

# Prediction of thyroid Disease Using Data Mining Techniques

Bibi Amina Begum

PG Scholar

Department of Computer Science and Engineering  
MSRIT

Bangalore ,India

bibiainalr@gmail.com

Dr.Parkavi A

Assistant Professor

Department of Computer Science and Engineering  
MSRIT

Bangalore ,India

parkavi.a@msrit.edu

**Abstract**—Classification based Data mining plays important role in various healthcare services. In healthcare field ,the important and challenging task is to diagnose health conditions and proper treatment of disease at the early stage. There are various diseases that can be diagnosed early and can be treated at the early stage. As for example, Thyroid diseases. the traditional ways of diagnosing thyroid diseases depends on clinical examination and many blood tests.The Main task is to detect disease diagnosis at the early stages with higher accuracy.Data mining techniques plays an important role in healthcare field for making decision,disease diagnosis and providing better treatment for the patients at low cost. Thyroid disease Classification is an important task.The purpose of this study is predication of thyroid disease using different classification techniques and also to find the TSH, T3,T4 correlation towards hyperthyroidism and hypothyroidism and also to finding the TSH, T3,T4 correlation with gender towards hyperthyroidism and hypothyroidism.

**Keywords**—Data mining techniques,classification model, thyroid diseases,Decision tree,knn,svm,Naïve Bayes,ID3 classification

## I. INTRODUCTION

Diagnosis of Thyroid Disease is very tedious and difficult tasks. The diagnosis thyroid disease in the traditional way includes clinical examination and the many blood tests. But then the main task is to diagnosis the disease at early stages with high accurate percentage. In medical field ,Data mining plays[1] a crucial role for diagnosis of disease. Data Mining provides many classification techniques for the prediction of disease accuracy. The gathered patient data collected from many health care organization is useful for the risk factors analysis for many diseases.

In healthcare and medical science ,the applications based on Data mining are very beneficial and important. The large amount of data gathered from health care organization has no organizational value unless transformed into most useful information and knowledge[13],which could be helpful in cost controlling, increasing the profits, and high quality maintainance of patient healthcare.Classification Algorithms is one of the most important applications in the field of data mining, which can be usefull for decision making in many real world problems.

## II. OVERVIEW OF THYROID

The thyroid is an endocrine gland.The function of thyroid gland is to produce of thyroid hormones[11].It reaches to all other organs through the bloodstream and control metabolism and growth development [5].The important functions of thyroid gland includes respiration,blood circulation,gut movements,temperature control,muscle functioning ,digestion and functioning of brain.Any dys function in the thyroid gland [12]may affects the normal physiological functioning human body [3].The thyroid hormone affects the growth and development depending on the amount of secretion.When the production of thyroid hormone is very less and is referred as hypo-thyroidism.When the production of thyroid hormone is very high then this type of thyroid disease is referred as hyper-thyroidism [2].

### A. Thyroid Harmones

The thyroid gland produces are tri-iodothyronine (T3)and L-thyroxine (T4) and [6].The thyroid hormones regulates various metabolic activities such as generation of heat, the consumption of carbohydrates, protein and fats. The pituitary gland controls production of tri-iodothyronine and L-thyroxine hormones.The Thyrotropin-Stimulating Hormone [10] from pituitary gland is released when thyroid hormone is required and circulates through the bloodstream to reach thyroid gland. TSH then stimulates the thyroid glands for the production of T4 and T3 hormones [6].The production of thyroid hormone are controlled by the feedback system [6] of pituitary gland.The TSH production is less when T3,T4 are more in the circulation and TSH production is more when T3,T4 are less.

### B. Thyroid and its Health Effects

Thyroid disorder are most common endocrine disease,across the worldwide. In an Indian survey 42 million people are suffering from this diseases [4].The Thyroid diseases is different from others in terms of their the relative visibility, predication ease, medical treatment mode accessibility [4] [5].The inappropriate production of thyroid hormone affects health conditions.

