JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON VARIOUS METHODS TO ENHANCE THE QUALITY OF HEALTH INFORMATICS USING BIG DATA

[1] Bi Bi Hafsa, [2] Mrs.Shreedevi S
[1] Student, [2] Asst Professor
[1] Dept. of Computer and Science Engineering, JNNCE,
[2] Dept. of Computer Science and Engineering, JNNCE, Shivamogga, India

Abstract: **In this era of big data, high volumes of a wide variety of valuable data can be easily generated or collected at a high velocity. Healthcare industry is an indispensable entity in the real world where large volumes of data is accumulated from time to time. So the Data world which is ruling us today not only established its identity in the field of computer sciencebut also other interdisciplinary sectors such as healthcare, economics, media and communication etc. This greatly limits thehealth providers' efforts to improve quality and efficiency.**

Index Terms **- Big Data, Big data in healthcare, Analytics opportunities, Methodology, Performance Analysis.**

I. INTRODUCTION

Big Data has gained a huge attention past few years, ever since the data is digitized. The information available over the distributed architecture has evolved instant accessibility across the globe. This helps in providing evidence based medicine to the patients instantly, using live analysis to simulate spreading or viral diseases, providing proactive care to the patients and reducing readmission rates to the hospital.(Manpreet Singh et.al 2017). Identifying and managing patients most at risk within the health care system is vital for governments, hospitals, and health insurers but they use different metrics for identifying the patients they perceive to be at most risk. Hospitals focus on re-admission rate and cumulative risk of death during hospitalization. Accurately predicting these indicators could assist in allocating limited resources and thus improve the hospital's operational efficiency. (Yang Xie et.al 2016) The demand for storing and processing large amounts of data is growing at faster rate due to rapid advancement in technology (Manish Kumar Pandey et.al 2016).

II. BIG DATA

Big Data has gained a huge attention from past few years, ever since the data is digitized. Big data analytics is playing a foremost role in healthcare sectors by helping them to progress patient-focused services, to detect and avoid diseases, to deliver better cure etc (Carson K. Leung, et.al ,2019). Big Data technologies provide instant solution and huge application to the biomedical problems. The instant service via prediction analysis proves out to better than the World Health Organisation (W.H.O.) and Centre for Disease Control (C.D.C.).(Manpreet singh et.al ,2017). In this era of big data, high volumes of a wide variety of valuable data can be easily generated or collected at a high velocity. Characteristics of these big data can be described by commonly known 3 V's, 5 V's, 7 V's, and even 42 V's (Carson k leung et.al, 2019). Big data analytics will help us to analyse and find the pattern between the data sets by the help of which we can improve the state of the current healthcare system.(Rahul kataraya et.al ,2020). Big data is making possible the huge tasks that were before impossible, like preventing disease spreading and crime, personalizing healthcare, quickly identifying business opportunities, managing emergencies, protecting the homeland, and so on. More reliable and powerful Clinical decision support systems (CDSS) are required to reduce the time required for diagnosis and increase the diagnosis accuracy.(Chitra Pasupathi ,2016).

III. BIG DATA ANALYTICS IN HEALTHCARE

It's about the importance and influences of big data analytics in healthcare commonly known as Healthcare Analytics.

Assisting Diagnosis - Big data analytics is used in identifying elongated diseases based on the symptoms interpretations, tests done and outcomes obtained.

Detect areas of scarcity in healthcare sectors - Big data analytics techniques can be used to detect areas of deficiency in healthcare sectors to expand healthcare services.

Sign of High-Risk Patients - Healthcare predictive analytics helps in smoothing running of the administrative operations in the hospitals, patient care situations, enhancing the management of prolonged diseases.

Avoiding Security Risks - Big data analytics can be used to learn historic data and recent fraud activities in the health industry and help them to prepare themselves to stop future likelihoods of being attacked by risks of similar type.

Better Customer Care Services - Healthcare companies today are providing quality care through portable devices such as health monitoring wristwatches, mHealth apps etc.(Praveen Kumar Sadineni et.al 2020).

