the model and can be replaced by any other dataset. With slight changes it can predict other disease as well. This sample data is chosen in the work as it comes from a legitimate source and is freely available online. it has been used in many past researches on big data science and machine learning. The target field is normalized and restricted to values 1 and 0. One of the principal benefits of this dataset is that it contains no categorical values so dataframe operations and preparing data for building ML model becomes easier. The data-set is small for training purpose and with time availability of huge datasets can prove better in terms of performance and prediction. The original data-set comprises of 76 features, and yet all documented experimental studies particularly use subset of 14 of them. In specific, the Cleveland data-set is the only one which is used by ML researches till date.

Table I Attribute names and description of heart disease dataset

Sno.	Attributes	Description
1	Age	age in years
2	Sex	1 for male 0 female
3	Cp	chest pain type Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic
4	trestbps	resting blood pressure (in mm Hg
5	chol	serum cholestoral in mg/dl
6	fps	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7	restecg	resting electrocardiographic results
8	thalach	maximum heart rate achieved
9	old peak	ST depression induced by exercise relative to rest
10	exang	exercise induced angina (1 = yes; 0 = no)
11	slope	the slope of the peak exercise ST segment
12	Ca	exercise induced angina (1 = yes; 0 = no)
13	thal	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14	target	Label class 1 for presence and 0 for absence of heart disease

B. Methodology

In the structural health observing coronary illness prescient examination is led using Apache Spark's MLlib. Different classification strategies are utilized to assemble model and parameters like accuracy is figured. The classifiers utilized in this investigation are 'LR', RF, DT and GBT. However, the main focus of the study remained the streaming data the built model is trained with existing heart disease database then live streams are tested against the built model. The data streams need to be sent in predefined format as per the data-set used for training the model. The structure of the work flow is shown in the fig 3.

C. Machine learning Models

The first criteria to create a heart disease tracking system is developing an (ML) model that actually predicts the disease when certain attributes are given in predefined format. We had various options to choose the classification algorithms for developing a model. Sparks MLlib endorses all classifiers which can handle large sets of data as well. With regard to offline model, four classifiers have been tested to construct model in order to pick the best one for our final monitoring system. The four classifiers used are Logistic Regression, Decision Tree, Random Forest and Gradient Boosted Trees (GBT). The accuracy and other parameters were calculated while the heart disease data-set was split in 70% and 30% in training and testing data. The algorithm uses different arguments like for LR maxIter, regParam, for GBT maxIter and for DTC, RF maxDepth, maxBins and numtrees. The resulting accuracy and other parameters like specificity and sensitivity for these algorithms while using the defined parameters are shown in table 2.

Table II performance values of the classifiers

Algorithm	Specificity	Sensitivity	Accuracy
LR MaxIter=10	82	94	89
DT MaxDepth=3 maxBins=10	87	80	83
RF maxDepth 4 maxBins = 64 numtrees=20	79.48	94	87.64
GBT MaxIter=10	79.4	90	85.3

As seen in table 2 above that the highest accuracy and sensitivity is achieved by logistic regression and then random forests, GBT and decision trees. So, for the live streaming data, training model is built using Logistic regression for the final health care monitoring system. Generally, Sensitivity depicts how good the procedure predicts one category and Specificity depicts how good the procedure predicts the another. Whereas Accuracy would calculate how well each category are predicted by the study. Accuracy, sensitivity, specificity is computed using equations 1 ,2 and 3.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Senstivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

The confusion matrix describes a classification model's performance; it contains details regarding observed and forecast classifications conducted by a classification algorithm. Fig 2 defines confusion matrix in which TN represents true negative, TP true positive, FP false positive and FN false negative.

	Predicted No 0	Predicted Yes 1
Actual No	TN	FP
Actual Yes	FN	TP

Figure 1 Depiction of confusion matrix

True positive (TP): We expected yes (they have the disease) in these and the disease is real. True negatives (TN): We were expecting no, and they do not have any illness. False positives (FP): Yes expected, yet they don't actually have the illness. (And is said to be "Type I error") False negative (FN): We were expecting no, but the disease still resides. (Also said to be "Type II Error").

