Heart Disease Prediction System

Department of Computer Engineering

G H Raisoni College of Engineering and Management, Pune, India

Ms. Shubham Bapusaheb Bhavar

G H Raisoni College of Engineering

& Management, Pune

Department of Computer Engineering

Mr. Satish Shriniwas Bankar Subhedar

G H Raisoni College of Engineering Engineering

& Management, Pune

Department of Computer Engineering Engineering

Mr. Salman Allauddhin

G H Raisoni College of

& Management, Pune

Department of Computer

Ms. Shubham Shivraj Golegavkar

G H Raisoni College of Engineering

& Management, Pune

Department of Computer Engineering

Mrs. Indhra Muthuswamy

Assistant Professor, Dept.

of Computer Engineering

G H Raisoni College of Engineering

& Management, Pune

ABSTRACT*:* In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person instantly that might create unexpected consequences for the human life. So diagnosing patients correctly on time is the most challenging task for the medical fraternity. The heart disease treatment is quite high and not affordable by most of the patients particularly in India. The research scope is to develop an early prediction treatment using data mining technologies. Nowadays every hospital keeps the periodical medical reports of cardiovascular patients through some hospital management system to manage their health-care. The data mining techniques namely decision tree and random forest are used to analyze heart attack dataset where classification of more common symptoms related to heart attack is done using c4.5 decision tree algorithm, alongside, random forest is applied to improve the accuracy of the classification result of heart attack prediction. In this system various data mining technologies are applied to make a proactive approach against failures in early predictions diagnosis of the disease. We proposed an automated system for medical diagnosis that would enhance medical care and reduce cost. Our aim is to provide a ubiquitous service that is both feasible, sustainable and which also make people to assess their risk for heart attack at that point of time or later.

Index Terms— Storke prediction, Random forest algorithm, KNN, ANN, C4.5 algorithms.

KEYWORDS: Data Mining, Cardiovascular Disease (CVD), Heart Attack, Machine Learning.

I. INTRODUCTION

Data mining has already established as a novel field for exploring hidden patterns in the huge datasets. Medical science is another field where large amount of data is generated using different clinical reports and other patient symptoms. Data mining can also be used heavily for the same purpose in medical datasets also. These explored hidden patterns in medical datasets can be used for clinical diagnosis. However, medical datasets are widely dispersed, heterogeneous, and huge in nature. These datasets need to be organized and integrated with the hospital management systems. Cardiovascular diseases are one of the highestflying diseases of the modern world1. According to world health organization about more than 12 million deaths occurs worldwide, every year due to heart problems. It is also one of the fatal diseases in India which causes maximum casualties. The diagnosis of this disease is intricate process. It should be diagnosed accurately and correctly. Due to limitation of the potential of the medical experts and their unavailability at certain places put their patients at high risk. Normally, it is diagnosed using intuition of the medical specialist. It would be highly advantageous if the techniques will be integrated with the medical information system.

II. REVIEW OF LITERATURE

Aditi Gavhane, Isha Pandya ``Prediction of Heart Disease Using Machine Learning'' IEEE Conference Record 2018 IEEE

The Heart Disease Prediction System using Machine learning algorithm, viz. MLP provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence author use Multi Layered Perceptron (MLP) in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users.

S. Kiruthika Devi, S. Krishnapriya and Dristipona Kalita ``Prediction of Heart Disease using Data Mining Techniques'' Indian Journal of Science and Technology, Vol 9(39), October 2016

The accuracy of the algorithms used in each technique can be enhanced by hybridizing or combining algorithms to a single algorithm which may not be accurate for weakly classified sets of data, and is expected to make quicker and more precise decisions.

M. Lavanya, Mrs. P. M. Gomathi, ``Prediction of Heart Disease using Classification Algorithms'' International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 7, July 2016
Medical related information are huge in nature and it can be derived from different birthplaces which are not entirely applicable in feature. The research undertook an experience on application of varies data mining algorithm to predict the heart attacks and to compare the based method of prediction. The predictive accuracy determined by J48, REPTREE, naive bayes, neural networks, CART. The overall objective is to study the various data mining techniques available to predict the heart.

Hlaudi Daniel Masethe, Mosima Anna Masethe, ``Prediction of Heart Disease using Classification Algorithms'' WCECS 2014, 22-24 October, 2014, San Francisco, USA
The research undertook an experiment on application of various data mining algorithms to predict the heart attacks and to compare the best method of prediction. The research results do not presents a dramatic difference in the prediction when using different classification algorithms in data mining. The experiment can serve as an important tool for physicians to predict risky cases in the practice and advise accordingly. The model from the classification will be able to answer more complex queries in the prediction of heart attack diseases. The predictive accuracy determined by J48, REPTREE and SIMPLE CART algorithms suggests that parameters used are reliable indicators to predict the presence of heart diseases.

