



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Special Issue 2, December 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

Intelligent of Thyroid Disease Prediction using Attributes of Machine Learning

Mrs.R.Vanitha¹, Mrs.T.Thivya Sindhu,

Department of Computer Science, Mannar Thirumalai Naicker College, Madurai, Tamilnadu, India¹.

Asst. Professor, Department of Computer Science, Mannar Thirumalai Naicker College, Madurai, Tamilnadu, India².

ABSTRACT: This paper has been written to access a source of reference for the research scholars who want to be ready to work in the area of prediction of thyroid disease using attributes of machine learning. Thyroid hormones affect the thyroid organ, which helps control the human body's digestion. Hypothyroidism or Hyperthyroidism is a major problem in India which is affected due to malfunctioning of thyroid hormones. Prediction of Thyroid Disease is a very difficult task at beginning stages with high accuracy. Thyroid gland is one of the most valuable organs in our body. The secretions of thyroid hormones are culpable in monitoring the metabolism. Due to the different machine learning techniques, compared widely used three algorithms namely logistic regression, decision trees and k nearest neighbour (KNN) algorithms to predict and diagnosis their activities in terms of accuracy.

KEYWORDS: Machine learning Decision Tree, Thyroid disease, k-nearest neighbour, Logistic regression, Prediction Model, Machine Learning Algorithms.

1. INTRODUCTION

Hyperthyroidism and hypothyroidism are most one of the two common diseases of the thyroid in the human body that releases thyroid hormones in measured proper rate of body's metabolism. The most common symptoms of thyroid are irregular body temperature, very low or high pulse rate, abnormal blood pressure, saturated weight gain or loss, swelling of hands and legs, unexpected hair loss, depression, mood changes frequently, metabolism, low memory power, skin diseases, heavy sweating, irregular stomach movements, hands shaking etc. These symptoms are expected to cardiac arrest. Thyroid disease may create unpredictable pulse rate and harm of the heart muscle in the human body.

Thyroid gland diseases are most endocrine disorders in the world, then leads to diabetes, advised by the World Health Organization. Hyperthyroidism and Hypothyroidism are affected about 20% and 10% of individuals person. In early disease, prediction and diagnosis according to doctors, are vital in preventing disease progress and even in death stage.



Fig 1. Thyroid System

The Fig 1 above depict the Thyroid systems. Hyperthyroidism is a major problem of disorder in which the thyroid gland releases so many thyroid hormones. Hyperthyroidism is affected by an increase in thyroid hormone rate level. Machine learning algorithms is one of the best solutions to find problems that are difficult to solve. Classification is a data extraction technique used to predict and identify more diseases, such as thyroid disease, which we are researched and classified here because machine learning algorithms play a significant role in classifying thyroid disease and because these algorithms are high performing and efficient and aid in classification. Then the application of computer learning and artificial intelligence in medicine fields back to the early days of the field, there has been a new movement to consider the necessary for machine learning- driven health care process.

II. OVERVIEW CONCERNING THYROID

The thyroid is an endocrine organ as shown in above figure 1. The capability concerning thyroid member is in accordance with abandon of thyroid hormones. Its extents in imitation of each single vile part thru the circulatory provision then control digestion or improvement. The extensive elements regarding thyroid part incorporate breath, blood flow, gut developments, temperature control, muscle working, assimilation and working concerning cerebrum. Any colourings action in the thyroid part might also affect the ordinary physiological work ethnic body. The thyroid hormone impacts the progress and improvement relying on the excuse on immunity. It is of two types hypothyroidism and hyperthyroidism.

Thyroid and its fitness impact: In an Indian study 42 million people are experiencing the basic endocrine infection. The Thyroid ailments are unique in relation in conformity with others as a long way as much their relative deceivability, utterance ease, scientific remedy mode availability. The incorrect advent over thyroid hormone influences wellbeing conditions.

Hyperthyroidism Increment creation in the thyroid hormones causes hyperthyroidism Graves.' disorder is some of the immune system difficulty up to expectation causes hyperthyroidism. The warning signs are glacial pores and skin, increment affectability to heat, weight reduction, raise pulse, hypertension, neck extension, anxiety, menstrual intervals abbreviate, visit gut traits then fingers trembling.