IV. BIG DATA AND OPPORTUNITIES IN THE HEALTH SECTOR

Here are five practices where Big Data can support and change the Healthcare business sector entire route of practice.

A. Health Tracking - Big Data helps in the complex analysis of the vitalities of the patient such as sleep, heart rate, activity, etc. This monitoring provides important data on vital metrics such as blood pressure control, pulse rate and more that will enable an entity to achieve protective healthcare strategies.

B. Cost drop - Cost drop Predictive big data help to reduce medical expenditure by forecasting the level of a technique with staff allocation appreciation. The will help hospitals schedule their investments productively and reduce their income levy.

C. Focus on improving High-Risk Patients - Focusing on improving High-Risk Patients If all medical records are archived, the hospital will have a valuable information source for identifying the normal group of patients visiting the hospital andappreciating their ongoing issues(Sathya Balaji et.al 2020).

V. OBSTACLES TO EFFECTIVE QUALITY

Here are the challenges that healthcare organizations encounter hitting obstacles to effective big data execution-

- Research data collection
- Security and patient health information protection
- Data collected in legacy systems
- Resources cons traints for practical execution ((Sathya Balaji et.al 2020).

VI. RESEARCH METHODOLOGY

Data set and summary Statistics, Data manipulation and aggregation Predictive methods Performance measures (Yang Xie, Davidet.al 2016). The terms basically divided into two parts, [1] The first one is application cluster which is oriented as well load balancer

[2] The second one is storage cluster which is responsible for data transactions(Manish Kumar Pandey et.al 2016).Loading (image) and Insert into array [size] While{ decomposition up to Level N},Biorthogonal Wavelet Analysis, Calculate N, Hard Threshold (NLevel detailed coefficients) Coefficients Remove, Coefficients Reconstruction(Huaiyu Wen et.al,2016).This process is based on K- Means Algorithm [Clustering Process] (Chitra Pasupathi et.al 2016). Big Data technologies provide instant solution and huge application (The instant service via prediction analysis proves better than WHO and data from national health register (Manpreet Singh et.al, 2017).)There has been a growing interest in using smart home technologies for detecting human activity patterns for health monitoring applications. Detecting human activities in smart homes by means of analyzing smart meters data is studied in (Abdulsalam Yassine, Shailendra Singh 2017).Big Data in healthcare domain has great potential to help derive meaningful insights after analysis. Big Data in Healthcare can drive clinical decision support, disease surveillance, epidemic controland population health management (Minerva Panda et.al, 2017).

Some tools and techniques have been designed to visualize patterns involving sets of items (i.e., item sets) and related co-occurring entities (Carson K. Leung, Yibin Zhang et.al,2019).To examine this type of big dataset, the IT sector uses the Hadoop platform for a wide variety of methods that have been developed to record, organize, and analyse this type of data More efficient tools are needed to extract meaningful output from big data.(Sunil Kumar et.al, 2019). There has been a growing interest in using

smart home technologies for detecting human activity patterns for health monitoring applications. Detecting human activities in smart homes by means of analyzing smart meters data(Jun Su Lee, Lee Carroll et.al, 2019).

The Proposed smart recommender system in healthcare domain. They used multiple learning methods with a hybrid approach. Their methods include usage of collaborative and content-based filtering, usage of context of users and dynamic filtering(Muhib Anwar Lambay et.al 2020).On the research database different big data tools and algorithms are used to obtain the predictions.

These predictions will help to predict the problems and take the appropriate measures which will eventually reduce thecost (Rahul Katarya et.al, 2020).

VII. SYSTEM DESIGN AND IMPLEMENTATION

We began by importing the dataset into the Data-frame. Used the Pandas tool to clean the dataset before applying the pre-processingprocedures. The dataset is divided into training (70%) and testing (30%) datasets, and ML methods are applied to the divided (training & testing) dataset. Any supervised system can be evaluated using the train-test split technique.

