IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Predicting and Visualizing the Heart Diseases by Machine Learning Algorithms with Big Data

R. Kannan[1*], V. Vasanthi[2]

[1]Department of Computer Science, Rathinam College of Arts & Science, Coimbatore, India
[2]Department of Computer Science, Rathinam College of Arts & Science, Coimbatore, India

[*]*Corresponding Author: dschennai@outlook.com, Tel.: +91-9003865448*

Available online at: www.ijcseonline.org

Accepted: 10/Jun/2018, Published: 30/Jun/2018

***Abstract*—** Nowadays, most of the researchers are focusing and inspiring the healthcare and medical industries. Because, globally humans are affected by various diseases. Especially, the heart diseases are major defect for humans, which is unpredictable and it may happen any moment in the human life. The Machine Learning (ML) algorithms and Big Data technologies are providing the complete and effective solutions for the healthcare and medical industries to predict and diagnosis the various diseases. Moreover, it helps to protect the human life by the accurate prediction results in real-time. Purpose of this paper, to predict the heart diseases automatically by segmenting and classifying the patient's heart data in real-time with the help of machine learning algorithms, big data, wireless heart monitor and smart phones. Finally, this research helps to predict, visualize and monitor the patient's data in remotely and alerting to the heart specialists and health care professionals to save the patient's life.

***Keywords*—** machine learning, big data, predicting heart diseases, visualizing heart diseases

I. INTRODUCTION

Machine Learning and Big Data algorithms are an active role in the healthcare and medical industries. The growth of healthcare and medical industries data are giving the opportunities to improve the patient's health. The cardiac failure is the basic cause for the human death in every country. The heart specialists can predict and diagnosis the cardiac diseases by the machine learning algorithms and big data technologies with accuracy and efficiency. Therefore, they can take very less time to predict whether the person defects by cardiac diseases or not. The current research in the study of cardiac-related disease compartmentalization to find the disguise medical information.

The Machine learning techniques allows to monitor and visualize the heart patient's data to alert the healthcare professionals, family members, nearest hospitals and ambulance when the patient suffers from the heart attack. It helps to automatically find the hidden reasons and efficiently give the resolutions for healthcare professionals. This research prominences the heart diseases prediction technique in terms of its accuracy and percentage of error by using the machine learning algorithms, big data, and heart monitoring systems.

A. Heart Diseases

Heart diseases are life frightening disease, and it should be contemplating as a global health precedence. Moreover, it resides a stress on patients, caretaker and healthcare systems. At present, almost 30 million of people are living with the heart diseases globally. Although the survival rates as of present globally 48.9 million people affected by heart defects by birth defects. The reasons and heart defects are cholesterol deposits, high blood sugar, poor in hygiene, physically inactive, unhealthy diet, smoking and hand change smoking, overweight, blood pressure, viral infection compare with the survival rates worse than any other diseases.

B. Machine Learning (ML)

The modern computers can learn and work itself by utilizing the machine learning algorithms for completing the critical tasks which process cannot be process simply done by humans such as Netflix movies rating, Amazon online recommendations, diagnosis in medical imaging, autonomous robotic surgery. The machine learning algorithms are four different types such as supervised learning, unsupervised learning, semi supervised learning and reinforcement learning. The below fig.1 shows the contradiction between the traditional programming and machine learning to provide the best results.

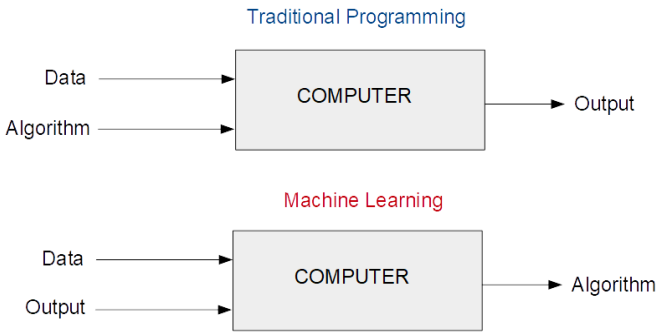

Figure.1 Overview of Machine Learning

C. Big Data

Big data is the distributed and parallel processing file system that allows the large and various types of data to store and access with less time (Fig.2). In addition, it also supports the more complex databases which it cannot possible to process by historical database management systems [11]. In additionally Big data gives many more benefits and opportunities like data security, cost reduction, and faster decision-making. Big data contains several tools like Hadoop, Map Reduce, Hive, Pig, Spark etc.