### 1) Hyperthyroidism

Increase production in the thyroid hormones causes hyperthyroidism. Graves' disease is one of the autoimmune disorder that causes hyperthyroidism[9]. The symptoms[14] are dry skin, increase sensitivity to temperature, thinning of hair, weight loss, increase heart rate, high blood pressure, excess sweating, neck enlargement, nervousness, menstrual periods shorten, frequent gut movements and hands trembling [3].

### 2) Hypothyroidism

Decrease production in the thyroid hormones causes Hypothyroidism. The medical term hypo means deficient or less. The causes for hypothyroidism are inflammation and thyroid gland damage. The Symptoms includes obesity, low heart rate, increase in cold sensitiveness, neck swelling, dry skin, hands numbness, hair problem, heavy menstrual periods and digestive problems. And these Symptoms may worsen over period if not treated [3].

## III. LITERATURE SURVEY

Deepika Koundal et al.[6] have studied the existing the earlier automatic tools for diagnosis of disease at the easier stage in an efficient way. Also the metrics study about the different evaluation of performance and also investigations on the trends and future developments are studied.

Nikita Sigh and Alka Jindal [7] have compared Support Vector Machine with K-Nearest Neighbor and Bayesian and concluded Support Vector Machine better than KNN and Bayesian with an accuracy about 84.62%. KNN found the nearest neighborhood automatically. The results is represented by graph with object as each vertices. The probability classification is done using Bayesian which indicates the sample data belongs to a class.

Edgar Gabriel et al.[8] have proposed a texture-based segmentation i.e two parallel versions of a code for Fine Needle Aspiration Cytology thyroid images is the most important first step in identifying a fully automated Computer Aided Design solution. The code is developed in MPI version to exploit computer resources such as PC clusters.

Preeti Aggawal et al.[9] listed the method for an automatic segmentation. The study shows the summary obtained by applying specific algorithm (automatic) segmentation and automatic tools on both thyroid US as well on lung CT [7]. For segmentation of thyroid US images they have used Analyze 10.0 and Mazda.

Eystraits G[10] have provided system TND (Thyroid Nodule Detector) using a technique called computer-aided diagnosis (CAD). During thyroid UltraSound examinations, a nodular tissue detection is used in ultrasound (US) and thyroid images videos acquired.

Won-Jin Moon et al.[8] have evaluated to differentiate between benign and malignant thyroid nodules using the accuracy of diagnostic ultrasonography (USG). They concluded that the important criteria and presence of calcification is shape, margin, echogenicity from benign nodules is discrimination of malignant.

S. Yasodha et al.[11] have proposed hybridization of Class-Attribute Contingency Coefficient (CACC)- Support Vector Machine techniques. The combination of CACC and SVM classification techniques are applied on thyroid data. When compared to other traditional models, the accuracy of the proposed model is better.

Alfonso Bastias et al.[4] have aimed at developing a machine learning classifier using AIS for diagnosis of health condition and of the proposed classifier for capability investigation. The proposed classifier successfully improved the thyroid gland disease identification process.

Gurmeet et al. [3] has proposed NN training diagnosis model for the of the thyroid disease. It aims in developing the general model for identifying any kind of disease. The objective of this paper is to thyroid disease diagnose by using three different artificial neural network algorithm having different framework, characteristics and accuracy.

Ali keles et al. [7] proposed an expert system for predicting of thyroid that is known as Expert System for Thyroid Disease Diagnosis (ESTDD). This expert system diagnose thyroid diseases through neuro fuzzy rules with 95.33% of accuracy.

## IV. DATASET DESCRIPTION

Dataset is taken from UCI machine learning repository [15]. Database consists of patients thyroid records. Each thyroid patients record consists of 15 attributes listed below.

Attribute can be Boolean (true/false) or continuous valued are in given below table I.