Abhishek Ta, ``Heart Disease Prediction System Using Data Mining Techniques''
In this study, the aim was to design ma predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset that is capable of enhancing the reliability of heart disease diagnosis using echo-cardiography.

Sellappan Palaniappan, Rafiah Awang ``Intelligent Heart Disease Prediction System Using Data Mining Techniques''
A prototype heart disease prediction system is developed using three data mining classification modeling techniques. The system extracts hidden knowledge from a historical heart disease database. DMX query language and functions are used to build and access the models. The models are trained and validated against a test dataset. Lift Chart and Classification Matrix methods are used to evaluate the effectiveness of the models. All three models are able to extract patterns in response to the predictable state. The most effective model to predict patients with heart disease appears to be Naive Bayes followed by Neural Network and Decision Trees.

Aditya Methaila, Prince Kansal,``EARLY HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES''
In this paper the focus is on using different algorithms and combinations of several target attributes for effective heart attack prediction using data mining. Decision Tree has outperformed with 99.62% accuracy by using 15 attributes. Also the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction. Association classification technique apriori algorithm, was along with a new algorithm MAFIA was used. Straight Apriori-based algorithms count all of the 2k subsets of each k-item set they discover, and thus do not scale for long item sets. They use ``look a heads'' to reduce the number of item sets to be counted. MAFIA is an improvement when the item sets in the database are very long.

III. EXISTING SYSTEM

In Existing PHR system data is outsourced on data server over internet. For privacy and security of information, encryption techniques and Cryptography is used to provide privileges. For smoother access privileges, the traditional public key encryption (PKE) may costs excess key management overhead and scalability issue drawbacks associated with existing system, which include:

- The high upfront acquisition costs
- Ongoing maintenance costs, and
- Disruptions to work-flows that contribute to temporary losses in productivity.

IV. PROPOSED SYSTEM

In this paper, Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. In this system using a android application user can uploads his/her health record in cloud storage. In cloud storage first user data is processed then training and testing data apply classification C4.5 algorithm to testing data and predict heart attack as shown in figure 1. The application of machine learning and data mining techniques shows it provides better performance results. This app will predict the heart attack from the chest pain at an early stage and guide the person to take treatment early such as to get their ECG done as early as

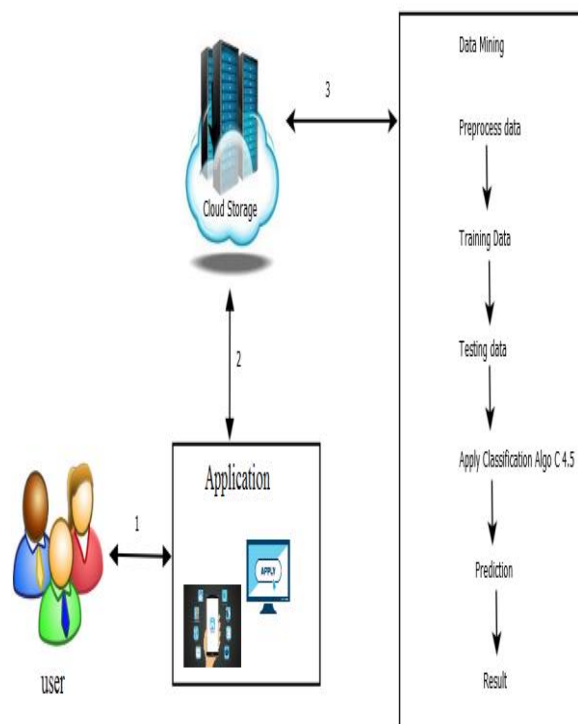possible and get evaluated from a doctor to make diagnoses of heart attack.