Hypothyroidism Reduction introduction among the thyroid hormone's motives Hypothyroidism. The Symptoms incorporates corpulence, low pulse, and increase between tranquil affectability, neck expanding, dead skin, arms deadness, hair issue, violent menstrual intervals and stomach associated issues. Thyroid hormones: The thyroid body produces are triiodothyronine (T3) or L-thyroxine (T4).

The thyroid hormones manage distinct metabolic exercises, for example, age regarding warmth, the utilization over sugars, protein and fats. The pituitary organ controls advent about triiodothyronine and L-thyroxine hormones. The Thyrotropin Stimulating Hormone from pituitary part is discharged now thyroid hormone is required then circles through the habit rule according to enter at thyroid organ. TSH at so much point animates the thyroid organs because the advent concerning T4 then T3 hormones. The beginning stage of thyroid hormone is confined by way of the input arrangement on pituitary organ. The TSH creation is less then T3, T4 are greater of the dissemination and TSH advent is more when T3, T4 are less.

Advantages of proposed system:

- We can predict the dimensions using best classifier.
- Dynamic nature in prediction.
- We can predict on our own by collecting the readings from clinical test

III. METHODOLOGY

The Dataset Description:

Database contains concerning suffered thyroid records. Every thyroid patient's document is comprising of 15 characteristics files beneath. Dataset is committed from UCI AI storehouse. Characteristic execute remain Boolean then steady esteemed are addicted beneath. Below figure 2 shows Thyroid Gland description, and fig 3 shows the data set Hypothyroid.csv.

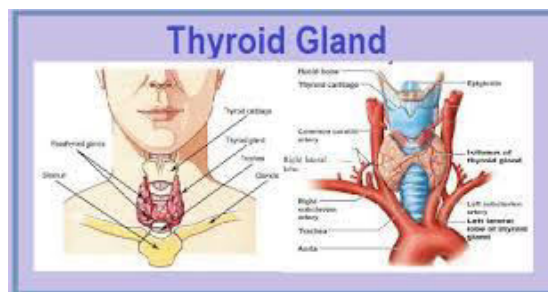


Fig2: Thyroid Gland

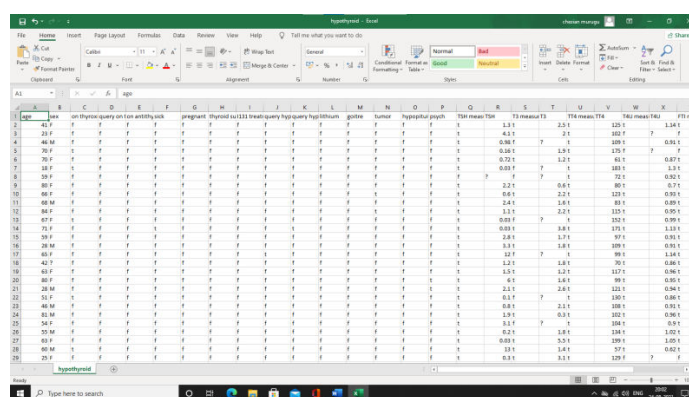


Fig 3: data set Hypothyroid.csv.

Table 1. Shows the attributes and patient description who has affected by Thyroid diseases. By analysing the above research work process, it is found that regularly used medical attributes to perform expert work for the diagnosis of thyroid diseases are given below in below table.1. From these attributes every researcher has pick out the attributes to perform work for thyroid disease diagnosis.

Table 1. Attributes and Patient description

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

Machine LearningTechniques:

Machine learning hasfive main different algorithms those are explained below:

1.Artificial Neural Network:

Neural network is familiar and a logical approach in training the absolute, distinct as well as vector valued functions and is a parallel system based on nervous system for learning real -esteemed, discrete-esteemed and vector-esteemed functions and is a parallel system based on human that have frequent corresponding alter elements fundamentally known to be as the neurons, working in a agreement way to solve final problems. Backpropagation is the most regularly worn learning techniques in ANN. It is a three-layered architecture that is placed in the procedures in the

neural networks. It is contained of 3 layered architectures i.e. input layer, hidden layer and an output layer. The foremost layer that is the input layer fuelled the inputs into this layer, the second layer i.e. a hidden layer- accords the output from the input layer and finally an output layer, gives the network's prediction. This miniature network helps to organize the new data.