SN	Attribute Name	Value Type
1	Age	continuous
2	Sex	m,f
3	On_thyroxine	f,t
4	Query_on_thyroxine	f,t
5	Thyroid_surgery	f,t
6	Query_hypothyroid	f,t
7	Query_hyperthyroid	f,t
8	Pregnant	f,t
9	Goitre	f,t
10	TSH value	continuous
11	T3 value	continuous
12	TT4 value	continuous
13	T4U value	continuous
14	FTT value	continuous
15	TBG value	continuous

**Table I. Data Description**

## V. CLASSIFICATION TECHNIQUES IN DATA MINING

### A. Decision Tree

A Decision tree[6][8] has 3 types of node such as internal node that represents test attribute, the classes or class attribute are denoted by the leaf node, the top most

is denoted by the root node of the tree. To construct the decision tree C4.5 and ID3 algorithms are used. The Advantages of using Decision tree is to identify and eliminate the redundant data known as “tree pruning” to improve the accuracy of the classification. The decisions are made on attribute with the highest normalized data also it can applied to both continuous and discrete values. On the other hand the disadvantages includes, for large data bases the efficiency and scalability are low.

### B. Backpropagation Neural Network

Backpropagation is a neural network algorithm. It consists of three different layers, input layer - the inputs are given here, hidden layer - the input to hidden layer can be the outputs with weights [5], number of hidden layer's arbitrary, output layer - the input to the output layer is from hidden layers, which eliminates prediction of the network's. Thus the advantages includes high accuracy, Very flexible for noisy and when the data is inconsistent, easy update of weights. The Disadvantages of Backpropagation Neural Network are representation of knowledge, it is difficult for humans to interpret, Knowledge. Decreases the accuracy of the network by the removal of weighted links. Selection of training dataset is difficult.

### C. Support Vector Machine

One of the type of learning system algorithm is Support Vector Machine[8], which is used to perform classification in a better accurately and uses 2 class classifier, referred as hyper plane as “decision boundary or decision surface”. The hyper plane separates positive training sample with the negative training data sample in a plane. The advantages includes an easy extend, used for pattern reorganization, quadratic optimization problem can be formulated. The some other disadvantages are suitable only for real valued space. It allow only 2 classes for classification using binary method and several strategies for multiple class classification. For user its very hard to understand Hyper plane.

### D. Density-based clustering

The density based clustering algorithm falls under data clustering algorithm: A space is considered with given set of points it groups together points that are closely packed together i.e., points with many closely neighbors. The most common clustering algorithms and also most cited for scientific literature is density based algorithm. It is opposite to k-means, using an R\*tree. In Density based clustering algorithm, an unassigned object is chosen from the given data set classification method like Hierarchical multiple classifier is used to classify the given dataset. Thus it is an efficient way to classify an data with accurate information in reduced time and cost[13].

## VI. PROPOSED WORK

The thyroid Dataset is taken from UCI data repository site. The Database consists of thyroid patient records. The Patients record is having different attributes described in

the data set description and different data mining techniques are applied to get the predication of thyroid disease and then Linear regression is performed to obtain the which hormone among TSH,T3,TT4 affect the male and female.And also which among the TSH, T3,TT4 influence the hypothyroidism and hyperthyroidism.

In healthcare services data mining technique is mainly used for making decision, disease diagnosing and giving better treatment to the patients at comparately low cost.Classification of thyroid disease plays is an important task in the prediction of disease.Dimensionality reduction may be done as a future work so that number of blood test the thyroid will be reduced and also time required to diagnose disease.

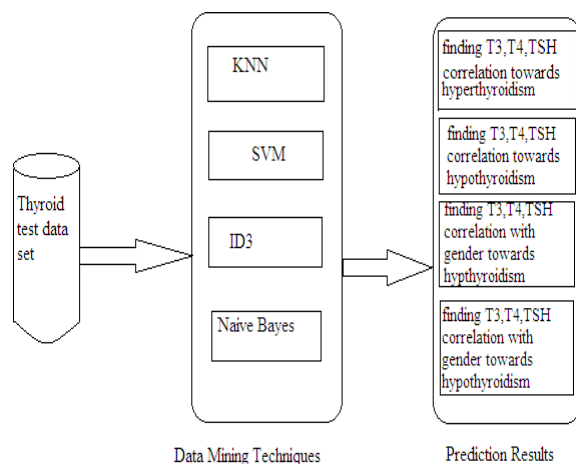


Fig.1 Proposed Classification model for thyroid disease prediction.

## VII. CONCLUSION

Thus this survey is needfull to identify how to predict the thyroid disorder at earlier stage using data mining techniques.DataMining classification algorithms are used to diagnose the thyroid problems and gives different level of accuracy for each techniques.These techniques help to minimize the noisy data of the patient's data from the data bases.Data mining Algorithms such as KNN, Naïve bayes, Support vector machine ,ID3are considered for the study.These various algorithm results are based on speed , accuracy and performance of the model and cost for the treatment.Also these classification of effective data are helps to find the treatment to the thyroid patients with better cost and facilitates the management.

## REFERENCES

- [1] Roshan Banu D, and K.C.Sharmili "A Study of Data Mining Techniques to Detect Thyroid Disease" International Journal of Innovative Research in Science, Engineering and Technology (Vol. 6, Special Issue 11, September 2017
- [2]Irina IoniNă and Liviu IoniNă" Prediction of Thyroid Disease Using Data Mining Techniques"The Classification Technique for Talent Management using SVM, (ICCEET), 978-1- 4673-0210-4/12, pp. 959-963, 2017

- [3] Khushboo Taneja, Parveen Sehgal, Prerana "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology (E-ISSN: 2321-3264) Vol. 3, No. 2, April 2016
- [4] Hanung Adi Nugroho, Noor Akhmad Setiawan, Md. Dendi Maysanjaya," A Comparison of Classification Methods on Diagnosis of Thyroid Diseases"IEEE International Seminar on Intelligent Technology and Its Applications,2017
- [5] M.N Das ,Brojo Kishore Mishra ,Shreela Dash," Implementation of an Optimized Classification Model for Prediction of Hypothyroid Disease Risks" 2017 IEEE Conference on Big Data and Analytics (ICBDA)
- [6] Rasitha Banu,G.— "Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data MiningTechnique", IJTRA
- [7] Izdihar Al-muwaffaq and Zeki Bozkus" MLTDD : USE OF MACHINE LEARNING TECHNIQUES FOR DIAGNOSIS OF THYROID GLAND DISORDER" Proceeding of Annual International Conference of the IEEE Engineering in Medicine and Biology Society,2017
- [8] [8]Umadevi S , Dr .JeenMarseline K.S,"Applying Classification Algorithms to Predict Thyroid Disease" International Journal of Innovative Research in Science, Engineering and Technology (Vol. 6, Special Issue 10, September 2017
- [9] Dr. Srinivasan B , Pavya K " Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov - 2016
- [10] K. Rajam and R. Jemina Priyadarsini"A Survey on Diagnosis of Thyroid Disease Using Data Mining Techniques" ,IJCSMC, Vol. 5, Issue. 5, May 2016, pg.354– 358.
- [11] Rouhani Modjtaba and Mansouri Kamran "Multi- classification approach for detecting Thyroid attacks", International Journals Pharm Bio Sci, pp. - 4(3),July 2017.
- [12] Gurmeet Kaur, Er.Brahmaleen Kaur Sidhu, — Comparison of anfis neural network with several other anns and svm for diagnosing hepatitis and thyroid diseases ,IRJET Volume1, Issue 11, July-2016
- [13] Saeed Shariati and Mahdi Motavalli Haghighi, Proposing Neural Network Efficient Training Model for Thyroid Disease Diagnosis In Proceedings of IEEE IACSIT 2017, pages 526-528.
- [14] Aytürk Keleş and Keleş, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov -2017 34.1 (2017): 242-246
- [15] [http://archive.ics.uci.edu/ml/machine-learning-databases/ thyroid-disease/hypothyroid.](http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/hypothyroid)