Figure 1. System architecture
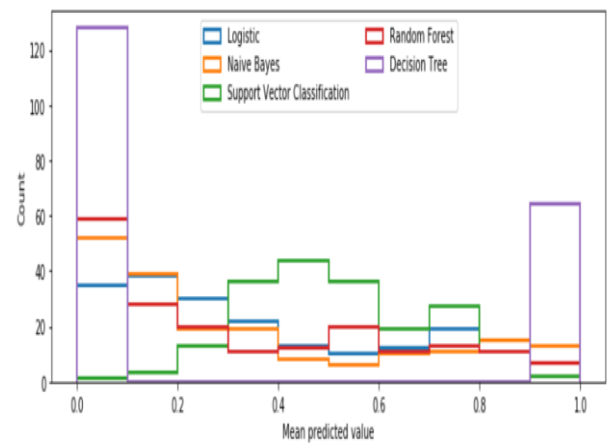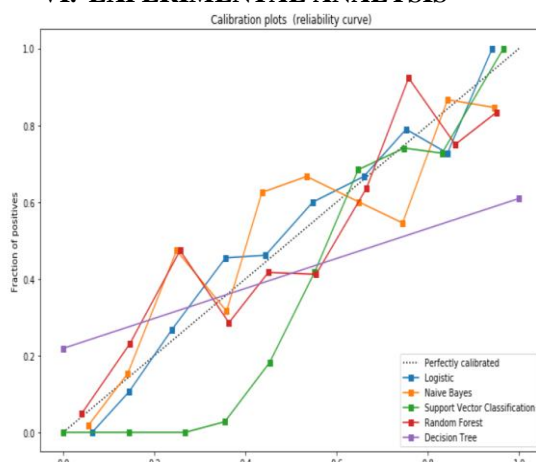
V. SYSTEM REQUIREMENTS

A. Software Requirement
1. Operating System: Windows 7 or above
2. Programming Language: Python 3.7
3. IDE: Python IDLE

B. Hardware Requirement
1. Processor: Pentium Processor Core 2 Duo or Higher
2. Hard Disk: 250 GB (min)
3. RAM: 1GB or higher
4. Processor Speed: 3.2 GHz or faster processor

VI. EXPERIMENTAL ANALYSIS

VII. CONCLUSION

The symptoms of heart disease are familiar for everyone but no one can predict when it may happen and also the prediction will not give any instance status i.e. normal or abnormal. Our research intention is to notify the patients that neither normal nor abnormal at its early stage. The application of data mining algorithms shows it better performance results among themselves. This app will predict the heart attack any obvious symptoms like chest pain and guide the person to take treatment as early as possible.

In future along with heart attack prediction other life disease like cancer can also be included by adding few more parameters of the patient.

REFERENCES

1. Stefano Bromuri , Serban Puricel, Rene Schumann , Johannes rampf,Juan Ruiz and Michael Schumacher, -An expert Personal Health System to monitor patients affected by Gestational Diabetes Mellitus: A feasibility study,— Journal of Ambient Intelligence and Smart Environ-ments 8 (2016) 219237G [1].

2. Gyorgy J. Simon,Member, IEEE,Pedro J. Caraballo, Terry M. Therneau, Steven S. Cha,M. Regina Castro and Peter W. Li ,-Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus,—IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEER-ING [2].

3. Han Wu, Shengqi Yang , Zhangqin Huang, Jian He, Xiaoyi Wang,- Type 2 diabetes mellitus prediction model based on data mining,— Informatics in Medicine Unlocked 10 (2018) 100107 [3].

4. Xue-Hui Meng a, Yi-Xiang Huang a, Dong-Ping Rao b, Qiu Zhang a, Qing Liu b, - Comparison of three data mining models for predicting diabetes or prediabetes by risk factors,— Kaohsiung Journal of Medical Sciences (2013)29, 93e99 [4].

5. Kung-Jeng Wang, Angelia Melani Adrian a, Kun-Huang Chen a, Kung-Min Wang b,- An improved electromagnetism-like mechanism algorithm and its

application to the prediction of diabetes mellitus,— Journal of Biomedical Informatics 54 (2015) 220229 [5].

6. Jinn-Yi Yeh, Tai-Hsi Wu b, Chuan-Wei Tsao, - Using data mining tech-niques to predict hospitalization of hemodialysis patients,— Decision Support Systems 50 (2011) 439448 [6].

7. Nongyao Nai-arun, Rungruttikarn Moungmai , - Comparison of Classifiers for the Risk of Diabetes Prediction,— 7th International Conference on Advances in Information Technology [7].

8. Asha Gowda Karegowda , M.A. Jayaram, A.S. Manjunath, - Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients,— IJEAT Volume-1, Issue-3, February 2012 [8].

9. Darshan K R, Anandakumar K R, A Comprehensive Review on Us-age of Internet of Things (IoT) in Healthcare System — 2015 IEEE ICERECS&T [9].

10. Chih-Hua Tai, Daw-Tung Lin, A Framework for Healthcare Every-where: BMI Prediction using Kinect and Data Mining Techniques on Mobiles, 2015 16th IEEE International Conference on Mobile Data Management[10].

11. Ritika Chadha , Shubhankar Mayank, Prediction of heart disease using data mining techniques, 2016 OF CSIT Springer [11].

12. Eleni I. Georga, Vasilios C. Protopappas, Stavroula G. Mougiakakou,, Short-term vs. Long-term Analysis of Diabetes Data: Application of Machine Learning and Data Mining Techniques, IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 2003 [12].