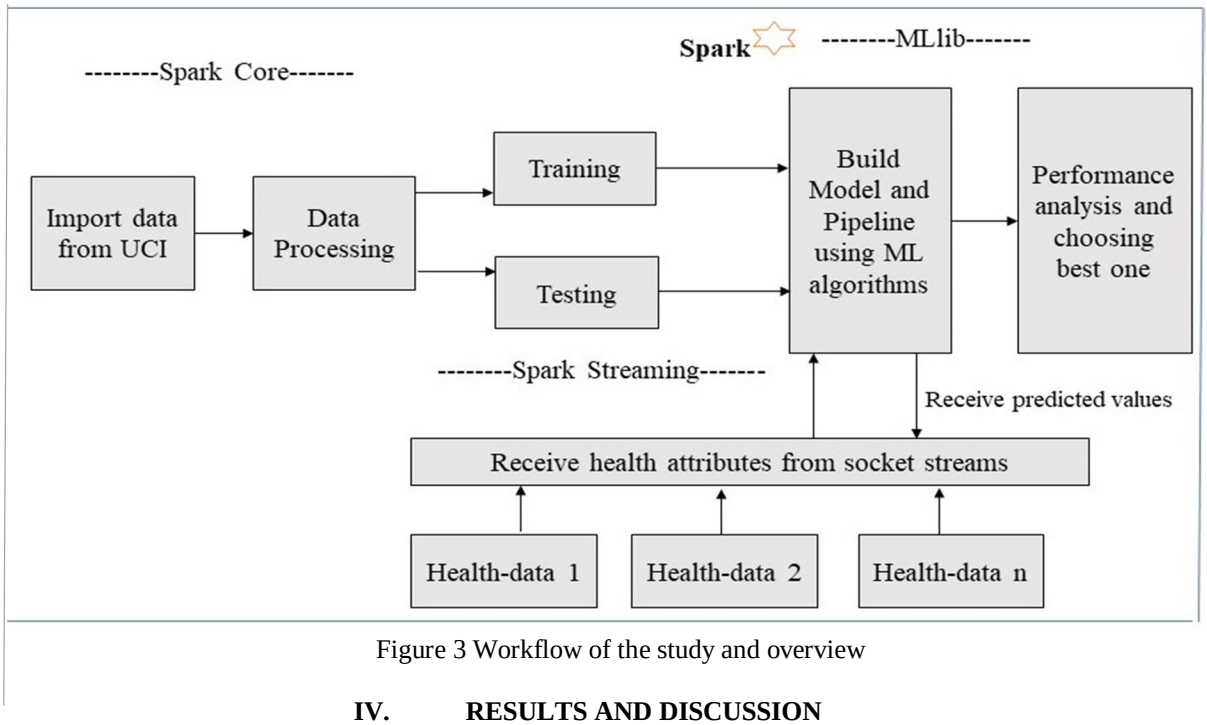


Figure 3 Workflow of the study and overview

IV. RESULTS AND DISCUSSION

The Confusion matrices acquired from all the classifiers can be seen in table III, and the accuracy comparison graph is shown in fig 5. As shown in the graph highest accuracy in the offline model where data-set is randomly divided into 70:30 ratio with the parameters shown in Table II taking into consideration all attributes, the LR and RF Classifier shows high accuracy of 89 and 87. i.e. LR proved better than RF also when taking streaming data into consideration.

Table III confusion matrices of four Algorithms

LR		DT		RFC		GBT	
32	7	34	5	31	8	31	8
3	47	10	40	3	47	5	45

Thus, Final model is constructed using LR Classifier and is chosen for health monitoring system. The LR ROC (Receiver operating characteristic curve) is shown in fig 4. Data input stream sent for prediction analysis is done by the use of TCP message via Net-cat server. In this analysis, a message in predefined format is sent to the model through specific port, it receives a message and implements several transformations and then applies Logistic Regression to show the predicted values.