2. Support Vector Machine:

Support vector machine is measured as a various research algorithm that helps in performing the study in a precise way. support vector machine is a method that is commenced with a concept of an ace of unscrambling hyper plane to aid in the delivery for sampling of data. A hyper plane or several planes are formed by the support vector machine classifier in high dimensional space. The drill data samples are being divided as a positive and negative data sample by the hyper plane. SVM helps the investigators in performing the analysis in a precise way. SVM is the discriminative algorithm, whose output is hyper plane which is used to catalogue the new classes. This hyperactive plane in two dimensional classes is a line intense a plane in binary parts. A single hyper plane or multiple planes can be formed by SVM in the high dimensional space.

Support vectors are those examples which are close to maximum-margin hyperplane. The training points in SVM are characterized $\{(R, S, R, S, R, N, S, N, 1, 1, 2, 2, \dots, \{, \})\}$ (1) In the equation (1), R_i is K-dimensional space vector and S_i represents the class to which a given vector is fitted. The division of training data by the hyper plane is signified by the general form $W \cdot R + B = 0$ (2) In equation (2), "W" denotes the K-dimensional space vector, vertical to hyperplane and "B" is scalar. Two hyperplanes which fit in to two dissimilar classes can be illustrated by following equations $W \cdot R + B = 1$ (3) $W \cdot R + B = -1$ (4) The main advantage of the SVM is that it avoids the over appropriate of data and rises the prediction accuracy.

3. Decision Tree:

Tree-like graph is castoff in decision tree classifier. A decision tree is confidential by its 3 nodes i.e. internal nodes, leaf nodes, and the root nodes. The internal node means as the test on an attribute, the leaf node connotes as the delivery of the class and the root node connotes as the tree that has the top most node. The two most universal algorithms that are used in the as presences of a decision tree for diagnostic and predictive model of thyroid diseases are C4.5 and ID3. Researchers use Decision Tree broadly in health care field particularly to diagnose various thyroid diseases.

A decision tree comprises 3 nodes i.e. root node, internal node and leaf node. The internal node achieves test on given attribute, based on the test the classes are allocated to leaf nodes. The root node visits on the top of the decision tree. Decision trees have the measurements to handle high dimensional data easily.

4. K- Nearest Neighbour:

When we given a training tuple KNearest Neighbour basically stores it and waits until it is given a test tuple. Hence it is a "lazy learner" as it stores the working out tuples or the "instances", they are also known as "Instance-Based Learners" [21-22]. Thus, k is a positive integer and decides how many neighbours impact the classification. "Closeness" delineates as a distance metric such as "Euclidean Distance" or "Manhattan Distance". Nearest near (KNN) KNN is some about the best regular AI calculations utilized because order, it companies an information point organized of whether its neighbours are characterized. KNN stores every alone reachable process and establishments a density measure.

The k-Nearest Neighbours tab (or KNN because short) is an easy tab in harmony with know yet in accordance with execute. The usage choice keeps obvious because of suggestion troubles yet wish stand shown exploiting the Iris blossoms classification issue. The model because of KNN is the whole making dataset. At the point then a port end is required because of a concealed facts example, the KNN scheming choice seem via the training dataset because of the k-most comparative cases. The predict faith over the almost similar occasions is summed above then lower back so the hope because of the inconspicuous occurrence.

5. Naive Bayes:

Naive is one of the most climbable and expert algorithms in data mining techniques. The Naive Bayes is known as excited learner because they have the ability of building a model immediately after a physical activity set is given. Naive Bayes classifier is constructed on Bayes theorem grounded on conditional probability. Naive Bayes theorem is stated as $P(C_j | D) = \frac{P(C_j) \prod P(D_i | C_j)}{P(D)}$ (5) In the equation (5), $P(C_j | D)$ denotes the probability of

instance „d“ being in class C_j . $P(d|C_j)$ denotes the probability of generating instance „d“ given a class C_j . $P(d)$ denotes the probability of instance „d“ occurring.

