A review on Medical Disease Diagnosis using Evolutionary and Classification Techniques

Amita Vishwakarma[1], Prof. Pooja Meena[2], Prof. Chetan Agrawal[3]

[1]M. Tech Scholar, Department of CSE, RITS, Bhopal (India)

[2]Professor, Department of CSE, RITS, Bhopal (India)

[2]Head & Professor, Department of CSE, RITS, Bhopal (India)

[1]amitavishwa06@gmail.com, [2]meena.pooja1@gmail.com, [3]chetan1nonly@gmail.com

ABSTRACT

Now a day's healthcare organization faces major challenges in the provision of cost, quality of services, patient's life decision, and detection of any diseases at early stages. These days many hospital generate huge amount of patient's data for many diseases, which play a key role for the treatment of diseases and get recovered from its. In this paper we present the survey for the health diseases diagnosis using evolutionary and other classification algorithm.

Keywords: - Evolutionary algorithm, Decision Tree, Neural Network, Feed Forward Back Propagation, Classification, Optimization.

INTRODUCTION

Used of computerization which generate huge amount of data for almost every field such as education, medical science, social service, marketing, business, e-commerce and sports etc., medical science is one of the very important for us because its directly related to human life. In this field computerized data for medical field is very-2 important for the treatment for any patients diseases. There are various diseases such as cancer, heart diseases, diabetes etc.

The heart disease is the leading cause of death worldwide and the diagnosis is a complicated task that requires both experience and knowledge [1]. Medical datasets contain a wealth of hidden information that can be important in making decision. Data mining is a knowledge discovery process to analyze data and extract useful information for effective decision-making. Data mining can be a useful tool in healthcare.

Multi-objective evolutionary algorithm (MOEAs), also known as multi-objective optimization algorithms (MOOAs), is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints; they are a population based search.

Neural networks are an approach to computing that involves developing mathematical structures with the ability to learn. The methods are the result of academic investigations to model nervous system learning [5]. Neural network is a types of classifier which is a classified the data value on the basis of pattern matching, and produce the results on the basis of their matched pattern of input data. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze [3].
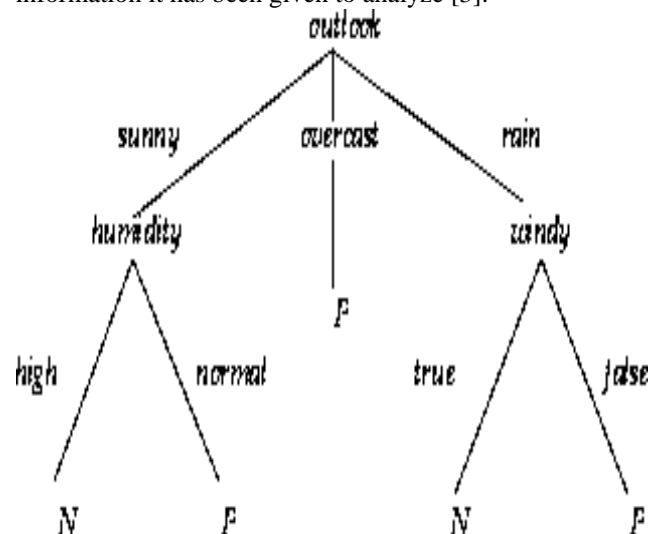

Figure 1: Basic algorithm for inducing a decision tree from training tuples.

The rest of this paper is organized as follows in section II we discuss about the dataset. In section III we discuss about the rich literature survey for the various diseases diagnosis using data mining and optimization techniques. In section IV we present a problem statement. In section V we discuss about the previous comparative study for the various techniques and authors. In section VI we shows a future scope for this research work, And finally In section VII we conclude the about our paper which is based on the whole literature survey journey also discuss the future scope.

II DESCRIPTION OF DATASET

In this section we discuss about the dataset which we used for the diseases detection in the field of health care. There are all these 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 data mining, optimization techniques.

Cancer Dataset

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.

Heart Dataset

This database contains 13 attributes (which have been extracted from a larger set of 75.

S.N	Attribute Name	Description
1	Age	Age in years
2	Sex	Male=1, Female=0
3	Cp	Chest pain type
4	Trestbps	Resting Blood pressure upon hospital nadmission
5	Cholestrol	Serum Chol
6	FBS	Fasting blood sugar>120 (true=1,false=0)
7	RestingECG	Resting Electrocardiographic Results
8	Thalach	Maximum Heart Rate
9	Exang	Does the patient experience angina as a result of exercise (value 1:yes, value 0:no)
10	Old peak	ST depression indeced by exercise relative to rest
11	Slope	Slope of the peak exercise
12	Thal	Value 3: Normal, value 6: fixed defect, value 7: reversible defect
13	CA	Number of major vessels colored by fluoroscopy (value 0-3)
14	ConceptClass	Angiographic disease status

Table 1: Description of heart Dataset attributes.