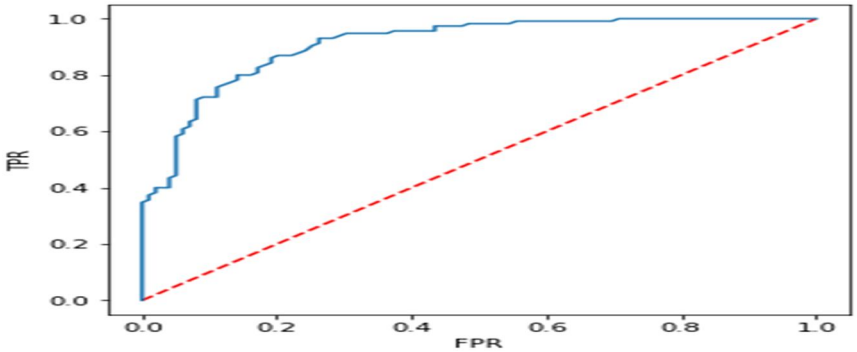


Figure 4 ROC curve of LR

Our work builds on a framework that uses distributed machine learning for stream data in real time. Spark has the ability to use various technologies together, such as Spark core, Spark SQL, MLlib, and Spark streaming. This research is conducted on virtual box with Ubuntu operating system and with sparks python API called pyspark. Results of some of the ingested data streams in predefined format are shown in table IV. Analysis of real time data streams is performed by the proposed model. However, the data-set used for training model in this work is small and with available large datasets of heart disease, model can be trained more well with high accuracy. The dataset used in this work can be changed by other disease datasets and can predict the same for that particular disease. The data streams are ingested from Netcat server in the proposed system but with slight change data can be sent to it from other social sites like twitter.

