

# Thyroid Disease Detection and Classification using Machine Learning Techniques: A Review

Anuradha Shyam<sup>1\*</sup>, Mohanrao Mamdikar<sup>2</sup>, Pooja Patre<sup>3</sup>

<sup>1,2,3</sup>Computer Science and Engineering, Vishwavidyalaya Engineering College, Lakhapur, India

\*Corresponding Authors: anuradha750shyam@gmail.com Mobile: +91-9406120923

DOI: <https://doi.org/10.26438/ijcse/v7i3.11211125> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 22/Mar/2019, Published: 31/Mar/2019

**Abstract**—: The thyroid is one in every of the foremost necessary organ in our body. It secretes thyroid hormones that area unit to blame for dominant metabolism. The less secretion endocrine causes adenosis and far secretion of thyroid causes glandular disease. For deciding, data processing technique is principally employed in tending sectors, sickness identification and giving higher treatment to the patients. during this paper we've got bestowed an summary and comparison of assorted existing data processing techniques used for thyroid diseases identification. most ordinarily used techniques area unit call Tree, Support Vector Machine and Neural Networks that has been resulted as a high accuracy. the most objective of this study are to hold out the survey of knowledge mining techniques accustomed identification of assorted thyroid ailments, to gift the techniques used and its accuracy.

**Keywords**:-Thyroid diseases, Neural network, Support Vector Machine, Decision tree, KNN, Learning Vector Quantization, etc.

## I. INTRODUCTION

In the life science field, the foremost monotonous and difficult task is to produce unwellness identification at early stage with higher accuracy. knowledge mining is outlined as a method of extracting usable knowledge from an outsized knowledge set of any data to seek out some patterns. prophetic data processing plays a necessary role in medical field for identification totally different diseases. strip-mined pattern provides a basis for the analysis of risk factors for several unwellness.

The thyroid may be a butterfly-shaped organ set within the front of the neck just under the Adams apple. The organ wraps round the cartilaginous tube (trachea) and encompasses a form that's almost like a butterfly fashioned by 2 wings (lobes) and hooked up by a middle half (isthmus). The organ [1] works sort of a little industrial plant that uses iodine (mostly from the diet in foods like food and salt) to supply thyroid hormones. These hormones facilitate to control the body's metabolism and effects processes, like growth and different necessary functions of the body [1]. the 2 most vital hormones are tetraiodothyronine (T4) and thyroid hormone (T3), representing ninety nine.9% and 0.1% of thyroid hormones severally. The secretion with the foremost biological power is truly T3. Once free from the organ into the blood, an outsized quantity of T4 is regenerate to T3 - the active secretion that affects the metabolism of cells throughout the body [1]. Thyroid disorders embody such

unwellness and conditions as graves disease, thyroid nodules, Hashimoto's redness, trauma to the thyroid, thyroid cancer and birth defects. These embody being born with a defective thyroid or while not a thyroid [4]. Thyroid disorders will cause the thyroid to become hyperactive (hyperthyroidism) or hypoactive (hypothyroidism). Thyroid disorders lead to a speed of the body's chemical processes and metabolism with symptoms like weight gain, fatigue and depression [1,2,3]. Cases of the disorders are a lot of common in ladies than in men. Hashimoto's redness happens most frequently in females between the ages of thirty and fifty years and seems to own a genetic element as a result of it will run in families. individuals over the age of fifty WHO have high blood pressure or coronary-artery disease are in danger for diagnosis [2,3,4]. they're typically misdiagnosed as depression, aging, or different causes of fatigue, fatigue, or forgetfulness. Therefore, a thyroid disorder may be a risk that has got to not be dominated go in identification of conditions like depression, fibromyalgia, lupus, sleep disorders, and numerous different conditions. as luck would have it, identification of thyroid issues is comparatively specific by blood tests of hormone levels. Thyroid disorders are treatable, but, may be serious if untreated. Any suspicion of thyroid issues has to be confirmed promptly by a doctor.