III RELATED WORK

In this section we discuss the about rich literature survey for the Intrusion detection techniques based on the various research paper which are highly cited from various reputed organization such as IEEE transactions, Elsevier, Springer and other international journal research papers. In the next section we also discuss the survey in the tabular format.

There are various authors who used the diseases diagnosis using various clustering and classification algorithm, some author also used evolutionary algorithm to find the accurate accuracy and other parameters which are related to patients for the any diseases. In our base papers author used a types of neural network classifier i.e. feed forward back propagation neural network with the optimization methods i.e. particle of swarm optimization and find the better accuracy than other existing techniques such as other classification techniques for example decision tree, association rule etc.
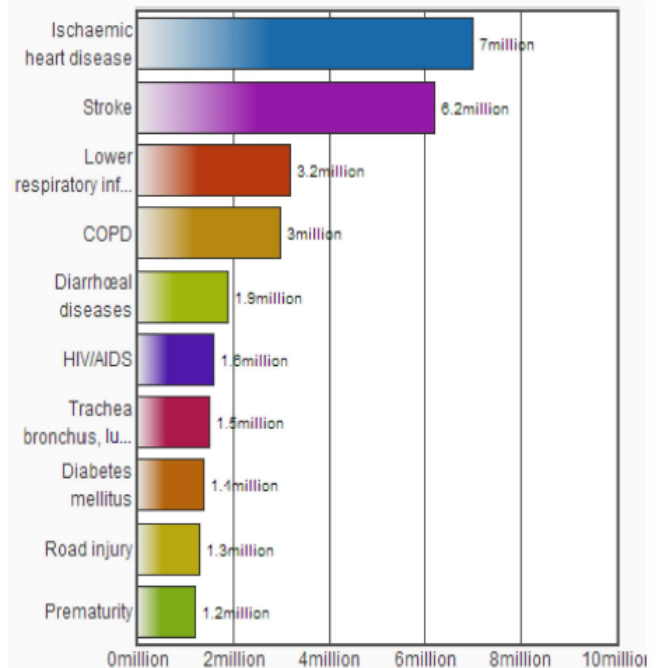

Figure 2: Ten leading causes of death in the world over the past decade [1].

IV PROBLEM STATEMENT

For the purpose of medical disease data classification various machine learning algorithm are applied, such as clustering, classification, and regression. Two of the most critical and well generalized problems of medical data are its new evolved feature and concept-drift. Since a medical data is a fast and continuous event, it is

assumed to have infinite length. Therefore, it is difficult to store and use all the historical data for training. Sometimes the field also suffer from the less accuracy due to generated from various classifier which is not provide good results however we can enhance the result accuracy using some optimization methods also.

V COMPARATIVE STUDY

In this section we discuss the comparative study for the data mining and machine learning techniques using in the medical diseases diagnosis in medical science. In the previous paper medical diseases diagnosis author improve the accuracy or detection rate using particle swarm optimization and feed forward back propagation neural network. Some author earlier also used data mining techniques for the detection of diseases at early stage and save the life of patients.

Authors	Algorithm	features	accuracy	feature ranking
Panday and Godara [5], 2012	Improved MLP	13	%82.8	No
Subbalakshmi., et al [13],2011	Naïve Bayes	15	%81	No
Nahar., et al [7], 2013	SMO	13	%86	No
Kumari and Godara[8], 2011	SVM	13	%84.12	No
Srinivas., et al [9], 2010	ODANB	13	%83.70	No
El-Rashidy., et al [6], 2010	Fuzzy Clustering	13	%63.15	No
Abdullah and Rajalaxmi [10], 2012	Decision Tree	13	%63.33	No
Soni [11]., et al, 2011	Association Rules	13	%81.51	No
Khemphila و Boonjing [12], 2011	Back-Propagation MLP	8	%89.56	Yes
Our work	PSO+FFBP	8	%91.94	Yes