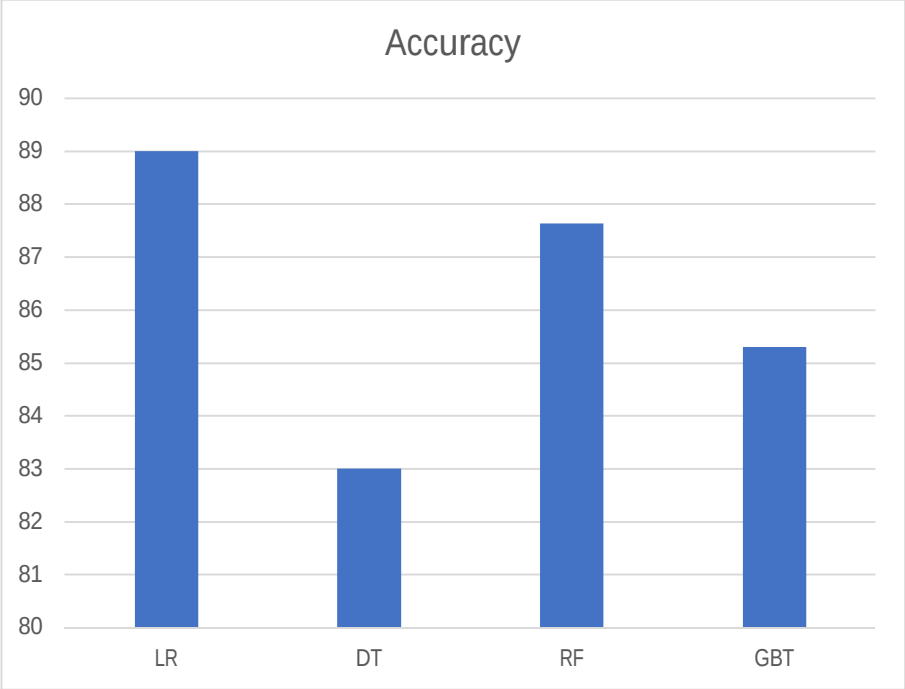


Figure 5 Graph of accuracy Obtained by used Algorithms

Table IV Results of some of data streams for heart disease prediction

Stream Text (“Age_Sex_Cp_testbps_chol_fbs restecg_thalach_exang oldpeak_slope_ca_thal”) “Where ‘_’ represents space”	Predicted value
TWEET_APP 63 1 3 145 233 1 0 150 0 2.3 0 0 1	0
TWEET_APP 37 1 2 130 250 0 1 187 0 3.5 0 0 2	0
TWEET_APP 70 1 1 145 233 1 0 150 0 2.3 0 0 1	1
TWEET_APP 41 0 1 130 204 0 0 172 0 1.4 2 0 2	1
TWEET_APP 56 1 1 120 236 0 1 178 0 0.8 2 0 2	1
TWEET_APP 66 1 1 120 236 0 1 178 0 0.8 2 0 2	1
TWEET_APP 60 0 2 135 210 1 1 171 0 0.9 2 0 1	0
TWEET_APP 29 0 3 160 285 1 1 187 0 1.3 1 0 1	0
TWEET_APP 57 1 1 167 281 1 0 182 0 1.5 1 1 1	1

We have compared this work with [Nair et.al (2018)][4] and [Ed-daoudy et al. (2019)][2] as both these researches have worked with same dataset with aim to build prediction system. When analysing [Nair et.al (2018)] they have used decision tree classifier with different values for parameters maxDepth and maxBins computed generalization error at different instances and finally used decision tree with parameter maxDepth 6 and maxBins 100 for final prediction system as it shows least generalization error. However, we tried four algorithms with their respective parameters and we got maximum accuracy when using LR i.e. 89%. Which is 6% higher than [Nair et.al (2018)][4].

On the other hand, when analysing and comparing [Ed-daoudy et al. (2019)][2] they have used Random forest classifier for final prediction system and have computed all the three parameters as we have computed. The values for the parameters they got using RF are accuracy 87.50%, Specificity 88.37%, sensitivity 86.66%. When we tried RF, we got the same accuracy as that of their model. But we have selected LR for our prediction system as it proves better in terms of accuracy, specificity and sensitivity as well we got 2% more accuracy and 8.5% more sensitivity than their study.

V. CONCLUSION

Healthcare data rises at an unprecedented pace in different and from unreliable sources from time to time. Heart disease has become one of the major illnesses in the world from which people die. So, a system must be built that will detect this disease earlier. Machine learning with the spark's Big Data engine can manage massive quantities of data and apply machine learning to the same. The distinction between this study and other research is that they have used platforms such as Kafka Casandra flume etc. with spark. whereas our entire system uses spark only so that fewer technology and cost-effective approach is followed.

VI. FUTURE SCOPE

With increasing the number of the diseases at an unprecedented pace, the main emphasis is on early detection and real-time response with the first preference. So, we can think of many things which need high attention specially in healthcare to work in future and with available ML and AI based techniques a complete healthcare monitoring application is needed which can predict various diseases based on input attributes given by individual which are known to him. with little modification the built application can predict other diseases as well. As a future scope web application can be built which can collect information and ingest those data attributes into spark model and predict disease and finally gives output on dashboard.