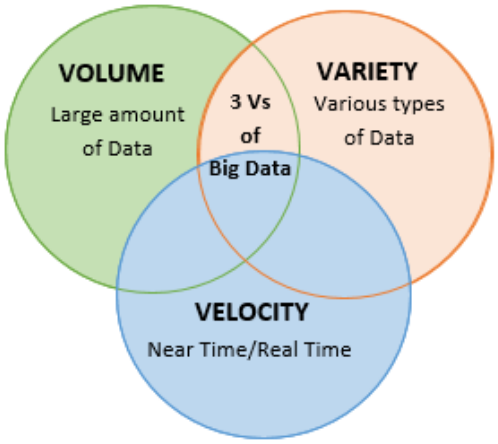

Figure.2 Big Data File System

II. MATERIALS AND METHODS

The main objectives of this paper is to develop the heart diseases prediction system for the heart patients, heart specialists, health care professionals and hospitals. In this system, we use the Logistic regression, ROC curve and Big Data Spark with heart monitoring device to indicate automatically the patient and health care professionals when the person suffers from heart attack with complete patient data by 24/7 online. In the same can track and visualize the data by the patients and health care professionals based on their location. Moreover, we are extracting the data from HRMS and comparing the same with AHA dataset to improve the prediction of the extracted patient's data.

A. Wireless Heart Rate Monitoring System (HRMS)

Wahoo X is a wireless heart rate monitoring sensor system (Fig.3) that extracts the Heartbeat rate, Rhythm, calorie, distance, steps in walking and running smoothness, body temperature and present location of the patient's data [14]. Wahoo X Bluetooth consumes low energy and stand for long time when connects with Android and IOS mac. From Wahoo X, we have obtained 15 heart patients data with 21 set of variables such as Name of patient, Age, Sex, the patient contact no, patient location, hospital name, hospital contact no, specialist name, contact no, heart features and etc..

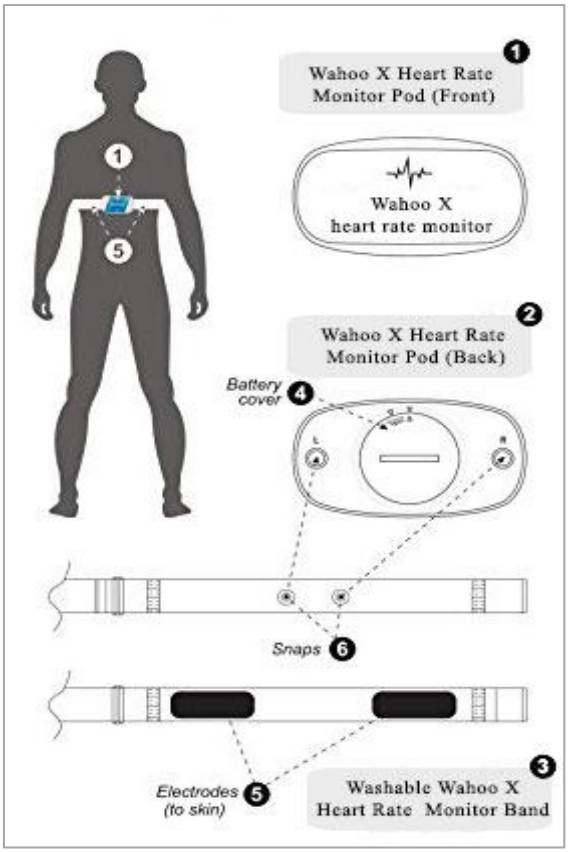

Figure.3 Wahoo X HRMS

B. American Heart Association (AHA) Dataset

In this paper, we have utilized the American Heart Association (AHA) dataset to compare with the on time patients data of HRMS dataset to alarm the heart patient abnormality [13]. In addition, we have discussed and confirmed the classification of this data with heart specialists (shows below table 1).