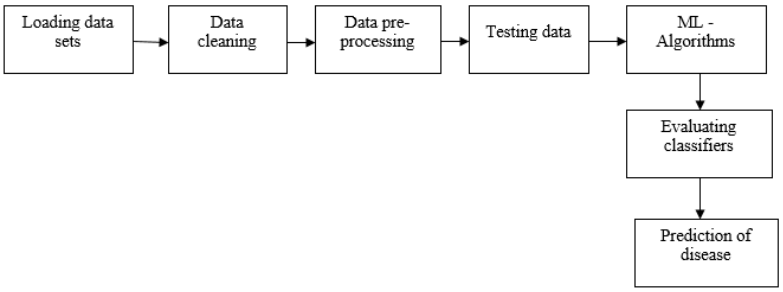


Figure 1: Block diagram of Heart Disease Prediction System

Stroke Disease Dataset attributes include name and description. It's used to address classification and regression difficulties. The dataset is separated into two subsets: the training dataset, which comprises the bulk of the data and is used in order to fit the Corresponding model, and the second, which contains the remaining data. The testing dataset is the second subset, which comprisesa smaller fraction of the data and is in such a way of used to assess the model's fit structure. Finally, the classification models wereevaluated to use the parameters Area Under ROC and Accuracy.

The following process of acquiring information about patients is known as dataset collection. Attribute selection mention to the process of identifying useful traits for the diagnosis of heart disease. Following the identification of accessible data assets, they arefurther picked, cleansed, and transformed into the appropriate form. Different classification approaches will be used on which wasbefore information to forecast the reliability of heart disease, as mentioned. The accuracy metric analyses the accuracy of several classifiers.

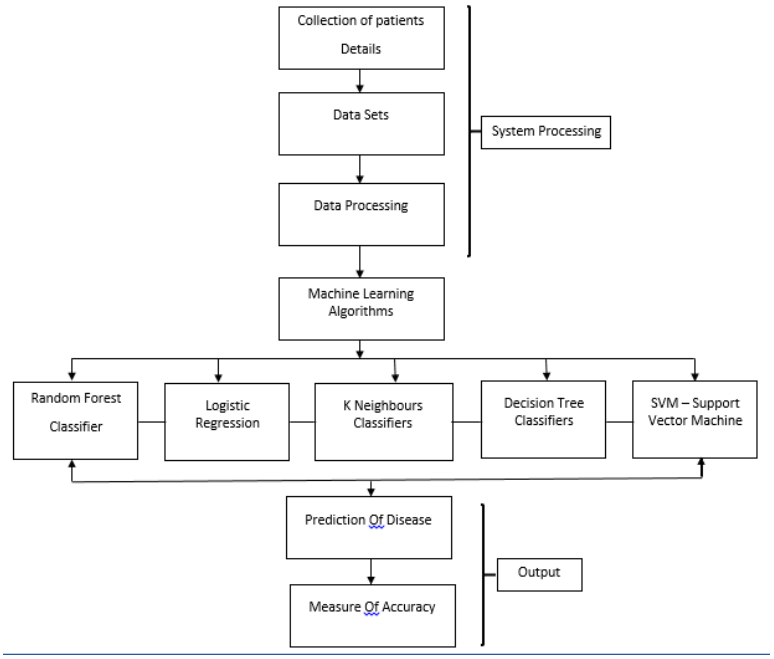


Figure 2: Proposed System

VIII. PERFORMANCE ANALYSIS

To predict cardiac disease, several ML techniques such as SVM, Decision Tree, Random Forest, Logistic Regression, and K Neighbours Classifiers are employed in this research. The Heart Disease UCI dataset has 76 variables in total, however only 14 of them are familiar heart disease. For this project, several patient characteristics such as gender, chest pain type, fasting blood pressure,serum cholesterol, exang, and so on are taken 31 into account. The accuracy of individual algorithms must be measured, and the algorithm with the highest accuracy is used to forecast heart disease. Various assessment measures like as accuracy, confusion matrix, precision, recall, and f1-score are used to evaluate the experiment. Accuracy- Accuracy is defined as the ratio of accurate predictions to total inputs in the dataset. Accuracy is defined as (TP + TN) /(TP+FP+FN+TN).