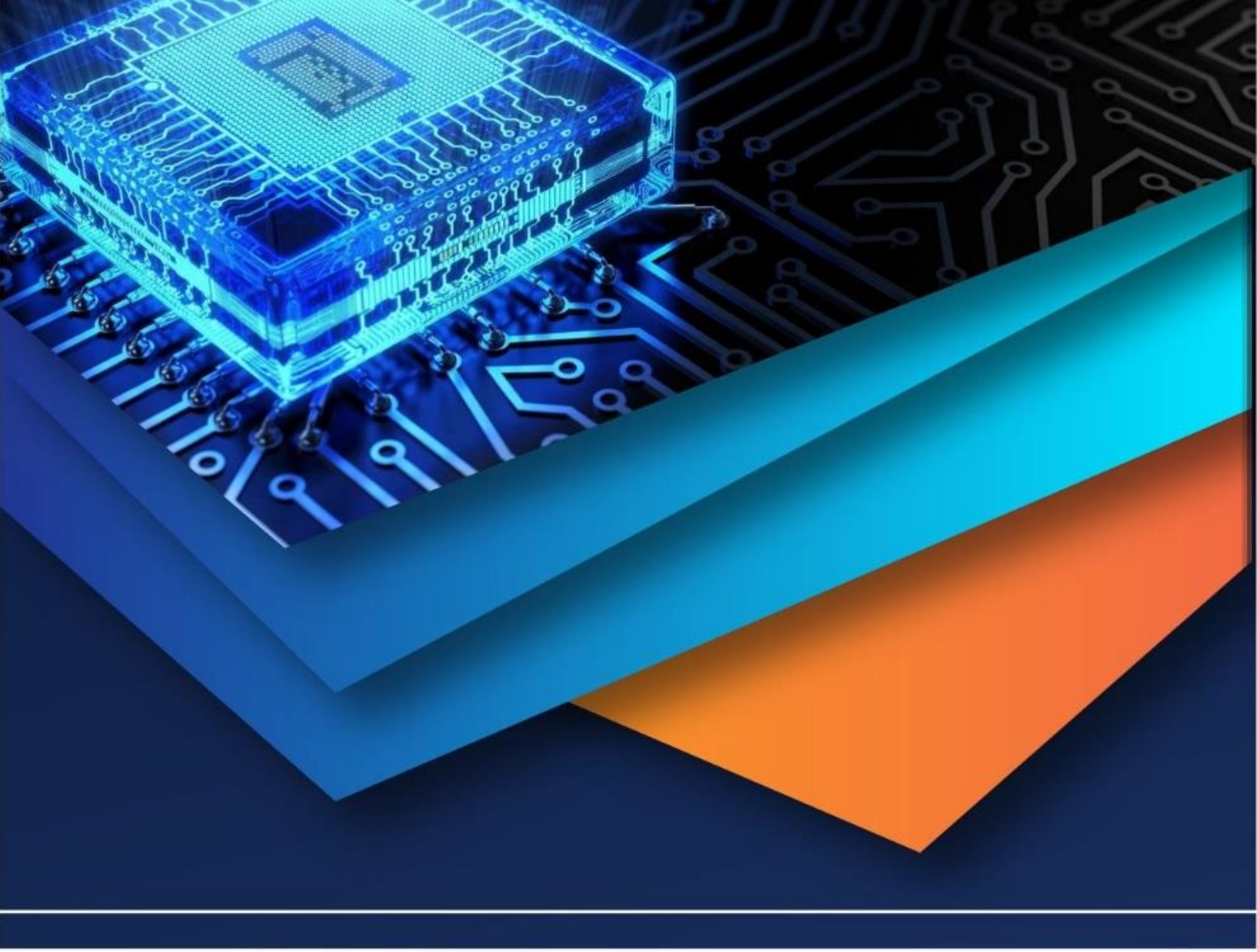
VII. ACKNOWLEDGMENTS

I would like to thank Dr. Sehijpal Singh, Principal and Dr. Parminder Singh, HOD CSE Guru Nanak Dev Engineering College (GNDEC) Ludhiana for providing us an opportunity to conduct this research. To Prof. Amanpreet Singh Brar and Prof. Hardeep Singh kang, Department of Computer Science and Engineering, (GNDEC), Ludhiana, to provide academic feedback, direction and support throughout the entire period of the project.