Table. 1 AHA Predicted Heart Diseases Data

Sinus Rhythm Type	Threshold value of HR, BP and hotness
Normal	60<=HR<=100 (Beats /minute), BP=100-140/60-80 mmHg, & Hotness =36.5-37.5^{o} C
Bradycardia	HR<=60 (Beats /minute)
Tachycardia	HR>=100 (Beats /minute)
Hypertension (Stage 1)	BP=Sys/Dys>=140/90 mmHg
Hypertension (Stage 2)	BP=Sys/Dys>=150/95 mmHg
Hypotension	BP=Sys/Dys<=00/60 mmHg
Fever	Hotness >=37.8 ^{o} C
Hypothermia	Hotness <=35.0 ^{o} C

C. Logistic Regression

The Logistic regression algorithm includes two type of methods called regression and classification for evaluating the dataset. It contains one or more independent variables that concludes an outcome. Meanwhile, the logistic regression searches the best fitting method to describe the relationship between response and predictor variables [10]. However, this method can give the two possible outcomes as in the Fig.4.

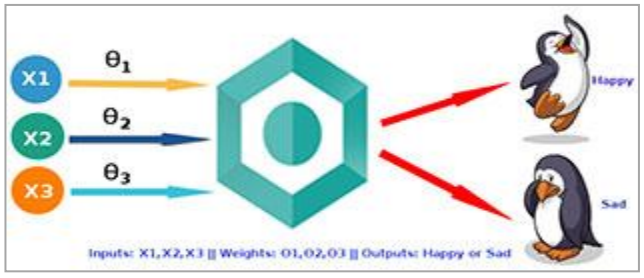

Figure.3 Logistic Regression Model

D. ROC Curve

There are more techniques available to evaluating the classifier performance. ROC curve is one of the major and popular technique used to improve the performance effectively by the model wide evaluation measure from the confusion matrix [9].

Model wide evaluation measure separates the datasets into two parts as positive and negative and use the below performance measures to measure the performance of separated a dataset [9].

- Specificity - to measure the negative part of a dataset
- Sensitivity - to measure the positive part of a dataset.

The Calculation of Model-wide evaluation based on the threshold values of scores (including discriminant values, posterior probabilities, etc.,) with labels, which has been produced by the ML Algorithms [9].

E. Machine Learning Software

As time goes on, there are best tools and software's are available to predict the heart diseases in the market. Here, we will discuss some of the favourable open source software name called as "R programming". R is the most popular statistical computing software package in the worldwide [4]. It is also easy to use a range of intuitive packages for big data, Machine learning (ML), and Internet Of Things (IOT) which has traditionally been used in the number of growing healthcare industries, commercial, government organizations, academics, and research [11]. R is the open-source, flexible, and customizable. Over 2,500 R packages contributed as extensions of innovative Analytics tools in the world.

III. EXPERIMENTAL EVALUATION

In this section, we have applied and evaluated the below six process (Fig 4) to achieve the results.

1. Extracting the real patient's data for every one minute from the heart monitor and send it to Smart Phone via Bluetooth.
2. Transfer the patient's data from Smart Phone to Big Data via the internet for every five minutes.
3. Use the Logistic regression and Spark ML framework in R programming for classifying and labeling the patients data on time.
4. Validate the real patient's data with American Heart Association (AHA) Dataset by ROC curve and boosting algorithms.
5. An alarm will send to the patient with the details of nearest heart specialists and hospitals via Smartphone when that patient abnormal stat is found after comparison. Meanwhile, the heart specialist and hospital get intimated with details of patient abnormal state and their location.
6. In addition, the patients, heart specialists and hospitals can keep track and visualize the patient's details at any time.