## II. LITERATURE REVIEW

Nikita Sigh et.al [3], have all over that SVM is way far better classifier as compared to KNN and Bayesian.

Accuracy of SVM is regarding eighty four.62%. KNN found the closest neighborhood mechanically. It depicted by the graph every vertices having object. Bayesian supported the chance classification which supplies the sample information belongs to a category.

Taher A. et.al. planned a way for thyroid sickness identification victimization ANFIS. Classification of information from the University of American state, Irvine (UCI) machine learning information set repository was performed to judge the effectiveness of the Neural-fuzzy classifier on real-world information, and to facilitate comparison with different classifiers. [4]. The dataset contains three categories and 215 samples. These categories square measure appointed to the values that correspond to the hyper-, hypo-, and traditional perform of the ductless gland. Gurmeet Kaur1 G. et.al. planned a way for thyroid sickness identification victimization ANN classification technique[5]. during this work, they thought of the various artificial neural networks therefore on succeed the simplest result by scrutiny their performance on the premise of PSNR price to succeed in the simplest doable answer. 3 neural network algorithms are investigated for identification of Thyroid sickness. In [6] varied data processing techniques like mathematician web, MLP, RBF Network, C4.5, CART, REP tree and call stump square measure accustomed develop classifier for identification of hypothyroid sickness and yielded ninety nine.60% accuracy. during this paper a knowledge set of total 3704 instances has been utilized in that 3481 belongs to class negative,194 belongs to remunerated hypothyroid class , ninety five belongs to primary hypothyroid class whereas two belongs to secondary hypothyroid class. The twenty nine attributes were used for classification of information. the whole work is simulated in machine learning tools.

Behnam S. presents a study on thyroid sickness identification by victimization neural networks with second order coaching algorithmic program. The genetic algorithmic program was accustomed notice the optimum network structure with high classification accuracy. 3 totally different proportions of training/testing teams square measure fashioned. per the results, it absolutely was seen that neural network structures can be with success accustomed facilitate identification of thyroid disease[7]. F.Temurtas et.al.[8], realized the identification by multilayer, probabilistic, and learning vector quantization neural networks were enforced on thyroid sickness.Dataset obtained from UCI machine learning info was used and therefore the achieved accuracy ninety two.96 %,94.43% and 89.79% severally. The dataset that consists of the thyroid sickness measurements contains 3 categories with 215 samples within which one hundred fifty samples taken for traditional,35 taken for hyper and30 taken for hypo and five attributes were taken. In [9] A CAD system PCA-ELM is developed for helping the identification of thyroid

sickness. Experimental results established that the planned system performed considerably well in characteristic among gland disease, ad enosis and traditional ones. it absolutely was determined that PCA-ELM achieved the very best classification accuracy of ninety eight.1% and mean classification accuracy of ninety seven.73% using10-fold cross-validation. The experimental outcome showed that PCA-ELM performed better than PCA-SVM in terms of classification accuracy with shorter run time. The 5 attributes were used for this work.

In [10] have developed new classification model like ensemble of C four.5 and random forest for classification of thyroid information. Feature choice vie a crucial role to boost the performance of the developed model. The planned model offers high classification accuracy of ninety nine.47%. In [11] during this work, the information set used for this study on hypothyroid is taken from University of American state Irvine (UCI) data repository. the whole analysis work is enforced in more than. 2 data processing techniques like J48 and call stump tree square measure accustomed classify hypothyroid sickness and achieved high accuracy of ninety nine.57% and 95.38% severally. during this work twelve attributes square measure used. Qureshi M.A. et.al [12] proposes totally differentcall decision trees algorithms classify different thyroid-related diseases. initial they applied feature reduction ways to get rid of ten inappropriate options from twenty nine options. They found the between ninety seven.43% and 99.18% for various thyroid tasks. Ahmed J. et.al. [13] planned a system TDTD (that may be a distinctive technique) for prediction of missing values in medical datasets, finally for classification they used support vector machine. They achieved accuracy of ninety five.7% on UCI dataset of thyroid sickness. Azar A.T. et.al. [14] planned a fuzzy based mostly technique for feature choice and have reduction, finally for classification purpose fuzzy agglomeration is employed they achieved accuracy between ninety eight to ninety nine.5% on UCI dataset.

Keles et al. [15] determined thyroid illnesses to have a master framework that called ESTDD (master framework for thyroid ailment determination), We found fluffy standards by utilizing neuro fluffy strategy, which will be employed in ESTDD framework. The exactness was accomplished 95.33%. F.Temurtas [16] understood the analysis by multilayer, probabilistic, and learning vector quantization neural systems were actualized on thyroid infection dataset which was taken from UCI machine learning database was utilized and the accomplished precision 92.96 %,94.43% and 89.79% individually. In [17] A CAD framework PCA-ELM is produced for helping the conclusion of thyroid illness. Trial results demonstrated that the proposed framework performed altogether well in recognizing among hyperthyroidism, hypothyroidism and ordinary ones. It was seen that PCA-ELM accomplished the most noteworthy

order precision of 98.1%. In [18] deliberate methodology for prior conclusion of Thyroid ailment utilizing back proliferation calculation utilized in neural system. ANN has been produced dependent on back proliferation of mistake utilized for before expectation of ailment. In [19] The creators examined and looked at four characterization models: Naive Bayes, Decision Tree, Multilayer Perceptron

and Radial Basis Function Network. The outcomes show a huge exactness for all the arrangement models. In [20] they have grown new order demonstrate like gathering of C4.5 and irregular timberland for characterization of thyroid information. Highlight choice assumed a vital job to enhance the execution of our created model. Our proposed model gives high characterization exactness of 99.47.

**Table 3.1 Problem Identification**

Authors	Title	Findings
Qureshi M.A. et.al [12]	“Expert Advice Ensemble for Thyroid Disease Diagnosis”	1. Decision Tree Classifiers are used for classification. 2. Accuracy ranges from 98- 99.18% 3. UCI dataset used. 4. 10 features from 29 are used.
Ahmed J. et.al. [13]	TDTD: Thyroid Disease Type Diagnostics”	1. Support vector machine used for classification. 2. Accuracy ranges from 95.7% 3. UCI dataset used.
Azar A.T. et.al. [14]	Expert System Based On Neural-Fuzzy Rules for Thyroid Diseases Diagnosis	1. Fuzzy clustering is used for classification. 2. Accuracy ranges from 98-99.5%. 3. UCI dataset used. 4. for future work they suggest the Neuro – fuzzy model for better classification or for data mining
Suman Pandey et al [21]	Comparative Study on Classification of Thyroid Diseases	1. Average accuracy claimed 99.47 % 2. Total 29 features of UCI dataset are used for classification. 3. Random forest method used for classification
G. Rasitha Banu [22]	A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease	1. Accuracy of classification 95.38% 2. Total 12 features used for classification. 3. Decision tree classifier is used

Artificial neural systems (ANN) have developed because of reproduction of organic sensory system, for example, the mind on a PC. Counterfeit Neural systems are spoken to as a lot of hubs called neurons and associations between them. The associations have loads related with them, speaking to the quality of those connections[20]. These days neural systems can be connected to issues that don't have algorithmic arrangements or issues for which algorithmic arrangements are too perplexing to be in any way found. In others words the sort of issues in which data sources and yields factors does not have an unmistakable connection between them. Most neural system engineering has three layers in its structure. First layer is input layer which gives an interface the earth, second layer is shrouded layer where calculation is done and last layer is yield layer where yield is put away. Information is proliferated through progressive layers, with the last outcome accessible at the —output layerl. A wide range of sorts of neural systems are accessible and multi layer neural systems are the most prominent. MLP ubiquity is because of more than one shrouded layer in its structure which helps now and then in tackling complex issues which a solitary concealed layer neural system can't settle.

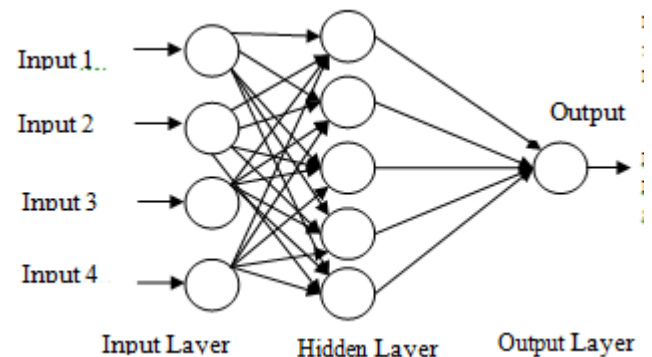


Figure1.1 Artificial Neural Network Model

**III. CLASSIFICATION TECHNIQUES IN DATA MINING**

**A. ANN-Based Support in Medicine**

**B. SVM-Based Support in Medicine**

“Support Vector Machine” (SVM) may be a supervised machine learning algorithmic program which may be used for each classification or regression challenges. However, it's principally employed in classification issues. during this algorithmic program, we tend to plot every knowledge item as a degree in n-dimensional area (where n is variety of options you have) with the worth of every feature being the worth of a specific coordinate. Then, we tend to perform classification by finding the hyper-plane that differentiate the 2 categories fine.Support Vectors square measure merely the

co-ordinates of individual observation. Support Vector Machine may be a frontier that best segregates the 2 categories.

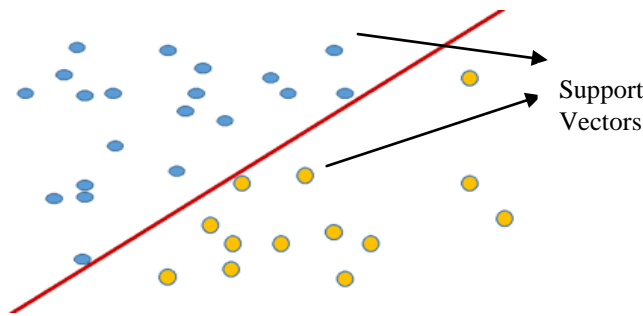


Figure 1.2 Support Vector Machine

**C. ANFIS-Based Support in Medicine**

ANFIS primarily based classifiers are booming in several application areas. but complicated application issues like reality medical image modelling analysis have emphasised the problems of feature set spatiality reduction and have linguistics interference, to capture the essential characteristics of given real image several options need to be extracted while not specific data of what property may best represent the image a priori. in fact generating a lot of options increase machine quality whereas not all such options square measure essential and a few might even cut back the figural power of the feature set because of mensuration noise. Amethod that may prompt a lot of important options, supported sample mensuration square measure thus extremely fascinating to support ANFIS support classification.

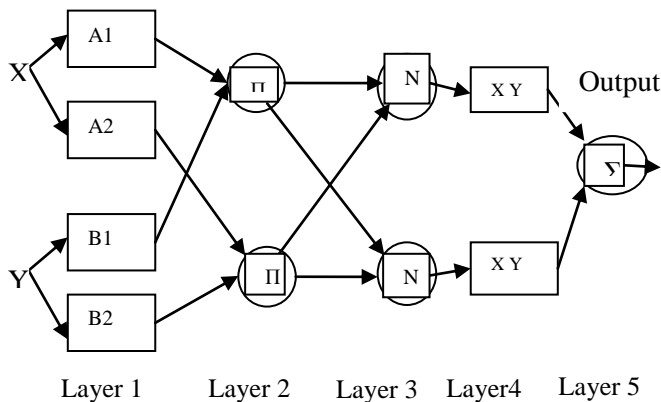


Figure 1.3 ANFIS Architecture

**D. Recurrent Neural Network (RNN)**

It is a category of artificial neural network wherever connections between units type a directed cycle. This creates an enclosed state of the network that permits it to exhibit dynamic temporal behavior. not like feed forward neural networks, RNNs will use their internal memory to method discretionary sequences of inputs. This makes them

applicable to tasks like world organisation segmental connected handwriting recognition or speech recognition.

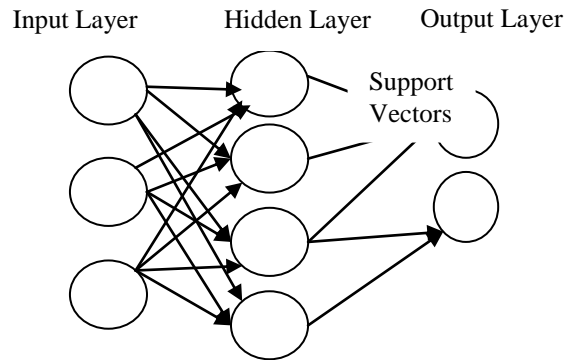


Figure 1.4 Recurrent Artificial Neural Networks

Some most commonly used classification techniques are: Artificial Neural Network, Support Vector Machine and Adaptive Neuro Fuzzy Inference System.

**E. QDA (Quadratic Discriminate Analyzer):**

In this case we've over 2 categories, the analysis employed in the derivation of the Fisher discriminate is extended to search out a topological space that seems to contain all of the category variability. Suppose that every of C categories includes a mean and also the same variance.

**IV. CONCLUSION**

In this survey we study number of data mining techniques which has been used by different researchers to detect or predict the thyroid disorder. Different Researchers have proposed different techniques to predict the thyroid disorder and different kinds of accuracy level as per used techniques. Here, comparison has been done on UCI dataset of thyroid disease and we found that hybrid classification models like neuro-fuzzy model and random forest with support vector techniques gives a better result than other single classifiers. So for future work we suggest to use hybrid modles rather that any single model for better classification result.

**REFERENCE**

[1] Dr. Sahai BS, Thyroid Disorders[online]. Available :<http://www.homoeopathyclinic.com/articles/diseases/thyroid.pdf>.  
 [2]<http://www.foxnews.com/health/2012/02/10/hypothyroidism-versus-hyperthyroidism.html>  
 [3]Nikita Singh, Alka Jindal, "A Segmentation Method and Comparison of Classification Methods for Thyroid Ultrasound Images", International Journal of Computer Applications (0975 – 8887) Volume 50 – No.11, July 2012  
 [4] Mary C. Frates, Carol B. Benson, J. William Charboneau and Edmund S. "Management of Thyroid Nodules Detected at US: Society of Radiologists in US consensus", conference statement management of thyroid nodules detected at US Volume 237, Number3.

- [5] 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
- [6] Shivaneepandey, RohitMiri, S. R. Tandan, "Diagnosis and Classification of Hypothyroid Disease Using Data Mining Technique", TJERT, June 2013.
- [7] AnupamShukla, PrabhdeepKaur, RituTiwari and R.R. Janghel, Diagnosis of Thyroid disease using Artificial Neural Network. In Proceedings of IEEE IACC 2009, pages 1016-1020.
- [8] FeyzullahTemurtas" A comparative study on thyroid disease diagnosis using neural networks" Expert Systems with Applications 36 (2009) 944–949.
- [9]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.
- [10] SumanPandey, 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.
- [11]G. RasithaBanu" A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease" International Journal of Computer Sciences and Engineering Volume-4, Issue-11 2016.
- [12]. Muhammad Anjum Qureshi, Kubilay Eksioglu, "Expert Advice Ensemble for Thyroid Disease Diagnosis", IEEE, 2017.
- [13] Jamil Ahmed, M. Abdul Rehman Soomrani," TDTD: Thyroid Disease Type Diagnostics", IEEE, 2016.
- [14]. Ahmad Taher Azar , Aboul Ella Hassanien, "Expert System Based On Neural-Fuzzy Rules for Thyroid Diseases Diagnosis", IEEE, 2018.
- [15] Keles, et al ESTDD: expert system for thyroid diseases diagnosis. 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] Zhang GP, Berardi. An investigation of neural network in thyroid function diagnosis. Health Care Management Science, 1998;1:29-37
- [21] 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.
- [22]G. Rasitha Banu " A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease" International Journal of Computer Sciences and Engineering Volume-4, Issue.