Naive Bayes (NB) In AI naïve bayes classifiers are a crew regarding frank probabilistic classifiers dependent about applying bayes hypothesis with sure self-sufficiency presumptions of the highlights. They are among the least complex Bayesian provision models. it shares a usual government as the nearness about a precise thing in group lamely after the proximity of partial mean element.

Data Flow process in various MachineLearningAlgorithms:

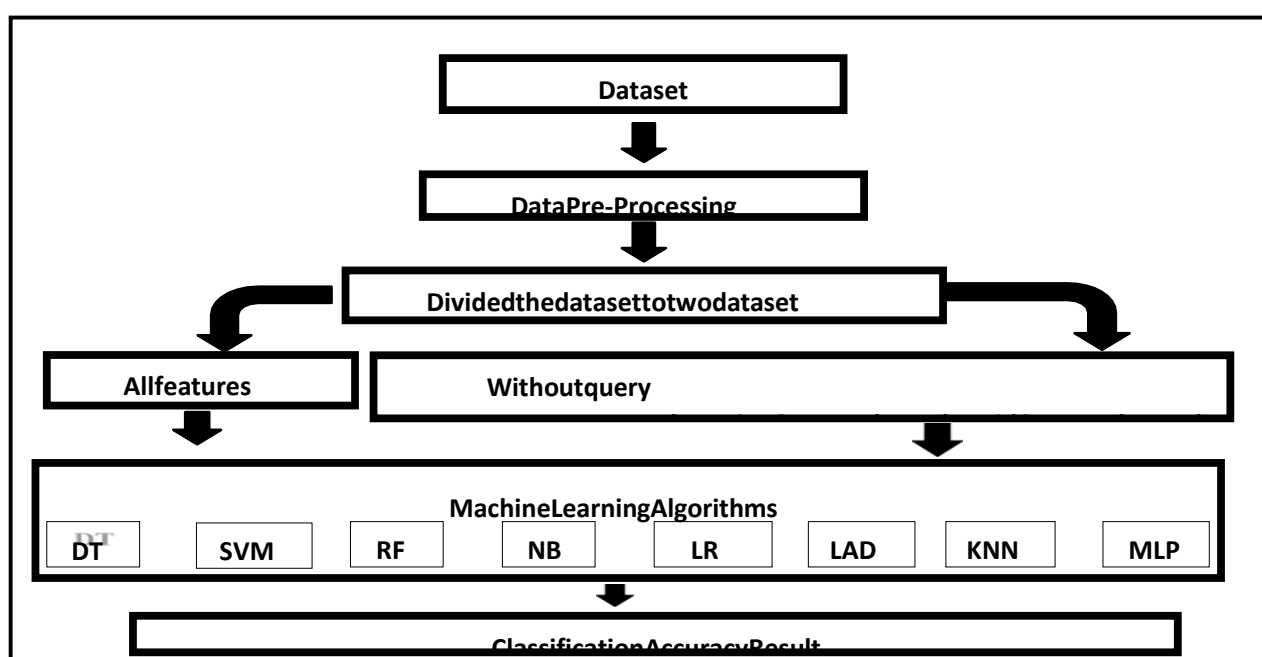
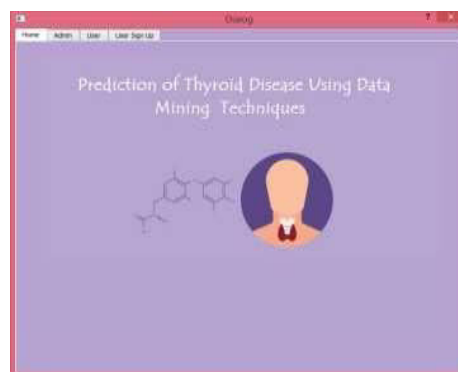


TABLE2
DISTRIBUTION OF INSTANCES IN DATASET

M a l e s			F e m a l e s						Tota l
			Pregnant			Non- Pregnant			
Normal	Hyp er	Hyp o	Normal	Hyp er	Hyp o	Normal	Hyp er	Hyp o	
192	199	201	90	90	90	207	199	196	

