COMPARATIVE STUDY OF EXISTING TECHNIQUES FOR DIAGNOSING VARIOUS THYROID AILMENTS

Umar Sidiq[1],Rafi Ahmad Khan[2]

[1]Ph.d scholar Mewar University Rajastan India.
[2]Assistant professor Kashmir University,J&K India

Abstract:- *The thyroid gland is one of the most important organ in our body. It secretes thyroid hormones which are responsible for controlling metabolism. The less secretion and much secretion of thyroid hormone causes hypothyroidism and hyperthyroidism respectively. In this paper an overview and comparison of existing data mining techniques used for diagnosing thyroid diseases is presented. The current study explores preliminaries behind the techniques and presents classification of various techniques based on their accuracy and number of attributes under investigation. The main focus of this study will be to carry out the survey of existing data mining techniques used to diagnosis of various thyroid ailments, to present the techniques used and its accuracy.*

***Keywords:-*Thyroid diseases, neural network, Support Vector Machine, Decision tree.**

1. Introduction:-

Data mining based applications are very valuable and essential in healthcare and medical science. In health care, there are large amount of data, and this data has no organizational value until converted into information and knowledge, which can help control costs, increase profits, and maintain high quality of patient care. In the health sectors data mining play an important role to predict diseases [1]. Data mining has provided different techniques and tools to extract valuable hidden patterns of data from complex medical databases with no trouble and are/is used to diagnose disease/diseases of a patient more accurately. In this paper main focus is to present the survey of existing data mining techniques and accuracy achieved used to diagnosis of various thyroid ailments . Performance of technique/techniques varies as the number of attributes is increased or decreased to be used as input. So this study presents number of attributes used by different researchers in their work along with accuracy.

Thyroid diseases are one of the most common endocrine disorders found in worldwide. In India, it is expected that about 42 million people in India suffer from thyroid diseases [2].The thyroid or thyroid gland is one of important and commonly called an endocrine gland present in the human body and is located in the human neck below the Adam's apple. The main purpose of thyroid is to produce thyroid hormones thyroxin (T4) and triiodothyronine (T3) into the blood stream as the principal hormones to control the body's metabolic rate and growth. The failure of thyroid hormone will leads to thyroid disorders or diseases [3].When this hormone is secreted very little it may lead to hypo-thyroidism. And when this hormone is secreted of too much it may lead to hyper-thyroidism [4]. Both excess and less thyroid hormone secretion causes health problems and sometimes may lead to death. Thyroid diseases are broadly dived into two types (i) Hyperthyroid: Increase in the hormone production can cause hyperthyroidism. In medical field, "hyper" indicates excess or too much. Hyperthyroidism occurs when the gland produces excess hormones. The most common cause for hyperthyroidism is the autoimmune disorder Graves' disease so also known as an overactive thyroid and it can cause a extensive range ofphysical changes. The symptoms that indicate the presence of hyperthyroidism includes loss of weight, high blood pressure, nervousness, increase in heart rate, an increased sweating, swelling in your neck, frequent bowel movements, shorter menstrual periods and trembling hands [3]. (ii)Hypothyroid: Decrease in the hormone production can cause hypothyroidism. In medical field,The term hypo means less or deficient/not enough. Hypothyroidism is a condition that the thyroid gland does not produce enough hormones. Inflammation and damage to the gland causes hypothyroidism.

The following sections of this paper are designed as follows. Section-2 reviews literature pertaining to data mining and applications of data mining techniques used for diagnosis of thyroid diseases. Section-2.1 describe important attributes used for diagnosis of thyroid diseases. Section-2.2 throws light on the comparative study. In Section-3 presents the conclusion of the paper. And at last the references are mentioned.

2. Review literature:-

In recent years, various works have been done for the diagnosis of various thyroid diseases using different data mining techniques by different authors. They tried to attain efficient methods and accuracy in finding out diseases related to thyroid by their work including datasets and different algorithms along with the experimental results and future work that can be done on the system to achieve more efficient results. This paper aims at analyzing different data mining techniques, tools and number of attributes that has been introduced in recent years for diagnosis of thyroid diseases with accuracy by different authors and achieved different probabilities for different methods.

