International Journal of Innovative Research in Technology and Management (IJIRTM), Volume-2, Issue-1, 2018.
www.ijirtm.com *ISSN: 2581-3404 (Online)*

A Survey on Health Care using Machine Learning Techniques

Faizan Ahmed Siddiqui[1], Dr. Rachana Dubey[2]
[1]M. Tech Scholar, Department of CSE, LNCT, Bhopal (India)
[2]Professor, Department of CSE, LNCT, Bhopal (India)
[1]fezsid@gmail.com, [2]rachanamishra812@gmail.com,

ABSTRACT
Machine Learning techniques are used to explore, analyze and extract medical data using complex algorithms in order to discover unknown patterns. In this study aimed at compare the performance analysis of various methods which is based on data mining techniques for the clinical support system and various types of diseases, developing an intelligent data mining system based on genetic algorithm optimized neural networks for the prediction of heart disease based on risk factors categories, for the implementation work we used the dataset from the UCI machine learning repository which include huge amount of dataset for the health care such as heart, diabetes, liver, breast cancer and statlog dataset etc.

Keywords: Machine Learning, UCI, Health Care, WHO, Ensemblling.

INTRODUCTION
Medical data mining and knowledge discovery is a relatively young and growing field that attracts many researchers [3]. Even though there are challenges that hinder the mining process, disease diagnosis is still an ever green field of study. The data obtained from the patients is the resource for medical data mining process [7]. Data mining or knowledge discovery in databases can be defined as the process for discovering the hidden patterns from amounts of data.

Data mining can be applied in various industries such as healthcare, risk detection, stock market, data analytics and prediction, traffic monitoring and fraud detection. In recent years, data mining has received much attention of health scholars and professionals. Some

benefits of data mining in health industries include providing fast and practical solutions to the patients at a lower cost, detection of illness caused and recommending the medical treatment methods, and building drug recommendation systems [2]. Data mining is the search for relationships and global patterns that exist in large databases, but hidden among the large amounts of data. Computer diagnosis of diseases are the doctor for the same instrument, the calculations for an engineer: design diagnostics does not replace the doctor, but it helps [12].

Therefore, important to develop mathematical methods of diagnostics and compare their effectiveness Artificial Intelligence can enable the computer to think. Computer is made much more intelligent by AI [15]. Machine learning is the subfield of AI study various researchers think that without learning, intelligence cannot be developed. There are many types of Machine Learning Techniques that are shown in below figure: 1[4].

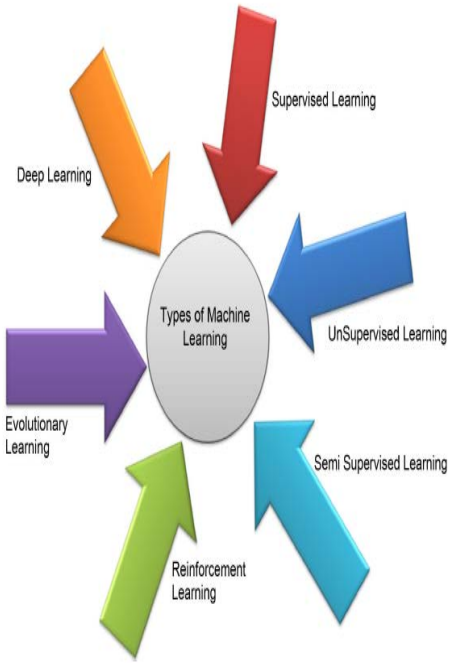

Figure 1: Types of machine learning techniques [4].

The above figure shows about the types of machine learning techniques which are basically supervised learning such as various classifier, up-supervised learning for example various clustering and types of clustering techniques, semi-supervised learning which is a combination of clustering and classification, Evolutionary learning such as genetic algorithm and deep learning etc. Data mining techniques are helpful in extracting & analyzing the complicated medical data using various kinds of techniques [4]. Practioners of medical science are also using these minute techniques in other field such as detection of cancer and stroke. Researchers have been applying various techniques of machine learning such as Artificial Neural Network, Back-Propagation neural network, decision tree classifier, ensemblling techniques; self organized map network and genetic algorithm for the optimization purpose [1]. Random Forest is a group of un-pruned classification or it is a regression trees which are made from the random selection of samples of the training data [3].

Diagnosis of Diseases by Using Different Machine Learning Algorithms Many researchers have worked on different machine learning algorithms for disease diagnosis. Researchers have been accepted that machine-learning algorithms work well in diagnosis of different diseases. Figurative approach of diseases diagnosed by Machine Learning Techniques is shown in below figure 2. In this survey paper diseases diagnosed by MLT are heart, diabetes, liver, dengue and hepatitis.