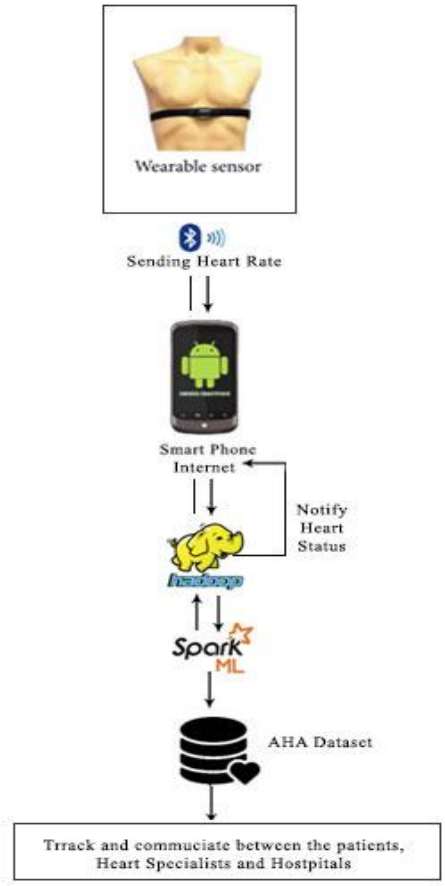

Figure.4 Architecture of the system

After many trial and errors, we have completed the all process with help of AHA dataset and heart specialists. We predicted the results from HRMS dataset with tuning parameters and compare AHA dataset. Eventually, the patients, hospitals and specialist can able to track the prediction data in smart phones and computers (Fig.5)

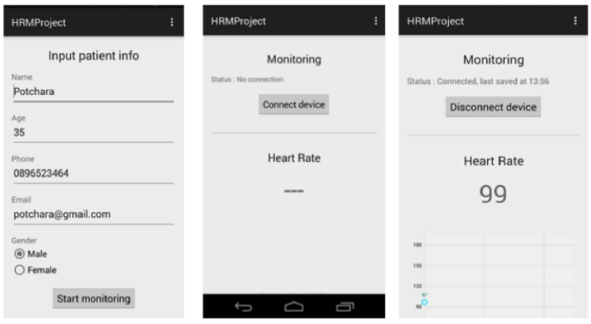

Figure.5 Heart Rate Monitoring System Application

Moreover, we have categorized the prediction data by the age wise for the alert messages (table 2).

Table.2 Age wise categorized Heart Diseases Data

Age	18-35	36-64	Above 64
Normal HR	72-75(BPM)	76-79(BPM)	70-73(BPM)
Bradycardia	HR<=55	HR<=60	HR<=65
Tachycardia	HR>=110	HR>=120	HR>=100
Hypertension	BP>=150/100	BP>=145/195	BP>=140/90
Hypertension	BP<85 mmHg	BP<96 mmHg	BP<117 mmHg
Fever	Temp>=37.2°C	Temp>=37.5°C	Temp>=36.9°C
Hypothermia	Temp<35.5°C	Temp<35.1°C	Temp<35.0°C

Finally, we have developed the tool for monitoring and visualizing the prediction data with help of R programming (shows fig.6).

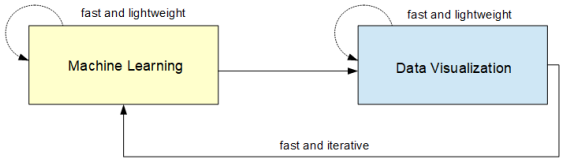

Figure.6 Machine Learning Vs Data Visualization

IV. RESULTS AND CONCLUSION

Twenty-one predicted variables from the fifteen patient's dataset are used to predict and diagnosis the heart disease. The performance of the logistic regression is compared to the accuracy obtained between AHA dataset and patient's real-time dataset to predict the heart diseases. A comparison of the area under the ROC and the accuracy of these model predictions show and performs with 0.97 % of accurate prediction.

After the comparison, this system can allow heart specialists and the patients, to visualize the heart's activities and to learn about the patient heart's activities 24/7 via online. If the patients get the cardiac failure symptoms then the alert will be sent to their family members, heart specialists and nearest hospitals with the details of abnormality via the mobile phone based on their location.

Furthermore, based on their earlier records first aid advice intimates to the patients and their relatives to protect them from out of danger.

This system can reduce the number of follow-up and hospital visits of the patients and it allows the patients/heart specialists/hospitals to check the patient's heart 24/7 online by the smart phones and computers (Fig.6). Hence, the death rate of heart patient by the heart failure can be reduced via this system through the effective and on time prediction.