Gurmeet kaur et al. [5] This paper presents a comparison of three artificial neural network algorithms viz. Multilayer Back Propagation (BPN), Radial Basis Function and Learning Vector Quantization to diagnosing thyroid disease using image dataset. Images are first processed to filter out noise to increase the learning rate of neural network through image processing steps. It has been found in this study that the radial basis function network outperformed over other algorithms as having the least no. of iterations and highest PSNR value.(2014)

Farhad Soleimanian Gharehchopogh et al.[6] considered a Multi-layer Perceptron (MLP) ANN using back propagation learning algorithm to classify Thyroid disease. By selecting an input layer with 5 neurons, a hidden layer with 6 neurons and an output layer with just 1 neuron, The suitable selection of activation function and the number of neurons in the hidden layer and also the number of layers are achieved using test and error method. We can be reached the accuracy level for thyroid disease to 98.6% and also increase the performance of ANN .In this work the data are taken from data set of UCI machine learning which includes 215 samples and each one has 5 effective parameters to classify thyroid disease.(2013)

In [7]Various data mining techniques like Bayes net, MLP, RBF Network, C4.5, CART, REP tree and decision stump are used to develop classifier for diagnosis of hypothyroid disease and yielded 99.60%accuracy. In this paper a data set with3772 instances from which 3481 belongs to category negative,194 belongs to compensated hypothyroid category , 95 belongs to primary hypothyroid category while 2 belongs to secondary hypothyroid category. The 29 attributes were used for classify the data. The entire work is simulated in WEKA tool. (2013)

In [8] number of experiments and comparisons to demonstrate the importance of feature selection for thyroid disease diagnosis are performed and also a comparison of two support vector machine learning algorithm(SVM with 3 feature subset as input (3 × FC) and SVM with 3 feature subset as input (10 × FC)) for diagnosis of thyroid disease is presented. The SVM algorithms were applied over thyroid disease dataset taken from UCI Machine Learning Repository and obtained 98.5 % and 98.62 % respectively. (2012)

In [9] the decision tree attribute splitting rules is using for the diagnosis of thyroid disorders. This method provides different splitting criteria(Information Gain, Gain Ratio, Gini Index, Likelihood Ratio Chi-Squared Statistics, Distance Measure for the construction) of decision tree . These Various splitting rule for decision tree attribute selection had been analyzed and compared. This helps to diagnose the thyroid diseases through the extracted rules. From this work, it is clear, that normalized based splitting rules have high accuracy and sensitivity or true positive rate. This work can be extended for any medical datasets. Further enhancement can be made by using various optimization algorithms or rule extraction algorithms. (2012)

In [10] the five different neural network algorithm have been trained and then a comparison of two machines learning techniques (RBF and PNN) is presented for the diagnosis of thyroid disease. The thyroid disease dataset of 222 cases selected and examined by specialists. The networks determine, based on pathological examinations and serological test results, the person is having thyroid disease or not and if so which type and grade of disease he or she is affected. We found out that the RBF network and PNN outperforms other network including GRNN, LVQ, and SVM. The classification accuracy was obtained about 91.1% and 99.5% respectively. (2009)

In [11] the two algorithms of of decision tree techniques (C4.5 & C5.0) were applied over thyroid disease dataset acquired from UCI Machine Learning Repository and also their comparison for diagnosis of thyroid disease is presented. The accuracy of C4.5 and C5.0 were more than 90% and 95% .(2010)

ANUPAM SKUKLA et al. [12] proposed the diagnosis of thyroid disorders using Artificial Neural Networks (ANNs). Three ANN algorithms; the Back propagation algorithm (BPA), the Radial Basis Function (RBF), and the Learning Vector Quantization (LVQ) Networks have been implemented in metlab for the diagnosis. The accuracy was obtained 92%, 80% and 98% respectively. The dataset of thyroid was obtained from UCI repository of machine databases [22].From this data-set, 187instanceshave been used for this work, each instance has five attributes and one class attribute. (2009)

In [13] SVM and PNN were used to classify hypothyroidism and hyperthyroidism. Feature selection is done by using GA. The comparative study was done on two datasets from DCI. Both PNN and SVM gave same efficient accuracy result of 100% on dataset one, but for second dataset classification accuracy of SVM was much better 99.02 % than PNN 96.8%.

In [14] a comparison of two fuzzy base techniques (i.e. AIRS and AIRS with fuzzy weighted algorithms) for the diagnosis of thyroid disease is implemented. The thyroid disease dataset which is taken from UCI machine learning respiratory. The robustness of this method with regard to sampling variations is examined using a cross-validation method. The achieved accuracy was 81% and 85% respectively. (2007)

Keles et al. [15] diagnosed thyroid diseases with a expert system that called ESTDD (expert system for thyroid disease diagnosis), We found fuzzy rules by using neuro fuzzy method, which will be emplaced in ESTDD system. The accuracy was achieved 95.33%. Beside it can be

benefited from this system for training of students in medicine. (2008)

F.Temurtas [16] realized the diagnosis by multilayer, probabilistic, and learning vector quantization neural networks were implemented on thyroid disease dataset which was taken from UCI machine learning database was used and the achieved accuracy 92.96 %,94.43% and 89.79% respectively. The dataset which consists of the thyroid disease measurements contains three classes with 215 samples (normal(150),hyper(35) and hypo(30)) and five attributes were taken.(2009)

In [17] A CAD system PCA-ELM is developed for assisting the diagnosis of thyroid disease. Experimental results proved that the proposed system performed significantly well in distinguishing among hyperthyroidism, hypothyroidism and normal ones. It was observed that PCA-ELM achieved the highest classification accuracy of 98.1% and mean classification accuracy of 97.73% using10-fold cross-validation. And also a comparative study was conducted between PCA-SVM and PCA-ELM. The experimental outcome showed that PCA-ELM outperformed over PCA-SVM in terms of classification accuracy with shorter run time. The five attributes were used for this work. (2012)

In [18] systematic approach for earlier diagnosis of Thyroid disease using back propagation algorithm used in neural network. ANN has been developed based on back propagation of error used for earlier prediction of disease. ANN was subsequently trained with experimental data and testing is carried out using data that was not used during training process. Results show that outcome of ANN is in good agreement with experimental data; this indicates that developed neural network can be used as an alternative for earlier prediction of a disease. The MATLAB Neural Network Toolbox software was implemented the classify the thyroid disease. In this work 29 attributes were used such as age , sex, thyroxin etc.(2015)

In [19] The authors analyzed and compared four classification models: Naive Bayes, Decision Tree, Multilayer Perceptron and Radial Basis Function Network. The results indicate a significant accuracy for all the classification models. The Decision Tree model outperformed over other classification models. In this work 29 attributes of dataset have been chosen and applied Feature Selection technique i.e. Chi-Square, The dataset is then filtered by applying the unsupervised discredited filter on the attributes to convert the continuous values into nominal and thus reduce the 29 attributes to 10 attributes.(2017)

In [20] we have developed new classification model like ensemble of C4.5 and random forest for classification of thyroid data. Feature selection played a important role to improve the performance of our developed model. Our proposed model gives high classification accuracy of 99.47 % with less number of features compare to other existing developed model. The dataset has been downloaded from the University of California at Irvine (UCI) machine learning repository [21] to demonstrate the technique. The dataset contain 30 attributes in thyroid which 29 features are considered as input and the 30th feature is considered as output. In this data set contains 7547 records in which 776 belong to thyroid and 6771 belong to non- thyroid data. The Thyroid data consisting of both hypothyroid and hyperthyroid data. (2015).

In [23] In this work, The data set used for this study on hypothyroid is taken from University of California Irvine (UCI) data repository. The entire research work is implemented in WEKA. Two data mining techniques such as J48 and Decision stump tree are used to classify hypothyroid disease and achieved high accuracy of 99.57% and 95.38% respectively. In this work 12 attributes are used. (2016)

Table no 1

Authors/ Author	Year	Technique/Techniques used	Accuracy achieved	Attributes used
Farhad Soleimanian Gharehchopogh et al. [6]	2013	MLP with back-propagation	98.6%	05
Shivanee Pandey et al.[7]	2013	MLP	94.035 %	29
		RBF	95.228 %	
		Bayes net	98.59 %	
		C4.5	99.57 %	
		CART	99.54 %	
		Decision stump	95.38 %	
		REP tree	99.57 %	
ANUPAM SKUKLA et al. [12]	2009	BPA	92%	05
		RBF	80%	
		LVQ	98%	
Feyzullah Temurtas[16]	2009	MLNN with LM (3 · FC)	92.96 %	05
		PNN (3 · FC)	94.43 %	
		LVQ (3 · FC)	89.79 %	
Li-Na Li et al [17]	2012	PCA-SVM	96.32 %	05
		PCA-ELM	98.10 %	
Suman Pandey et al [20]	2015	Ensemble of C4.5 and Random Forest 99.47	99.47 %	29
G. Rasitha Banu [23]	2016	J48	99.57 %	12
		Decision stump tree	95.38 %	
M. Nazari et al [8]	2012	SVM with 3 feature subset as input (3×FC)	98.5 %	03
		SVM with 3 feature subset as input (10 × FC)	98.62%	

2.1. General Used Attributes to diagnose thyroid Diseases:

By analyzing the above research work it is found that frequently used medical attributes to perform experimental work for the diagnosis of thyroid diseases are given below in below table no.2.Among these attributes almost every researcher has selected attributes to perform work for thyroid disease diagnosis. In some papers this is given that they use 05 attribute[6][12][16][17]and but in case of other research work used 12 attributes[23]and so on for diagnosis of thyroid diseases.

Table no 2

Attributes	Description
Age	In years
Sex	Male or female
TSH	Thyroid-Stimulating Hormone
T3	Triiodothyronine
TBG	Thyroid binding globulin
T4U	Thyroxin utilization rate
TT4	Total Thyroxin
FTI	Free Thyroxin Index

2.2 Performance Study

The outcome of predictive data mining techniques on the same or different data sets reveals that neural networks[10][6] outperforms over other techniques and sometimes support vector machine and decision tree are having almost similar accuracy in as of neural network diagnosing various thyroid diseases[13][8][7]. The summary of data mining techniques is given in below table no 3.

Table no 3

Technique used	Accuracy	Reference
Neural Network	99.5 %	[10]
	98.6 %	[6]
	99.1%	[10]
Support vector machine	99.02 %	[13]
	98.62 %	[8]
Decision Tree	99.5 %	[7]

Various kinds of data mining techniques are used by different researchers to perform their practical work which has emerged in recent years for efficient and effective thyroid disease diagnosis. The analysis shows that Classification techniques are most promising for efficient and effective thyroid disease diagnosis. Classification technique is based on machine learning. Classification method utilize mathematical techniques such as decision trees, neural network, linear programming and statistics. These methods are mostly used in machine learning, artificial intelligence and pattern recognition. Various types of classification techniques are KNN, Decision Tree, Support Vector Machine, Neural Network and Bayesian Method. Some kinds of classification techniques mostly used to diagnose thyroid diseases are defined below;

1) Neural Network: Neural network is a parallel system based on human nervous system having multiple interrelated processing elements known as neurons, functioning in unity to solve specific problem**.** Back propagation is three layered based on the neural network algorithm. An input layer - the inputs are fed into this layer, a hidden layer – give the input to the output layer and one output layer, which emits the network's prediction. This network model helps to classifying new data.

2) Support Vector Machine: Support Vector Machine is one type of learning system algorithm, which is used to perform classification more accurately. A hyper plane or multiple planes are created by the support vector machine classifier in high dimensional space .This hyper plane separates the positive and negative of training data sample

3) Decision Tree: Tree-like graph is used in decision tree classifier. A Decision tree has three nodes, internal node denotes the test on an attribute, leaf node denotes the classes or class distributions and root node is the top most node in a tree.C4.5 and ID3 are two important algorithms used to construct the decision tree for diagnosis of thyroid diseases. Researchers use Decision Tree widely in healthcare field particularly to diagnose various thyroid diseases.

3. Conclusion:

The objective of our work is to provide a study of different data mining techniques that can be employed in diagnosis of thyroid diseases. Various data mining classification techniques are defined in this work which has been used in recent years for efficient and effective thyroid disease diagnosis. The analysis shows that different technologies are used in all the papers showing different accuracy to each other. In most research papers it is shown that neural network outperforms over other techniques. On the other hand, this is also given that support vector machine and decision tree has also performed well. There is no doubt that researchers worldwide have attained a lot of success to diagnose thyroid diseases but it is suggested to decrease the number of parameters used by the patients for diagnosis of thyroid diseases. More attributes means a patient has to undergo more number of clinical tests which is both cost effective as well time consuming. Thus there is need to develop such type of algorithms and thyroid disease predictive models which require minimum number of parameters of a person to diagnose thyroid disease and saves both money and time of the patient.

References

[1] S. Vijiyarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey", International Journal of Computer Applications & Information Technology, vol. II, no. I, pp. 17–21, 2013.

[2]http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169866/(accessed dec 2015)

[3]http://www.foxnews.com/health/2012/02/10/hypothyroidism–versus-hyperthyroidism.html(accessed dec 2015

[4] Ozyilmaz,Lale, and Tulay Yildirim."Diagnosis of thyroid disease using artificial neural network methods".Neural Information Processing,2002.ICONIP'02.Proceedings of the 9th International Conference on Vol.4.IEEE,2002.

[5] Gurmeet kaur et al., "ARTIFICIAL NEURAL NETWORKS FOR DIAGNOSIS OF THYROID DISEASE", International Journal for Technological Research in Engineering, Volume 2, Issue 1, September-2014

[6] F. S. Gharehchopogh, M. Molany and F. D.Mokri, "USING ARTIFICIAL NEURAL NETWORK IN DIAGNOSIS OF THYROID DISEASE: A CASE STUDY", International Journal on Computational Sciences &Applications (IJCSA) Vol.3, No.4, August 2013.

[7] Shivanee Pandey, Rohit Miri, S. R. Tandan, "Diagnosis and Classification of Hypothyroid Disease Using Data Mining Technique", TJERT, June 2013.

[8] M. Nazari, F.Seiti, and M. Teshnehlab. "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification" International Journal of Electrical & Computer Sciences IJECS-IJENS ,(2012) Vol: 12 No: 01 pp.13-19

[9] J.Margret ,B. Lakshmipathi and S.A.Kumar , "Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules", International Journal of Computer Applications (0975 – 8887) Volume 44– No.8, April 2012.

[10] Modjtaba Rouhani, Kamran Mansouri, "Comparison of several ANN architectures on the Thyroid diseases grades diagnosis",ISCAIT,2009.

[11] A. Upadhayay "Empirical Comparison by data mining Classification algorithms (C 4.5 & C 5.0) for thyroid cancer data set", International Journal of Computer Science & Communication Networks, ,(2010). Vol 3(1), pp.64-68

[12] Anupam Shukla, Prabhdeep Kaur, Ritu Tiwari and R.R. Janghel, Diagnosis of Thyroid disease using Artificial Neural Network. In Proceedings of IEEE IACC 2009, pages 1016-1020.

[13] Fatemeh Saiti, Afsaneh Alavi Naini, Mahdi Aliyarishoorehdeli, Mohammad Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and S VM".

[14] K. Polat, S. Sahan, & S. Gunes. " A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis" Expert Systems with Applications. (2007) vol: (32), p.p 1141–1147.

[15] Keles,et al ESTDD: expert system for thyroid diseasesdiagnosis. Expert Syst. Appl. 34(1):242–246, 2008.

[16] Feyzullah Temurtas" A comparative study on thyroid disease diagnosis using neural networks" Expert Systems with Applications 36 (2009) 944–949.

[17]Li-Na Li,Ji-Hong Ouyang ,Hui-Ling Chen &Da-You Liu"A Computer Aided Diagnosis System for Thyroid Disease Using Extreme Learning Machine"J Med Syst (2012) 3327–3337.

[18]Prerana, Parveen Sehgal, Khushboo Taneja"Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology Vol. 3, No. 2, April 2015.

[19] S. Sathya Priya, Dr. D. Anitha "Survey on Thyroid Diagnosis using Data Mining Techniques" International Journal of Advanced Research in Computer and Communication Engineering Vol. 6, Special Issue 1, January 2017.

[20] Suman Pandey, Deepak Kumar Gour, Vivek Sharma" Comparative Study on Classification of Thyroid Diseases" International Journal of Engineering Trends and Technology (IJETT) – Volume 28 Number 9 - October 2015.

[21] Sheetal Gaikwad and Nitin Pise "An Experimental Study on Hypothyroid Using Rotation Forest" International Journal of Data Mining &Knowledge Management Process (IJDKP) Vol.4, No.6, November2014.

[22] UCI repository of machine learning databases,http.//archive. ics. uci. edu/ml/datasets. html accessed November

[23]G. Rasitha Banu " A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease" International Journal of Computer Sciences and EngineeringVolume-4, Issue-11 2016.