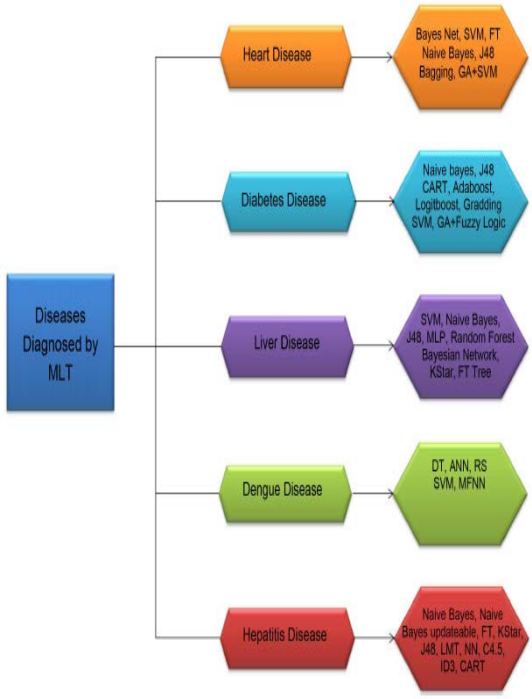

Figure 2: Diseases diagnosed by MLT [4].

The above figure shows the diseases diagnosis detection by the using machine learning techniques there are various types of diseases detection such as heart, diabetes, liver and dengue etc. all the diseases dataset are collected from the UCI machine learning repository for the research purpose. All the disease are human related and various researchers and organization currently doing the research for the same.

The rest of this paper is organized as follows in the first section we describe a introduction of about data mining and machine learning techniques for health care and specially heart diseases. in section II we discuss about the rich literature survey for the data mining and machine learning techniques for the heart diseases. In section III we discuss about the problem formulation and statement as we getting from the rich literature survey, In section IV we discuss about the dataset and their types, finally in section V we conclude the about our paper which is based on the literature survey and specify the future scope.

II RELATED WORK

In this section we discuss about the rich literature survey for the health care using machine learning and other techniques**.** Ankita Dewan and Meghna Sharma Et al. they develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. It can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions. S.Leoni Sharmila, C.Dharuman, P.Venkatesan Et al. gives a comparative study of different machine learning technique such Fuzzy logic, Fuzzy Neural Network and decision tree in classifying liver data set. The main aim of the classification techniques is to analyze the input data and to predict the accuracy for the future work. Ming Yuchi and Jun Jo et al. proposes a HR prediction model based on the relationship between HR and PA. The predictor has the potential to be used in various areas, such as: cardiopathy research and diagnosis, heart attack warning indicator, sports capability measure and mental activity evaluation. Basheer Mohamad Al-Maqaleh, Ahmed Mohamad Gasem Abdullah et al. they proposed an intelligent predictive system using classification techniques for heart disease diagnosis, namely, J48 decision tree, Naïve Bayes and Multi-Layer Perceptron Neural Network (MLPNN) are proposed. The main objective of this paper is to study these classification techniques to predict the heart disease and find the best technique of prediction. S.Poongothai, C.Dharuman and P.Venkatesan et al. proposes a comparison of fuzzy and hybrid techniques

like fuzzy genetic and neuro-genetic for the diagnosis of diabetes on a Pima Indian Diabetes dataset from UCI machine learning repository. Diabetes is one of the major challenging disease to the world.

III PROBLEM STATEMENT

Clinical decisions are often made based on doctors intuition and experience rather than on the knowledge rich data hidden in the database. [8] This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The word, Heart disease encompasses the diverse diseases that affect the heart. Heart disease is the major cause of casualties in the world. The various forms of cardiovascular disease are High blood pressure, coronary artery disease, valvular heart disease, stroke or rheumatic fever/rheumatic, heart diseases. As per the estimation of the World Health Organization (WHO), there are twelve millions of deaths occurred worldwide every year due to the cardiovascular diseases [13]. The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using the data mining modeling techniques and machine learning techniques and Ensemble learning combines homogenous machine learning approaches using various merging methods whose goal is to achieve enhanced accuracy over single classifier model. The Intelligent Heart Disease Prediction System can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database [19]. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot [25].

IV DATASET DESCRIPTION

In this section we discuss about the dataset which we used for the diseases detection in the field of health care. There are all The dataset types will be fetched from the UCI machine learning repository for the research purpose. In future we implement the diseases detection and improve the accuracy and other performance evaluation parameter with the help of all these dataset using machine learning techniques.

Breast cancer

This breast cancer domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia [16]. This data set includes 201 instances of one class and 85 instances of another class. The instances are described by 9 attributes, some of which are linear and some are nominal. They include attributes are class, age, menopause, tumor-size, inv-nodes, node-caps, dag-malig, breast, breast quad and irradiat.