Table 2: Comparative result analysis using different techniques [1].
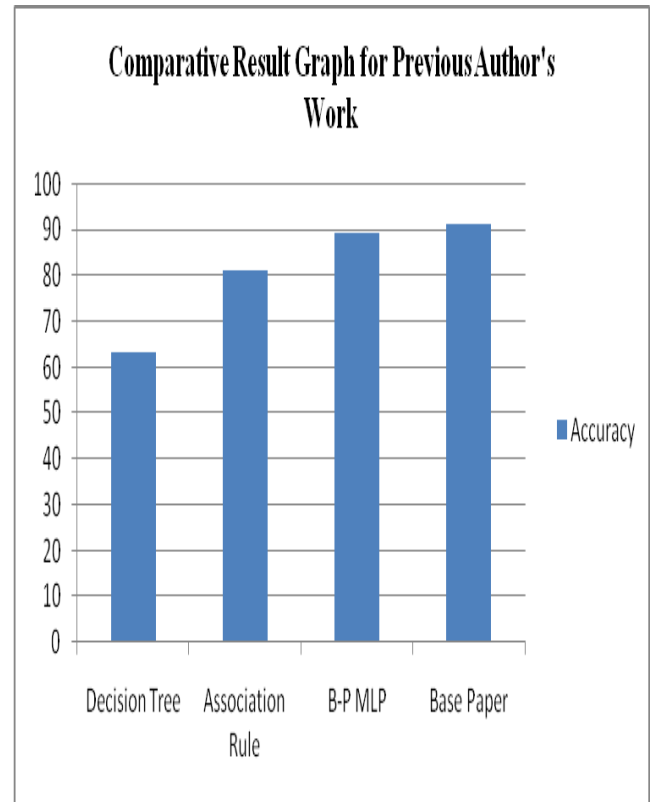

Figure 3: Compartive result graph for previous author work in the base paper.

VI RESEARCH SCOPE

Data mining and machine learning technique gives well know classifier for medical disease data classification. In cluster oriented ensemble classifier is suffered from a selection of optimal number of cluster for ensemble. The selection of optimal number of cluster improves the performance of cluster oriented ensemble classifier for medical disease data classification. Machine learning play big role in pattern recognition and network security. The recognition of pattern faced the series of training process. The training process of classification technique generates the accuracy performance of classifier and method of pattern recognition. The optimality of cluster is selected by heuristic function. For this process we may also use some optimization technique. By using this optimization technique we maintain the selection process of clustering technique and noise removal of boundary base class. Noise reduction and selection of optimal number of cluster in ensemble classifier used features sub set selection process using optimization technique. We also introduce a new feature sub set selection method for finding similarity matrix for clustering without alteration of classifier.

VII CONCLUSION AND FUTURE WORK

For the purpose of medical disease data classification various algorithm are applied such as machine learning, evolutionary algorithm, neural network and optimization techniques etc. The creation and selection of optimal number of cluster improves the performance of an algorithm for the medical disease data classification. Some optimization method discussed here and we plan to implement in future for the health care diagnosis such as some Evolutionary methods, swarm intelligence methods with classification techniques and compare some traditional classification methods for the medical science domain.

REFERENCES:-

[1] Majid Ghonji Feshki and Omid Sojoodi Shijani "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network", IEEE, 2016, Pp 48-53.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Delia-Maria FILIMON and Adriana ALBU "Skin Diseases Diagnosis using Artificial Neural Networks", IEEE, 2013, Pp 1-6.

[7] Sellappan Palaniappan and Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, 2008, Pp 108-115.

[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] Asha Gowda Karegowda and Darshan M "Optimizing Feed Forward Neural Network Connection Weights Using Artificial Bee Colony Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, 2013, Pp 452-454.

[13] Sultan Noman Qasem, Siti Mariyam Shamsuddin, Siti Zaiton Mohd Hashim Maslina Darus and Eiman Al-Shammari "Memetic multiobjective particle swarm optimization-basedradial basis function network for classificationproblems", Elsevier, 2013, Pp 165-190.

[14] Ashraf Osman Ibrahim, Siti Mariyam Shamsuddin, Nor Bahiah Ahmad and Sultan Noman Qasem "Three-Term Backpropagation Network Based On Elitist Multi-objective Genetic Algorithm for Medical Diseases Diagnosis Classification", Life Science Journal, 2013, 1815-1822.

[15]Mr. Deepak Dhanwani and Prof. Avinash Wadhe "Study Of Hybrid Genetic Algorithm Using Artificial Neural Network In Data Mining For The Diagnosis Of Stroke Disease", International Journal of Computational Engineering Research, 2013, Pp 95-100.

[16] Fabio Mendoza Palechor, Alexis De la Hoz Manotas, Paola Ariza Colpas , Jorge Sepulveda Ojeda, Roberto Morales Ortega, Marlon Piñeres Melo " Cardiovascular Disease Analysis Using Supervised and Unsupervised Data Mining Techniques" Volume 12, Number 2, February 2017. Pp 81-90.

[17] Syed Umar Amin, Kavita Agarwal and Dr. Rizwan Beg "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", IEEE, 2013, Pp 1227-1231.

[18] B V Sumana and T. Santhanam "Prediction of diseases by Cascading Clustering and Classification", ICAECC, 2014, Pp 1-8.

[19] V Krishnaiah, M Srinivas, Dr.G Narsimha and Dr.N Subhash Chandra "Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique", IEEE, 2010, Pp 1-7.