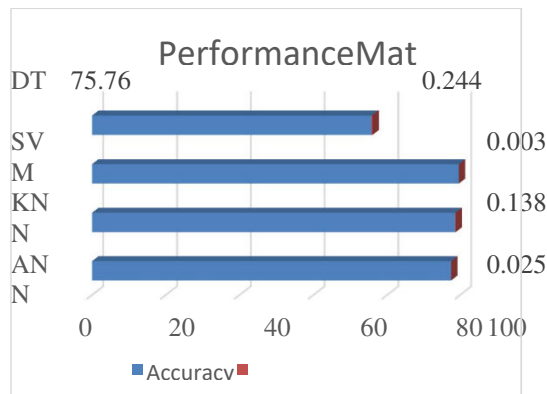
Figure :dialogflow

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



IV.RESULT AND DISCUSSION

The data sets for the thyroid disease have been controlled from the UCI machine learning repository. The principal phase contained of the subset selection that is executed by adapting mutual information and prediction of the thyroid datasets done using ANN. Specifically in the clarification of diseases neural networks are successfully enforced in the individual fields in the medical realm. The certainty of the study for the datasets of the thyroid diseases are assigned as the elected appearance by every feature selection algorithm



V. CONCLUSION

In this paper we have proposed to predict and diagnose the thyroid disorder in earlier stage using machine learning algorithms. Proposed algorithms help to minimize the noisy data of a patient. In the most research papers, it is shown that neural network outperforms over other techniques. Then, this is also given that support vector machine and decision tree has also performed in machine learning. Among, there is necessary to develop such type of algorithms and thyroid disease predictive models which require minimum number of parameters of a person to diagnose thyroid disease in human body and saves both money and time of the patient. We may have achieved the highest accuracy of 99.08% by SVM classifier. We may achieve the highest accuracy of 98.56% by naïve bayes classifier. Furthermore, this paper can be improved by applying various features selection algorithms to increase the performance of Thyroid Detection and prediction in person. In the future, the prediction algorithm can be further improved by adding many attributes to the dataset leading to more and more strong results in human body. From this paper, there is no doubt that researchers worldwide have a lot of success to prediction and diagnose thyroid diseases, but it is recommended to decrease the number of parameters used by the patients for prediction and diagnosis of thyroid diseases.



REFERENCES

1. S. Godara and R. Singh, "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", Indian Journal of Science and Technology, Vol. 910, March 2016.
2. Zhang, Guoqiang Peter, and Victor L. Berardi. "An investigation of neural networks in thyroid function diagnosis." Health Care Management Science Vol. 1, Issue 1, Sep 1998
3. Sunila, Rishipal Singh and Sanjeev Kumar. "A Novel Weighted Class based Clustering for Medical Diagnostic Interface." Indian Journal of Science and Technology Vol 9, Issue 44, Nov.2016.
4. Lavarello, Roberto J., Billy Ridgway, Sandhya Sarwate, and Michael L. Oelze. "Imaging of follicular variant papillary thyroid carcinoma in a rodent model using spectral-based quantitative ultrasound techniques." In Biomedical Imaging (ISBI), IEEE 10th International Symposium, pp. 732-735, 2013.
5. Savelonas, Michalis A., Dimitris K. Iakovidis, Ioannis Legakis, and Dimitris Maroulis. "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images." IEEE Transactions on Information Technology in Biomedicine, Vol. 13, Issue 4, July 2009.
6. Polat, Kemal, Seral Şahan, and Salih Güneş. "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis." Expert Systems with Applications, Vol 32, Issue 4, May 2007.
7. Kumari, Milan, and Sunila Godara. "Comparative study of data mining classification methods in cardiovascular disease prediction 1." IJCST, Vol. 2, 2011 ISSN: 2229-4333
8. Keleş, Ali, and Aytürk Keleş. "ESTDD: Expert system for thyroid diseases diagnosis." Expert Systems with Applications Vol 34, Issue 1, 2008
9. 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.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details