Diabetes

Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records. The automatic device had an internal clock to timestamp events, whereas the paper records only provided "logical time" slots (breakfast, lunch, dinner, bedtime). There are number of dataset attributes are 20 which are categorized on the basis of multivariate and time series [16].

Statlog (Heart) Dataset

This database contains 13 attributes (which have been extracted from a larger set of 75. The total 13 Attribute whish are age, sex, chest pain type (4 values, resting blood pressure, serum cholestoral in mg/dl, fasting blood sugar > 120 mg/dl, resting electrocardiographic results (values 0,1,2), maximum heart rate achieved, exercise induced angina, oldpeak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels (0-3) colored by flourosopy, thal: 3 = normal; 6 = fixed defect; 7 = reversable defect [16].

LIVER DATASET

The first 5 variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each line in the bupa data file constitutes the record of a single male individual. It appears that drinks>5 is some sort of a selector on this database. See the PC/BEAGLE User's Guide for more information.

V CONCLUSIONS AND FUTURE WORK

This comparative study classification processes in the diagnosis of patients that can suffer from Heart Disease. These techniques such as data mining and machine learning were compared and the best results obtained came from the all these method based on the selected metrics or some performance parameters. In future we plan to implement a model based on some machine learning techniques and ensemblling model for the diseases with reduce to error and risk factor and improve the accuracy.

REFERENCES:-

[1] Ankita Dewan and Meghna Sharma “Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification”, IEEE, 2015, Pp 704-706.

[2] Seyed Mohammad Jafar Jalali, Sergio Moro, Mohammad Reza Mahmoudi, Keramat Allah Ghaffary “A Comparative Analysis of Classifiers in Cancer

Prediction Using Multiple Data Mining Techniques" ISCTE-IUL, 2017. Pp 1-12.

[3] S.Leoni Sharmila, C.Dharuman, P.Venkatesan "Disease Classification Using Machine Learning Algorithms - A Comparative Study" International Journal of Pure and Applied Mathematics, Vol-114, 2017, Pp 1-10.

[4] Meherwar Fatima, Maruf Pasha "Survey of Machine Learning Algorithms for Disease Diagnostic" Journal of Intelligent Learning Systems and Applications, 2017, Pp 1-16.

[5] Tabreer T, Hasan, Manal H, Jasim and Ivan A. Hashim " Heart Disease Diagnosis System based on Multi-Layer Perceptron neural network and Support Vector Machine" International Journal of Current Engineering and Technology, Vol-7, 2017.

[6] Ming Yuchi and Jun Jo "Heart Rate Prediction Based on Physical Activity using Feed forward Neural Network", International Conference on Convergence and Hybrid Information Technology 2008International Conference on Convergence and Hybrid Information Technology, 2008, Pp 344-350.

[7] R.Sujatha, E.P.Ephzibah, Sree Dharinya, G. Uma Maheswari, V.Mareeswari Vamsi Pamidimarri " Comparative study on dimensionality reduction for disease diagnosis using fuzzy classifier" International Journal of Engineering & Technology, 2018, Pp 79-84.

[8] Basheer Mohamad Al-Maqaleh, Ahmed Mohamad Gasem Abdullah "Intelligent Predictive System Using Classification Techniques for Heart Disease Diagnosis" International Journal of Computer Science Engineering, 2017. Pp 145-151.

[9] M. S. A. Megat Ali and A. H. Jahidin "Hybrid Multilayered Perceptron Network for Classification of Bundle Branch Blocks", IEEE., 2011, Pp 149-154.

[10] Arijit Ukil, Soma Bandyopadhyay, Chetanya Puri and Arpan Pal "HEART-TREND: AN AFFORDABLE HEART CONDITION MONITORING SYSTEM EXPLOITING MORPHOLOGICAL PATTERN", IEEE, 2016, Pp 6260-6264.

[11] Theresa Princy. R and J. Thomas "Human Heart Disease Prediction System using Data Mining Techniques", ICCPCT, 2016, Pp 1-5.

[12] Aigerim Altayeva, Suleimenov Zharas and Young Im Cho "Medical Decision Making Diagnosis System Integrating k-means and Naïve Bayes algorithms", ICCAS, 2016, Pp 1087-1092.

[13] Rahul Joshi, Minyechil Alehegn " Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach" International Research Journal of Engineering and Technologyol-4, 2017. Pp 426-435.

[14] S.Poongothai, C.Dharuman and P.Venkatesan "A Comparison of Fuzzy Genetic and Neuro Genetic Hybrid Algorithm for the Classification of Diabetes Disease" International Journal of Pure and Applied Mathematics Volume 113, 2017, Pp 208-216.

[15] M.Akhil jabbar , Priti Chandra and B.L Deekshatulu "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection", IEEE, 2012, Pp 628-634.

[16] https://archive.ics.uci.edu/ml/index.php