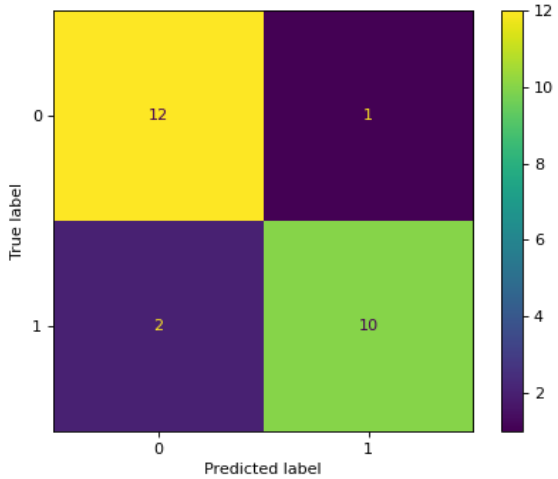


Figure 3: Confusion Matrix

Where : TP: True positiveFP: False Positive
FN: False NegativeTN: True Negative.
Correlation Matrix: In machine learning, the correlation matrix is used to pick features. It demonstrates the interdependence ofseveral properties.

IX. RESULTS AND ANALYSIS

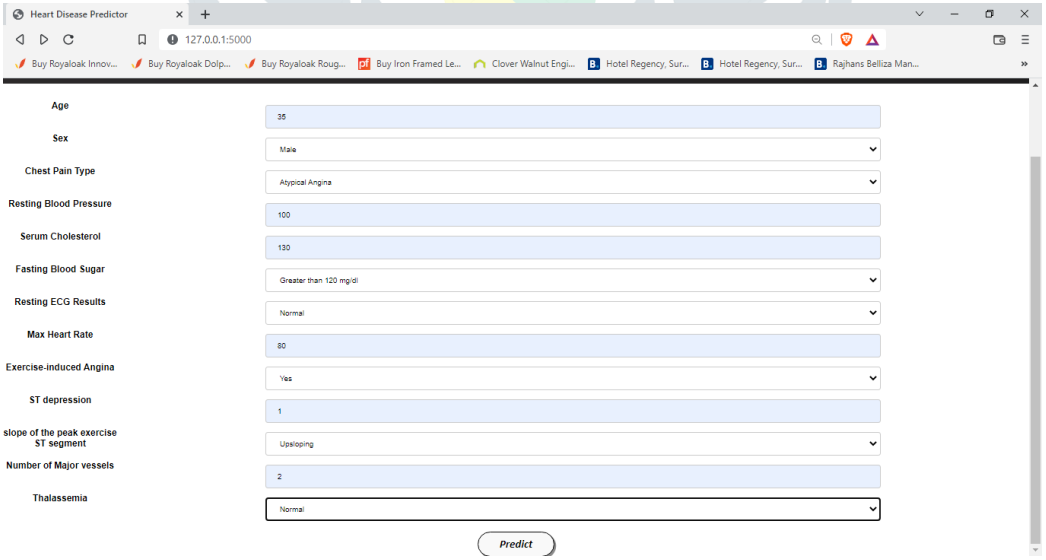


Figure 4: Describing example of Websites and inserting patient's details

The above figure - 4 which describes the websites of a Heart disease patients which it includes their particular details of the particular patient's.

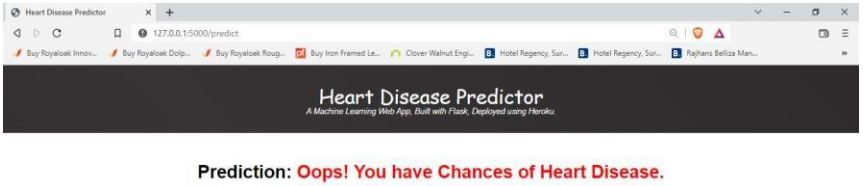
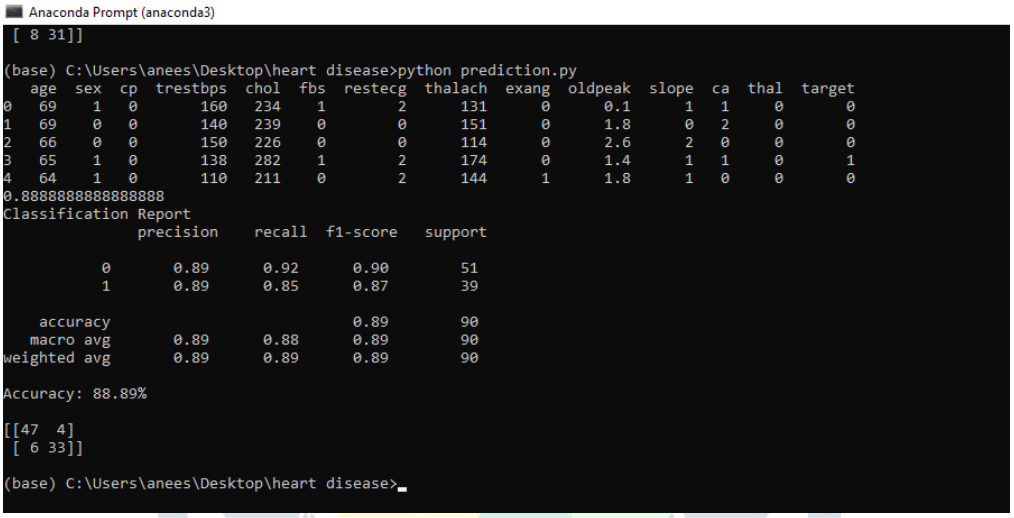


Figure 5: Predictions to be done as follows

In the above figure – 6, the following figure shows the prediction result of the patient through collecting their corresponding detailsand letting the patients know that they are having such disease or not.



```
Anaconda Prompt (anaconda3)
 [ 8 31]]

(base) C:\Users\anees\Desktop\heart disease>python prediction.py
   age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak  slope  ca  thal  target
0   69    1   0       160   234    1        2      131      0      0.1      1   1     0       0
1   69    0   0       140   239    0        0      151      0      1.8      0   2     0       0
2   66    0   0       150   226    0        0      114      0      2.6      2   0     0       0
3   65    1   0       138   282    1        2      174      0      1.4      1   1     0       1
4   64    1   0       110   211    0        2      144      1      1.8      1   0     0       0
0.8888888888888888
Classification Report
              precision    recall  f1-score   support

           0       0.89      0.92      0.90        51
           1       0.89      0.85      0.87        39

    accuracy                           0.89        90
   macro avg       0.89      0.88      0.89        90
weighted avg       0.89      0.89      0.89        90

Accuracy: 88.89%

[[47  4]
 [ 6 33]]

(base) C:\Users\anees\Desktop\heart disease>
```

Figure 6: The prediction table which shows Accuracy

In the above figure 6 represents the actual accuracy of the machine learning algorithm we used such as random forest Algorithm, Which it includes precision ,recall ,f1-score and support values as a classification report and presents a results in the form of confusionmatrix.

X. Presenting a comparison of Previous Accuracy and Present

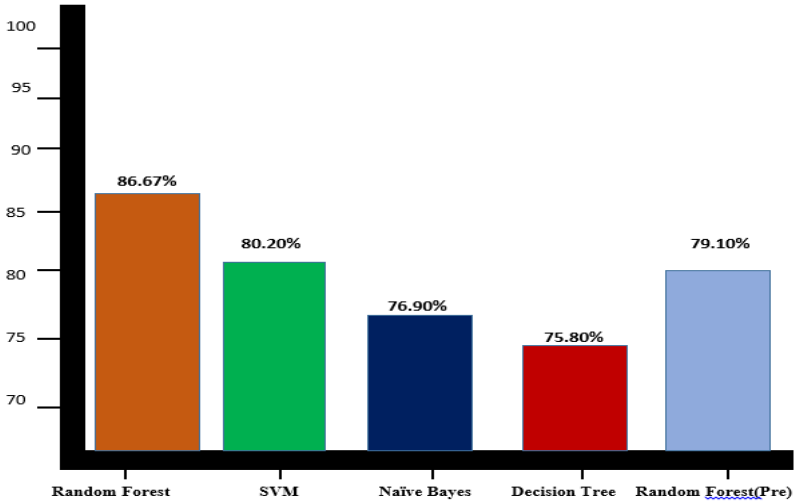


Figure 7: Presenting accuracy Chart.