REFERENCES

[1] R. Elshawi, S. Sakr, D. Talia, and P. Trunfio, "Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service," Big Data Res., vol. 14, pp. 1–11, 2018, doi: 10.1016/j.bdr.2018.04.004.
[2] A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 Int. Conf. Wirel. Technol. Embed. Intell. Syst. WITS 2019, pp. 1–5, 2019, doi: 10.1109/WITS.2019.8723839.
[3] H. Animesh, K. M. Subrata, G. Amit, M. Arkomita, and A. Mukherje, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review," Adv. Comput. Sci. Technol., vol. 10, no. 7, pp. 2137–2159, 2017.
[4] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction," Comput. Electr. Eng., vol. 65, pp. 393–399, 2018, doi: 10.1016/j.compeleceng.2017.03.009.
[5] A. Ed-Daoudy and K. Maalmi, "Application of Machine Learning Model on Streaming Health Data Event in Real-Time to Predict Health Status Using Spark," Int. Symp. Adv. Electr. Commun. Technol. ISAECT 2018 - Proc., pp. 1–4, 2019, doi: 10.1109/ISAECT.2018.8618860.
[6] M. S. E. Sendi, M. Heydarzadeh, and B. Mahmoudi, "A Spark-based Analytic Pipeline for Seizure Detection in EEG Big Data Streams," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2018-July, no. April 2019, pp. 4003–4006, 2018, doi: 10.1109/EMBC.2018.8513385.
[7] J. Chen et al., "A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment," IEEE Trans. Parallel Distrib. Syst., vol. 28, no. 4, pp. 919–933, 2017, doi: 10.1109/TPDS.2016.2603511.
[8] H. D. Masethe and M. A. Masethe, "Prediction of heart disease using classification algorithms," Lect. Notes Eng. Comput. Sci., vol. 2, pp. 809–812, 2014.
[9] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," IEEE Access, vol. 5, no. c, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
[10] H. Ahmed, E. M. G. Younis, A. Hendawi, and A. A. Ali, "Heart disease identification from patients' social posts, machine learning solution on Spark," Futur. Gener. Comput. Syst., 2019, doi: 10.1016/j.future.2019.09.056.
[11] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data," Knowledge-Based Syst., vol. 117, pp. 3–15, 2017, doi: 10.1016/j.knosys.2016.06.012.
[12] C. B. Gokulnath and S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease," Cluster Comput., vol. 22, pp. 14777–14787, 2019, doi: 10.1007/s10586-018-2416-4.
[13] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," Futur. Gener. Comput. Syst., vol. 82, pp. 375–387, 2018, doi: 10.1016/j.future.2017.10.045.
[14] M. Zaharia et al., "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," Proc. NSDI 2012 9th USENIX Symp. Networked Syst. Des. Implement., pp. 15–28, 2012.
[15] S. Maji and S. Arora, Information and Communication Technology for Competitive Strategies, vol. 40. Springer Singapore, 2017.

[16] C. Science and G. M. Faculty, "H e a r t D i s e a s e P r e d i c t i o n U s i n g M a c h i n e l e a r n i n g a n d D a t a M i n i n g T e c h n i q u e," 2009.
[17] "WHO official site." [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
[18] "Apache spark official site." [Online]. Available: https://spark.apache.org/docs/latest/ml-guide.html.
[19] "Dataset site." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.
[20] J. G. Shanahan, H. Street, H. Street, and S. Francisco, "Large Scale Distributed Data Science using Apache Spark," pp. 2323–2324.
[21] R. Lin, Z. Ye, H. Wang, and B. Wu, "Chronic Diseases and Health Monitoring Big Data: A Survey," IEEE Rev. Biomed. Eng., vol. 11, no. c, pp. 275–288, 2018, doi: 10.1109/RBME.2018.2829704.
[22] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, "A disease diagnosis and treatment recommendation system based on big data mining and cloud computing," Inf. Sci. (Ny)., vol. 435, pp. 124–149, 2018, doi: 10.1016/j.ins.2018.01.001.
[23] R. James, G., Witten, D., Hastie, T., Tibshirani, An Introduction to Statistical Learning - with Applications in R | Gareth James | Springer. 2013.
[24] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations.," Northeast SAS Users Gr. 2010 Heal. Care Life Sci., pp. 1–9, 2010.
[25] G. Manogaran, R. Varatharajan, and M. K. Priyan, "Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System," Multimed. Tools Appl., vol. 77, no. 4, pp. 4379–4399, 2018, doi: 10.1007/s11042-017-5515-y.
[26] H. Ahmed, N. Omran, S. F. Abd-El Ghany, E. M. G. Youn, N. F. Omran, and A. A. Ali, "Stroke Prediction using Distributed Machine Learning Based on Apache Spark Internet of Things View project An Integrated Smoothing Method For Fingerprint Recognition Enhancement View project Stroke Prediction using Distributed Machine Learning Based on Ap," Int. J. Adv. Sci. Technol., vol. 28, no. 15, pp. 89–97, 2019, doi: 10.13140/RG.2.2.13478.68162.
[27] P. Carbone et al., "Apache Flink™: Stream and Batch Processing in a Single Engine," vol. 36, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY