# Prediction of Thyroid Using Machine Learning Techniques

**Shiva Ranjani J, The Oxford College of Engineering Bangalore & India, shivaranjani4848@gmail.com**

**Monica S, The Oxford College of Engineering Bangalore & India, preeethureddy98@gmail.com**

**Suhas Shetty, Sapathagiri College of Engineering Bangalore & India, suhasshetty570@gmail.com**

**Abstract** **Thyroid disease is one of the leading causes of the emergence of medical diagnosis and prognosis, and its inception is a difficult medical research axiom. The thyroid gland is one of our body's most important organs. Thyroid hormone secretion is caused by metabolic regulation. Hyperthyroidism and hypothyroidism are two of the most common thyroid gland disorders. The thyroid gland produces thyroid hormones and regulates the body's metabolic rate. We used data cleansing techniques to generate enough raw data for an analysis of the patient's risk of developing the thyroid gland. Machine learning plays an important role in the disease prediction process. Based on data from the UCI machine learning repository, this white paper describes the analytical and classification models used in thyroid disease. When solving complex learning tasks like the, it is essential to have a good knowledge base that can be fixed and used as a hybrid model. This article also proposed various machine learning techniques and diagnoses for thyroid gland prevention. Machine learning algorithms, such as support vector machines (SVMs), artificial neural networks (ANNs), and decision trees, were used to predict a patient's risk of developing thyroid disease.**

**Keywords — thyroid disease, prediction models, machine learning algorithms.**

## I. INTRODUCTION

Computational biology advances are being used in the healthcare industry. This enabled us to collect previously stored patient data in order to predict medical illness. For early disease detection, various intelligent prediction algorithms are available. There are many records in the medical information system, but there is no intelligent system that can easily analyse the disease.

The machine learning algorithm will play an important role in solving the complex, non-linear problems associated with the development of predictive models over the course of hours. A predictive model is required for each disease to determine feature as accurately as possible, which can be chosen from a variety of datasets that can easily be used as a classification of healthy patients. Misclassification, on the other hand, may subject healthy patients to unnecessary treatment. As a result, the highest cardinality is the ability to predict any disease associated with thyroid disease.

The thyroid gland is an endocrine gland located in the throat. It aids in the secretion of thyroid hormone and essentially affects the rate of metabolism and protein synthesis by straightening at, the reduced part of the human neck under the Adam's apple. Thyroid hormones aid in the regulation of metabolism in the body by measuring how quickly the heart beats and how quickly the body's metabolism burns calories. Thyroid hormones, levothyroxine (abbreviated T4) and triiodothyronine, are produced by the thyroid gland (abbreviated T3).

These hormones are essential in manufacturing, as well as overall design and monitoring, to regulate body temperature [1] [2]. Thyroxine (T4) and triiodothyronine (T3) are two active hormones produced by the thyroid gland [1] [2] [3]. These hormones are required for protein metabolism. Absorption and transfer of energy in all parts of the body, as well as diffusion of body temperature. Iodine is considered the most important building material for the thyroid for these two thyroid hormones, i. H. (T3 and T4), and has been reflected in some specific problems, some of which are particularly widespread. Thyroid hormone deficiency causes hypothyroidism, while thyroid hormone excess causes hyperthyroidism. There are many causes associated with hyperthyroidism and hypothyroidism. There are kinds of medicines. Thyroid surgery is prone to ionizing radiation, constant thyroid sensitivity, iodine deficiency, and a deficiency of the enzyme that produces

 DOI : 10.35291/2454-9150.2022.0335 

thyroid hormone [4].

## II. LITERATURE SURVEY

There has been a lot of work done in recent years to diagnose the different thyroid illnesses. Many authors have made use of various data mining techniques. The authors demonstrated an adequate method and assurance in identifying disorders similar to thyroid disease through work that involves diverse datasets and algorithms coupled with future work to get effective and improved results. The purpose of this work is to analyse several strategies of data mining mechanisms and statistical attributes that have been popular in recent years for the interpretation of thyroid illnesses with the certainty by many authors to achieve varied possibilities and approaches. -there are various algorithms of machine getting to know counting random decision tree, naïve Bayes, SVM and ANN that are appreciably used within the common illnesses and inside the prognostic problems.

There are few functions which can be comprised of illnesses associated with heart disease [5], diabetes, Parkinson's, hypertension, the Ebola virus(EV) [19-20], diagnoses and forecasting, R-NA sequenced information analysis and allocation of biomedical imaging[23-25].regardless of, the development of a gadget gaining knowledge of-positioned ailment prediction mechanism and medical determination is a nontrivial task .

there may be crucial issues i.e., acquisition of facts, compilation and grouping which can be worn to teach the machine learning systems. within the actual pastime problems, estimation of large statistics sets in biomedical over a deep continuation are desired, and are essentially non-existent [12]

In [26] systematic approach for earlier designation of Thyroid illness victimisation back propagation algorithmic program employed in neural network. ANN delicate and establishes on back propagation of miscalculation that is being used for prior disease predictions. The impact of ANN is being trained with the empirical information and testing mechanisms that area unit borne out as a knowledge that wasn't in use throughout the method of coaching ANN concludes in a very sensible compliance with the preliminary information and indicates the advanced neural network that uses as a substitute for previous illness predictions. In the authors scrutinized and compare the four classification models specifically Naive Bayes, call Tree, Multilayer Perceptron and Radial Basis perform Network. The conclusion demonstrates a big accuracy for all the classification models. the choice Tree model exceeds by the opposite classification models. during this work twenty-nine attributes of data set is conscripted and enforced as a Feature choice technique i.e. Chi-Square, The datasets area unit being filtered by conducting the unattended coated filters on the attributes for conversion within the continuous values into nominal and thence cut back the twenty nine attributes to ten attributes

Machine learning (ML) may be a division of computer science and is infiltrated within the dimensions of research project at growing steps. Machine learning facilitates algorithms to review from expertise while not notably being prioritized [6]. Machine learning has been evoked by the input detonation that's connected with associate increasing machine capability, and classical medicine area unit a sophisticated homogenised recent knowledge science approach to strap the capabilities of the aesthetic knowledge [7]. to think about immense arrangements of knowledge, the actual tools explore in near clinically relevant liaison between input and output criterion. Factual analyses of surgical conclusions area unit eminently deceivable to amend surgical accords. Decisive aspects of surgical accords area unit description of the patient's comrade that aids from surgery within the arbitration. Machine learning permits computers to see from preceding knowledge to create meticulous predictions on current knowledge. The informative aspect makes terribly authoritative prediction algorithms that may copy the erst exotic communication in immense, convoluted sets {of knowledge |of knowledge |of information} and acclimatize to effective data aura [16].

The composite characteristics and also the curative procedures that square measure being employed within the thyroid disorders cater associate degree ample bunch of Byzantine and various information and therefore, a auspicious framework for the formulation of machine learning models [15]. This proposes associate degree ample probable for the use of machine learning models and braces a flourishing tendency towards rigorous medicines during which medicine square measure stitched to the actual patients. within the field of machine learning, an in depth divergence might be contrived amid supervised and unattended learning. supervised learning algorithms confirm from "labelled" coaching information to crop a model that accomplishes predictions on erst unreal information [8].

For unattended mechanism of learning, solely unlabelled information square measure possible and also the algorithms peeks to plus the analogies and devices, unattended learning algorithms might catch the Brobdingnagian variety of unlabelled genetics information as input and analyse erst anonymous assemblage of information. These algorithms might somehow be

 DOI : 10.35291/2454-9150.2022.0335 

dominant in antecedental erst arrangements in complicated information that don't seem to be primarily measurable by humans [9] and should be accustomed develop labels to finally train a supervised model. In standard programming, a technologist manually creates a group of data – "the programs" – to develop a crave output from a given set of input variables. In machine learning, the inputs square measure equipped beside the crave output and pc algorithms square measure inquired to derive the "rules from the classified coaching data". The processed learning method is associate degree adequate manner of deciphering Brobdingnagian abundance of information, planning hid communications in composite sets of information, and allering to dynamic aura [10-13].

In the learning mechanism, algorithms endeavour to quality the wonderful aggregation of input variables (features) and weights square measure enclosed to those options within the model, by that decreasing the inequality between the anticipated and substantial results. Machine learning is employed in coaching the system over huge databases, wherever the implemented machine learning techniques square measure recycled to develop abstraction devices or frame a model and use the accomplished devices or frame a model and use the accomplished devices or models in creating predictions within the future for anonymous cases [11-14]

## ARCHITECTURE OF THYRIOD PREDICTION SYSTEM

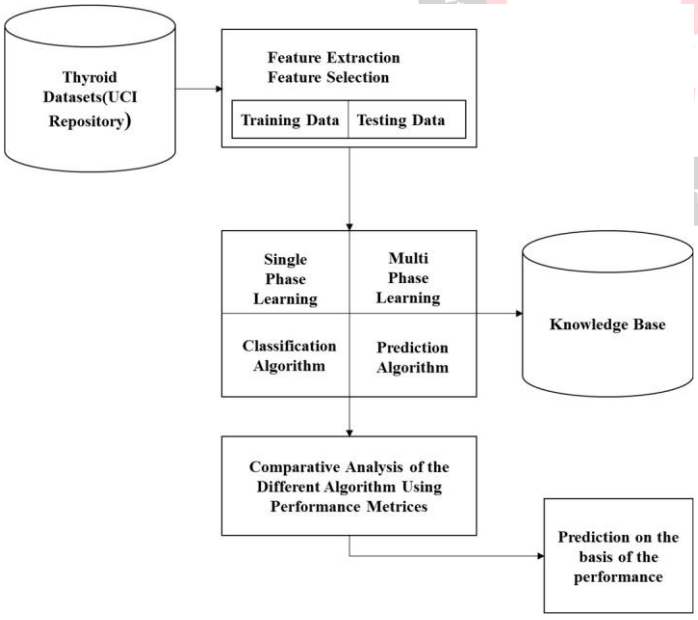



Fig1. Thyroid Prediction System

## III. METHODOLOGY

Supervised learning is Associate in Nursing info mining undertaking of inferring a perform from named training info. The coaching information comprised of an appointment of preparing illustrations. In managed adapting, each case is a couple comprising of Associate in Nursing info input object (commonly a vector) and also the desired output value (additionally referred to as the higher-up flag). A supervised learning calculation investigates the training info Associate in Nursing produces an in direct function, which might be used for mapping new illustrations [17]. Associate in Nursing ideal improvement can take into account the calculation to effectively decide the class names for unseen cases. this needs the taking in calculation to add up from the training information to hidden circumstances in an exceedingly "sensible" manner.

### *A. Attributes Used to diagnose thyroid Diseases:*

By analysing the on top of analysis work it's found that often used medical attributes to perform experimental work for the identification of thyroid diseases square measure given below in below table no.1.Among these attributes nearly each research worker has elite attributes to perform work for thyroid malady identification.

TABLE I. ATTRIBUTE FOR THE FEATURE SELSECTION

| 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 |

### *B. Performance Study of the proposed Algorithms:*

1. Artificial Neural Network: Neural network provides accustomed and a realistic approach in coaching absolutely the, separate furthermore as vector valued functions and may be a parallel system supported systema nervosum for learning real -valued, discrete-valued and vector valued functions and may be a parallel system supported human that have various corresponding alter parts essentially known to be because the neurons, working in a very agreement thanks to solve definite issues. Back propagation is that the most often worn learning techniques in ANN. it's a three layered design that's placed within the algorithms within the neural networks. it's comprised of three stratified designs i.e., input layer, hidden layer associate degreed an output layer. The foremost layer that's the input layer fueled the inputs into

 DOI : 10.35291/2454-9150.2022.0335 

this layer, the second layer i.e., a hidden layer- accords the output from the input layer associate degreed finally an output layer, beams the network's prediction. This miniature network helps to classify the new knowledge.

2. Support Vector Machine: Support vector machine is taken into account as associate degree various analysis algorithmic program that helps in activity the analysis in a very precise approach. Support vector machine is associate degree approach that's commenced with a plan of associate degree ace of separating hyper plane to aid within the distribution for sampling of knowledge. A hyper plane or multiple planes area unit created by the support vector machine classifier in high dimensional house. The coaching knowledge samples area unit being separated as a positive and negative knowledge sample by the hyper plane.

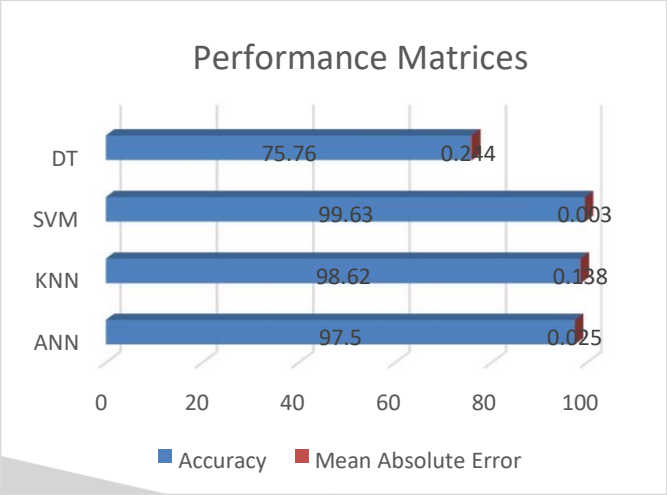
3. Decision Tree: Tree-like graph is employed in call tree classifier. a call tree is classed by its three nodes i.e., internal nodes, leaf nodes, and therefore the root nodes. the interior node connotes because the check on associate degree attribute, the leaf node connotes because the distribution of the category and therefore the root node connotes because the tree that has the highest most node. the 2 most extensive algorithms that area unit utilized in the as semblances of a call tree for diagnostic and prognostic model of thyroid diseases area unit C4.5 and ID3. Researchers use call Tree wide in tending field significantly to diagnose numerous thyroid diseases [18].

4. K- Nearest Neighbor: once given a coaching tuple K-Nearest Neighbor merely stores it and waits till it's given a check tuple. therefore, it's a "lazy learner" because it stores the coaching tuples or the "instances", they're conjointly called "Instance- primarily based Learners" [21-22]. Thus, k may be a positive whole number and decides what percentage neighbors influence the classification. "Closeness" delineates as a distance metric like "Euclidean Distance" or "Manhattan Distance".

## IV. Result and Discussion

The data sets for the thyroid diseases are possessed from the UCI machine learning repository. The work is dwelled with two totally different stages. The foremost section comprised of the set choice that's dead by adapting mutual info and prediction of the thyroid datasets done using ANN. Specifically within the interpretation of diseases neural networks square measure with success enforced within the distinctive fields within the medical realm. The certainty of the analysis for the datasets of the thyroid diseases square measure allotted because the elective look by each feature choice formula.

| Algorithm Used | Accuracy | Mean Absolute Error |
|---|---|---|
| ANN | 97.50 | 0.025 |
| KNN | 98.62 | 0.138 |
| SVM | 99.63 | 0.003 |
| DT | 75.76 | 0.244 |




Indeed, the growth of our unified representative may be a constructive mechanism to predict thyroid illness supported the restricted dataset that is available with America. The model will more be enhanced to any desired level by increasing the number of inputs and outputs and dynamic knowledge can be generated as additional knowledge are often fed thereto. In a nutshell, not solely we've developed a prototype integrated frame work to diagnose the thyroid disease however conjointly act as a call maker for diagnosing the thyroid illness.

## V. Conclusion

The intent of our work to be done more is to cater the analysis of individual techniques of machine learning that may be mobilized within the designation of thyroid diseases. There square measure various approachable analyses that square measure diagrammatical and square measure being employed within the latter years of adequate and competent thyroid illness designation. The analysis shows that totally technologies square measure utilized in all the papers showing different accuracies. In most analysis papers it's shown that neural network outperforms over alternative techniques. On the opposite hand, this can be conjointly providing support vector machine and call tree has also performed well. there's little doubt that researchers worldwide have earned plenty of success to diagnose thyroid diseases, however it's steered to decrease the number of parameters employed by the patients for designation of thyroid diseases. additional attributes mean a patient has got to endure a larger range of clinical tests that is each price effective still time intense. Thus, there's a

 DOI : 10.35291/2454-9150.2022.0335 

to develop such kind of algorithms and thyroid illness prophetical models that require minimum range of parameters of an individual to diagnose thyroid illness and saves each cash and time of the patient.

## REFERENCES

[1] L. Ozyılmaz and T. Yıldırım, "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, 2002) pp. 2033–2036.

[2] K. Polat, S. Sahan and 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,*vol. 32, 2007, pp. 1141-1147.

[3] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM," in *3rd International Conference on Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009*.

[4] G. Zhang, L.V. Berardi, "An investigation of neural networks in thyroid function diagnosis," *Health Care Management Scienc*e, 1998, pp. 29-37. Available: http://www.endocrineweb.com/ thyroid.html,(Accessed: 7 August 2007).

[5] V. Vapnik, Estimation of Dependences Based on Empirical Data, Springer, New York, 2012.

[6] Obermeyer Z, Emanuel EJ. Predicting the future— big data, machine learning, and clinical medicine. N Engl J Med. 2016; 375:12161219.

[7] Breiman L. StatisticalModeling: the two cultures.Stat Sci. 2001;16:199-231..

[8] Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol. 2017; 9:245-250

[9] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521: 452-459.

[10] Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S, Montazeri A. Artificial neural networks in neurosurgery. J Neurol Neurosurg Psychiatry. 2015; 86:251-256.

[11] Deo RC. Machine learning in medicine. Circulation .2015;132: 19201930.

[12] P.C. Austin, J.V. Tu, J.E. Ho, D. Levy, D.S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, J. Clin. Epidemiol. 66 (4) (2013)398–407.

[13] A.K. Pandey, P. Pandey, K.L. Jaiswal, A heart disease prediction model using Decision Tree, IUP J Comput. Sci. 7 (3) (2013) 43.

[14] S. Ismaeel, A. Miri, D. Chourishi, in: Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis, IEEE Canada International Humanitarian Technology Conference, 2015, pp. 1–3.

[15] L. Verma, S. Srivastava, P.C. Negi, A hybrid data mining model to predict coronaryartery disease cases using noninvasive clinical data, J. Med. Syst. 40 (7) (2016) 1–7.

[16] R. Rajkumar, K. Anandakumar, A. Bharathi, Coronary artery disease (CAD) prediction and classification—a survey, ARPN J. Eng. Appl. Sci. 11 (9) (2006) 5749–5754.

[17] Y.T. Lo, H. Fujita, T.W. Pai, Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations, J. Mech. Med. Biol. 16 (01) (2016) 1640010.

[18] B. Farran, A.M. Channanath, K. Behbehani, T.A. Thanaraj, Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study, BMJ Open 3

(5)(2013):e002457.

[19] M. Heydari, M. Teimouri, Z. Heshmati, S.M. Alavinia, Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran, Int. J. Diabetes Dev. Countries36 (2) (2015) 167–173.

[20] A. Colubri, T. Silver, T. Fradet, K. Retzepi, B. Fry, P. Sabeti, Transforming clinical data into actionable prognosis models: machine learning framework and fielded ployableapp to predict outcome of Ebola patients, PLoSNegl. Trop. Dis. 10 (3)(2016) e0004549.

[21] A. Jabeen, N. Ahmad, K. Raza, Machine learning-based state-ofthe-art methods for the classification of RNA-Seq data, in: N. Dey, A. Ashour, S. Borra (Eds.), Classification inBioApps. Lecture Notes in Computational Vision and Biomechanics, vol. 26, Springer,Cham, 2018.

[22] S.S. Ahmed, N. Dey, A.S. Ashour, D. Sifaki-Pistolla, D. Ba˘lasTimar, V.E. Balas, J.M. Tavares, Effect of fuzzy partitioning in Crohn's disease classification: a neuro-fuzzy based approach, Medical Biol. Eng. Compute. 55 (1) (2017) 101–115.

[23] Enas M.F. El Houby, A survey on applying machine learning techniques for management of diseases, Journal of Applied Biomedicine, January 2018

[24] Crohn's disease classification: a neuro-fuzzy based approach, Medical Biol. Eng. Compute. 55 (1) (2017) 101–115.

[25] 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.

[26] 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.

 DOI : 10.35291/2454-9150.2022.0335 