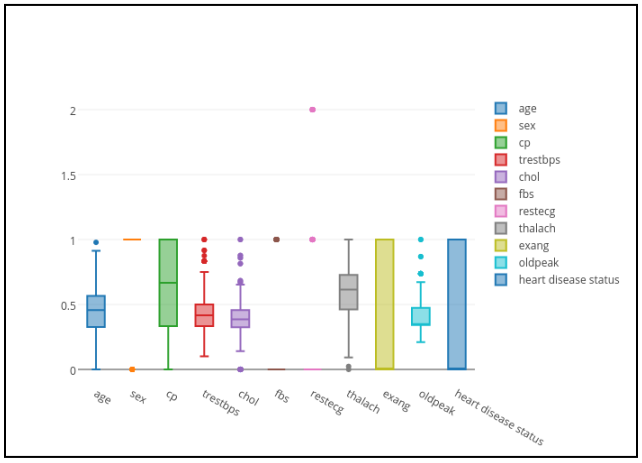

Figure. 6 Monitoring & Visualizing Heart Prediction Data

V. FUTURE WORK

This paper has summarized the methods for heart disease predication along with innovative machine learning algorithms. As a future work, we have planned to predict the heart diseases using Tensor flow of deep learning algorithms with more dataset. This Deep Learning Tensor Flow will automate and increase the process of prediction in terms of speed.

REFERENCES

[1] Pedro Domingos, "The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World," .

[2] Ian H. Witten, Eibe Frank , Mark A. Hall, Data Mining Practical Machine Learning Tools and Techniques third Edition, Morgan Kaufmann Publishers is an imprint of Elsevier, pp 978-0-12-374856-0

[3] Machine Learning for Dummies - John Paul Mueller, Luca Massaron - May 2016

[4] Machine Learning Using R - A Comprehensive Guide to Machine Learning - Karthik Ramasubramanian Abhishek Singh - 2017

[5] Understanding Machine Learning: From Theory to Algorithms by Shai Shalev-Shwartz and Shai Ben-David- Published 2014 by Cambridge University Press.

[6] coursera.org, ' Machine Learning', 2017. [Online]. Available: https://www.coursera.org/learn/machine-learning.

[7] Roberto Battiti Mauro Brunato "The LION Way Machine Learning Plus Intelligent Optimization" – 2014.

[8] Brendan Phibbs MD, The Human Heart: A Basic Guide to Heart Disease, Second Edition, University of Arizona College of Medicine, Tucson, Arizona, pp 978-0781767774.

[9] Introduction to the ROC (Receiver Operating Characteristics) plot, 2018, [Online]. Available: https://classeval.wordpress.com/introduction/introduction-to-the-roc-receiver-operating-characteristics-plot.

[10] Jason W. Osborne, Best Practices in Logistic Regression 1st Edition, University of Louisville, USA, pp 978-1452244792, 2015.

[11] Y. Lakshmi Prasad, Big Data Analytics Made Easy, pp 978-1-946390-71-4, 2015.

[12] Nick Pentreath, Machine Learning with Spark, pp 978-1-76326-651-9, 2015.

[13] American Heart Association, [Online]. Available: https://professional.heart.org/professional/ResearchPrograms/UCM_461443_AHA-Approved-Data-Repositories.jsp.

[14] Wahoo X Heart Rate Monitoring System (HRMS), [Online]. Available: URL: https://www.wahoofitness.com/devices/heart-rate-monitors.

Authors Profile

Mr.R.Kannan received Bachelor of Science from Madurai Kamaraj University in 1999, Master of Science from Bharathiyar University in year 2003 and Master of Philosopy from 2004. He is currently pursuing Ph.D in Department of Computer Science, Rathinam Colege of Arts and Scienc, Coimbatore, India since 2017. His main research work focuses on Machine Learning, Deep Learning, Big Data Analytics, Data Mining and Computational Intelligence based education. He has 3 years of teaching experience and 10 years of Industrial Experience.

Dr. V. Vasanthi pursed Ph.D from Karpagam University, Coimbatore, Tamil Nadu, India. She is currently working as Assistant Professor in Department of Computer Science, Rathinam Colege of Arts and Science, Coimbatore, India. She has published more than 20 research papers in reputed international journals including including IEEE and Springer and so on. it's also available online. Her main research work focuses on Machine Learning, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. She has 10 years of teaching experience and 7 years of Research Experience.