In the above figure 7, the following graph is intrigued based on the analysis of different kind of previous experimental projects which is used by using different Machine Learning Algorithms like Random forest algorithm, SVM algorithm, Naïve Byes Algorithm, Decision Tree.

XI. CONCLUSION

Today Big Data Analytics is benefiting Healthcare companies more broadly. Enhancing the quality of Healthcare services help in delivering personalized solutions to distinctive problems such as predicting epidemics, treating diseases, improving the value of life, increasing avoidable care etc. The present paper concludes that integrating machine learning techniques with Big Data Analytics improves the quality of healthcare services.

XII. ACKNOWLEDGEMENT

I would like to thank my Guide Mrs. Sreedevi.S, Asst professor, Dept. of Computer Science and Engineering, JNNCE and the Department of Computer Science and Engineering, JNNCE, Shivamogga for their expertise and assistance throughout all aspects ofmy study and for their help, support and such a great Cooperation.

XIII REFERENCES

[1] Yang Xie, Student Member, IEEE, G¨unter Schreier, Senior Member, IEEE, David C.W. Chang, Sandra Neubauer, Ying Liu, Stephen J. Redmond, Senior Member, IEEE, Nigel H. Lovell, Fellow,IEEE " Predicting Days in Hospital Using Health Insurance Claims" – 2016.

[2] Manish Kumar Pandey and Karthikeyan Subbiah "A Novel Storage Architecture for facilitating EfficientAnalytics of Health Informatics Big Data in Cloud ",Department of Computer Science, Institute of Science,Banaras Hindu University, IEEE International Conference on Computer and Information Technology-2016.

[3] Huaiyu Wen, Fadia Shah, Jian ping Li, Raheel Ahmed Memon, Faiza Shah4 "X-Ray ImageReliability Using Biorthogonal Wavelet Compression For Medical Big Data" 1institute Of Economic Information Engineering, Chengdu University -2016.

[4] Diagnosis Chitra Pasupathi and Vijaya Kalavakonda" Evidence Based Health Care System Using BigData for Disease", Prathyusha Engineering College, India, RMK Engineering College, India -2016.

[5] Manpreet Singh, Vandan Bhatia, Rhythm Bhatia,"BIG DATA ANALYTICS-Solution to Healthcare",2017 International Conference on Intelligent Communication and Computational Techniques (ICCT).

[6] Shailendra Singh, Atif Alamri, Abdulsalam Yassine "Mining Human Activity Patterns from Smart HomeBig Data for Healthcare Applications" Member, IEEE, Member, IEEE-2017.

[7] Minerva Panda, Syed Mohd Ali, Sanjog Kumar Panda, "Big Data in Health Care: A Mobile BasedSolution" Department of Computer Science and Engineering, IIIT Bhubaneswar, India 2017.

[8] Carson K. Leung*, Yibin Zhang, Calvin S.H. Hoi, Joglas Souza, and Bryan H. Wodi," Big Data Analysis and Services: Visualization of Smart Data to Support Healthcare Analytics", 2019 International Conferenceon Internet of Things.

[9] Sunil Kumar_ and Maninder Singh, "Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools", BIG DATA MINING AND ANALYTICS" - ISSN 2096-0654 – March 2019.

[10] Jun Su Lee, Lee Carroll, Carmelo Scarcella, Nicola Pavarelli, "Meeting the Electrical, Optical, and Thermal Design Challenges of Photonic-Packaging", IEEE Journal Of Selected Topics In Quantum Electronics, Vol. 22, No. 6, November/December 2016

[11] Muhib Anwar Lambay, Dr. S. Pakkir Mohideen "Big Data Analytics for Healthcare Recommendation Systems", Research Scholar & Associate Professor ,Department of Computer Science & Engineering, Crescent Institute of Science & Technology, Chennai, India- 2020.

[12] Rahul Katarya, Sajal Jain, "Exploration of Big Data Analytics in Healthcare Analytics", 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP).