Capitulation of Machine Learning Techniques for Detection of Auto immune Thyroiditis

S.Sathyapriya[1] ,Dr.D.Anitha[2]

[1]*Research Scholar, Department of Computer Science, Sri Ramakrishna*

Arts and Science College for Women, Coimbatore,(India)

[2]*Research Supervisor, Department of Computer Science, Sri Ramakrishna*

Arts and Science College for Women, Coimbatore,(India)

ABSTRACT

Even as hypothyroidism continues to grow significantly in the country, particularly among women, awareness its diagnosis remains shockingly low, a latest study shows one in ten adults in India suffer from Auto immune hypothyroidism, a recent survey conducted by Indian.The prevalence of hypothyroidism in the developed world is estimated to be about 4-5%. Thyroid disorders impair normal functioning of the thyroid gland causing abnormal production of hormones leading to hyperthyroidism. If left untreated, hypothyroidism can cause elevated cholesterol levels, an increase in blood pressure, cardiovascular complications, decreased fertility, and depression.According to the survey, the disease is more prevalent among women,. This review will focus on the epidemiology of five common thyroid diseases in India: (1) hypothyroidism, (2) hyperthyroidism, (3) goiter and iodine deficiency disorders, (4) Hashimoto's thyroiditis, and (5) thyroid cancer. This review will also briefly cover the exciting work that is in progress to ascertain the normal reference range of thyroid hormones in India, especially in pregnancy and children In pregnant women, thyroid disorders can lead to placental abnormalities and increased risks, Some parameter estimation methods, execution of the distinctive neural system models have been explored

Key word: Auto immune Hyperthyroidism,Hypothyroid

I.LITERATURE REVIEW

There are many people who have studied various medical data and analyzed methods and models for preprocessing and classifying

The data Ngan, Po Shun, et al (1999) introduced a system for discovering medical knowledge by learning Bayesian networks and rules. Evolutionary computationis used as the search algorithm. The Bayesian networks can provide an overall structure of the relationships among the attributes

Yeh, Wei-Chang (2012) improved simplified swarm optimization (SSO) to mine a thyroid gland dataset collected from UCI databases. Close Interval Encoding (CIE) is added to efficiently represent the rule structure, and the Orthogonal Array Test (OAT) is added to powerfully prune rules to avoid over-fitting the training dataset [16].

Chen, Hui-Ling, et al (2012) proposed expert system, Fisher Score Particle Swarm Optimization Support Vector Machines (FS-PSO-SVM) has been rigorously evaluated against the thyroid disease dataset, which is commonly used among researchers who use machine learning methods for thyroid disease diagnosis[20].

Azar et. al (2013) performed a comparison between hard and fuzzy clustering algorithms for thyroid diseases data set in order to find the optimal number of clusters. Different scalar validity measures are used in comparing the performances. K-means clustering; Kmedoids clustering; Fuzzy C-means; Gustafson–Kessel algorithm; Gath–Geva algorithm clustering results for all algorithms are then visualized by the Sammon mapping method to find a low-dimensional (normally 2D or 3D) representation of a set of points distributed in a high dimensional pattern space .

II.AUTO IMMUNO HYPERTHYROIDISM

Population studies have suggested that about 16.7% of adult subjects have anti-thyroid peroxidase (TPO) antibodies and about 12.1% have anti-thyroglobulin (TG) antibodies. In this same study of 971 subjects, when subjects with abnormal thyroid function were excluded, the prevalence of anti-TPO and anti-TG antibodies was 9.5% and 8.5%.

Arificial Neural Networks

Data mining in health care

Data mining refers to extracting unknown patterns from an enormous volume of data involving different methods and algorithms which exist at the intersection of fields such as artificial intelligence, machine learning, statistics and database systems (Piatetsky-Shapiro & Parker, 2011). Hospitals, clinics and medical analysis laboratories accumulate a large amount of patient data over the years. These data provide a basis for the analysis of risk factors for many diseases (various types of cancer, heart diseases, diabetes, hepatitis etc.). In literature are mentioned certain applications of data mining techniques in the health domain, some of them being presented in the following paragraphs. The authors have narrowed their research area on thyroid disorders and the examples given below are strictly about the related work described in literature, regarding the application of data mining for these classes of diseases. The majority of examples refer to diagnosing diseases of thyroid using decision trees, artificial neural networks, support vector machine, expert systems etc. For example, the diagnosis of thyroid disorders using ANN's

An ANN is a machine learning method which recreates the human cerebrum and is framed by number of handling units called neurons and builds up an association between them. The associations have weights connected with the neurons, representing the knowledge gained by the network3. A layer of a neural network is the one which contains neurons in the system. There are 3 unique layers of network. The primary layer is the input layer which acts as an interface with the user of the network and the system and the last layer is an output layer which generates output for a given input. There might be one or more hidden layers amongst the input and output layers. Neural Systems with multiple hidden layers are called multilayer perceptrons. The hidden layer associates to distinguish the perplexing examples in the input information. Neural Networks have been generally applied in a wide variety of problems. A portion of the noticeable and broadly utilized neural systems which go under the classification of feed forward neural systems, are outspread premise capacity systems, Bayesian

systems, fuzzy neural systems, back propagated neural systems, stream based systems. Learning should be possible in two ways. 1. Supervised Learning

2. Unsupervised Learning

Supervised learning is a technique where the output of the network is matched the teacher information. Many supervised learning algorithms are available among which back propagation algorithm is used for different types of applications.

Unsupervised learning is used in self organizing neural nets. As opposed to direct learning, unsupervised learning does not require an instructor. In this technique for learning, the input vectors are compared and assembled together to shape a group with comparable qualities. Proposed4 a model to diagnose the thyroid dysfunction using the Artificial Neural Networks (techniques) models

III.PRE-PROCESSING

Pre-processing is the import step need to be followed to achieve good classification of data. Preprocessing is nothing but cleaning of data to remove unwanted or bad data. A data set is said to be a bad data when it has incomplete values or missing data also the information which doesn't helps in classification are unwanted data. The major two techniques for getting good data from the thyroid data set which are used in this paper are Correlation based feature subset selection and wrappers technique.

IV.ARTIFICIAL NEURAL NETWORKS

Models like

(1)The cross validation classification method

(2)Variable Selection method

(3)The Regression – based method

The neural networks have shown best and accurate results in diagnosing the thyroid dysfunction.

Limitation: For large number of characteristics there is a need to build up a variable determination technique.

Proposed5 an important classification problem for thyroid function diagnosis by using the multivariate analysis and two notable approaches the Bayesian regularized networks and also the Self Organizing Map (SOM).

Thyroid sickness finding6 was done with different Neural Networks with various enactment capacities for comparing the performance of the networks. Every N/W was compared to decide the performance of the networks that gave the best results.

machine learning strategy including simulated insusceptible acknowledgment framework (AIRS) which is an arrangement framework to analyze the thyroid issue.

Thyroid sickness finding information set gave an 85% precision in past applications to enhance 20% exactness to utilize hybridized frameworks. A method with expert system for thyroid disease diagnosis. The method used for thyroid disease diagnosis called as STDD (System Thyroid Disease Diagnosis).

ANN structures on the thyroid ailment finding. To analyze the thyroid sickness, RBF, LVQ and SVM system models are utilized for the purpose. It is observed that, RBF Networks gave the best results for determination of thyroid organ capacity. RBF Networks outperformed other systems in analyzing the thyroid disease.
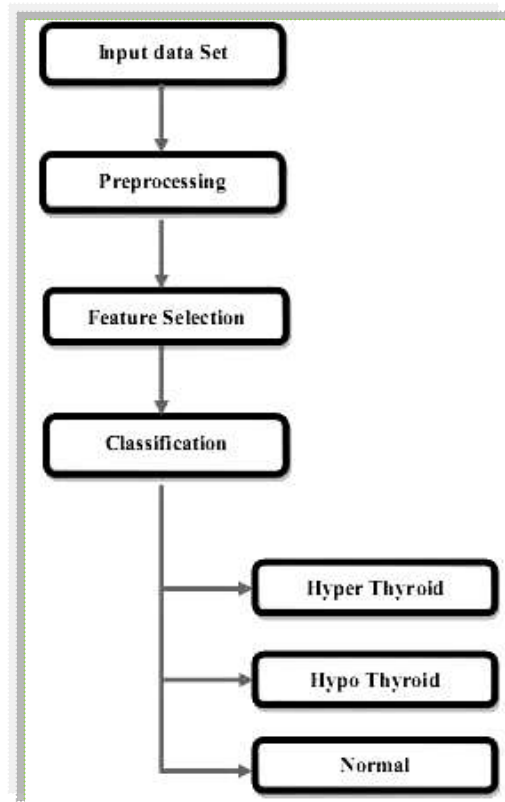
V.CONCLUSIONS

Auto Immune Hypothyroid is one of the most common diseases. It is affects almost every aspect of health. The thyroid produces several hormones, each of them must be produced by Thyroidin normal rang; to help cells convert oxygen and calories into energy. Since Thyroid datasets are uncertain data, missing attribute values, and continuous features, presented a reliable learning method and analytical study for diagnosing hypothyroid disease that can be used by doctors in other medical diagnosing algorithms. Our statistical results show that STDD algorithm is the best in reducing size of tree, time, attributes and increasing accuracy. Although data sets with modified similarity relation achieved good results The neural networks techniques have shown best and accurate results in diagnosing the thyroid dysfunction.

REFERENCES

[1] L. Ozyilmaz and T. Yildirim (2002). "Diagnosis of thyroid disease using artificial neural network methods," in: Proceedings of ICONIP'02 9th international conference on neural information processing (Singapore: Orchid Country Club) pp. 2033–2036.

[2] K. Polat, S. Sahan and S. Gunes (2007). "A novel hybrid method based on artificial immunerecognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis," Expert systems with Applications, vol. 32, pp. 1141-1147.

[3] G. Zhang, L.V. Berardi(2007). "An investigation of neural networks in thyroid functiondiagnosis," Health Care Management Science, 1998, pp. 29-37. Available: http://www.endocrineweb.com/thyroid.html, (Accessed: 7)

[4] Baris Senliol, et al (2008). "Fast Correlation Based Filter (FCBF) with a different searchstrategy", Computer and Information Sciences, ISCIS'08. 23rd International Symposium on. IEEE.

[5] Blake.C and Merz.L (1998). UCI repository of machine learning databases, Department of Information and Computer Science, University of California at Irvine, Irvine CA.

[6] Brown, G., Pocock, A., Zhao, M.-J., Lujan, M (2012).Conditional Likelihood Maximisation: AUnifying Framework for Information Theoretic Feature Selection. The Journal of Machine Learning Research (JMLR).

[7] Dash .M, Liu .H (1997). "Feature Selection for Classification", Intelligent Data Analysis, pp 131–156.

[8] Deng H and Runger G. (2012). "Feature Selection via Regularized Trees", in International Joint Conference on Neural Networks (IJCNN), IEEE.

[9] Dietterich, T (2000). "Ensemble methods in machine learning", Proceedings of the 1st International Workshop on Multiple Classifier Systems.

[10] Dr. Bhargava Neeraj, Sharma Girja, Dr. Bhargava Ritu and Mathuria Manisha (2013). "Decision Tree Analysis on J48 Algorithm for Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering (JARCSSE), Volume 3, Issue 6.

[11] Duch, W (2000). "Similarity based methods: a general framework for classification approximation and association", Control and Cybernetics.

[12] Guyon, I., Elisseeff, A (2003). "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3, Volume 3,pp 1157-1182

[13] Hall Mark A (2009). "Correlation based Feature Selection for Machine Learning", Thesis for the degree of Doctor of Philosophy at The University of Waikato.

[14] Hsu Hui-Huang, Hsieh Cheng-Wei, Lu Ming-Da (2011). "Hybrid feature selection by combining filters and wrappers", Elsevier Expert Systems with Applications Journal, ISSN 09574174.

[15]https://www.omicsonline.org/thyroid-disorders-therapy.php