

Analysis of Different Machine Learning Techniques used in the Detection and Prediction of Thyroid Disease

Aarathi Rajagopalan Nair^{1*}, Himangi Pande²

¹Student, Department of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, India

²Faculty, Department of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, India

Abstract: When a sufficient amount of hormone is not produced in the body, this state is generally termed thyroid disease. The thyroid is the hormone responsible for the smooth functioning of the body. Ever since the application of data science in the medical domain researchers are putting their heart and soul to detect thyroid diseases. Through this paper, we are trying to review certain work done by different researchers cooperating the machine learning algorithms. We try to study the dataset, consider the features and accuracy of different models.

Keywords: Thyroid disease, Machine learning, Support vector machine, KNN, PNN, ANN

1. Introduction

A small organ which is located in the neck encases the trachea also known as the windpipe [1]. This small organ is the thyroid gland. The shape of the thyroid gland is a butterfly shape, with two wings and a middle portion. This is entirely extended around the neck. The thyroid gland produces thyroid hormone that controls many of the important functions of our human body. When the functions of the thyroid are not properly done that can affect the human body entirely. There may be conditions like hyperthyroidism which arises is too much thyroid hormone is produced and when a little amount of thyroid hormone is produced which causes a condition called hypothyroidism. In both cases, it should be considered crucial and medical attention should be taken.

1) Functions of the thyroid hormone

As mentioned above thyroid hormones affect each part of the body and they are [8]:

- Weight gain or weight loss is regulated or controlled by the thyroid hormone.
- Heartbeat speed is monitored.
- Body temperature
- The rate at which the food travels around the digestive tract.
- Muscle contraction is controlled.

- Death of a cell and replacing it with new cells are also controlled.
- Growth in children
- Brain maturity in children

Thyroid disease can occur to anyone irrespective of the gender, age etc. which implies it can occur to men, women, infants, elderly person, teenagers. There are cases reported where small infant at the time of birth, baby was affected by hypothyroidism. Thyroid is very common disease. According to the world health organization thyroid disease is a very serious as it is affecting more than 200 million people around the world.

2. Reasons of Developing a Thyroid Disorder

1) Hypothyroidism

- It could be congenital
- Iodine deficiency
- Autoimmune as in Hashimoto thyroiditis
- Post-operative
- Pituitary dysfunction
- Drugs like lithium, clomiphene, lithium, spironolactone

2) Hyperthyroidism

- Any growth in thyroid like adenoma, carcinoma
- Autoimmune like graves' disease
- Exogenous administration of thyroid hormones

B. Symptoms of Thyroid disorder

1) Hypothyroidism

- Increase in weight
- Tiredness or fatigue
- Forgetfulness is experienced
- In women if menstrual flow is frequent and heavy
- Voice will be hoarse
- Cold temperature will be hard to handle

2) *Hyperthyroidism*

- Psychological condition such as anxiety, nervousness, and irritability will be experienced.
- Trouble during sleep
- Decrease in weight
- Thyroid gland may be enlarged
- In women menstrual cycle will be stopped or it might be irregular.
- Sensitive to heat
- Eye irritation or having problems with vision.

Diagnostician of thyroid disease: Thyroid disorder is hard to detect as the symptoms are very common with other disorders. But there are tests which includes blood test, image testing and physical examining. Here in this paper there is study of different paper to show the innovation that took place in the field of thyroid detection with the help of machine learning and deep learning.

3. Study of Different Papers

The main challenge that a medical professional faces, in identifying thyroid diseases is the presence of nodules that are difficult to palpate [9]. The patients can also have wide range of presentation, ranging from hypothyroid, Euthyroid, hyperthyroid with or without nodules. It is also important to identify structural and functional changes. Such that it can be identified at early stage such that it can be utilized in ruling out cancer.

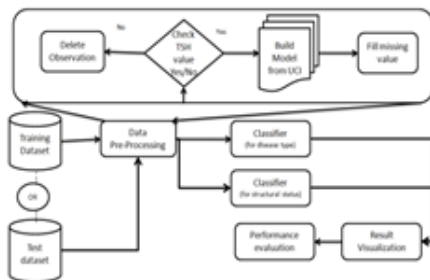


Fig. 1. A basic methodology how machine learning model completely works

In the particular paper the authors used a dataset which is taken from real world and was provided by Shaheed Muhtarma Benazir Bhutto medical university-Pakistan [1]. The dataset provided is used for both testing and training data. Cleaning of medical data is a tough task as it contains a lot of metabolic profile in which the values can end up in drastic changes. Therefore in the paper they have used a particular algorithm to fill up the missing value.

In the classification phase they divided the entire cleaned data set into two parts. First part of it is used for detecting the thyroid diseases which is mentioned above that is it may be hyperthyroid, hypothyroid or euthyroid. The other half of the divided dataset is used to check whether the person has got goiter or not. For both the cases they have used support vector machine for classification. Normal functioning of thyroid is necessary for overall wellbeing of the body. A thorough medical examination and a battery of blood test both

biochemical evaluations are necessary to diagnose thyroid related problems. Machine learning in classification plays a very significant problem. It also plays a very important role in medical domain. In this paper the authors tried to bring a detailed study using Naive Bayes and decision tree that is ID3.

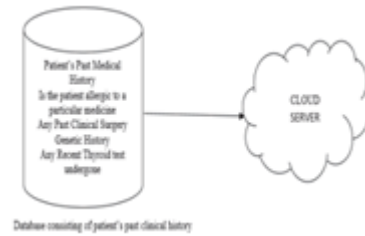


Fig. 2. Cloud server to store the medical details

In an another paper the authors have taken their dataset from Kaggle. In the dataset we have very different attributes which age(continuous), gender(M/F), Hyper Thyroid(F/T), Hypo Thyroid(F/T), pregnant(F/T), T3 value(continuous), TSH value(continuous), and T4 value(continuous) [2].

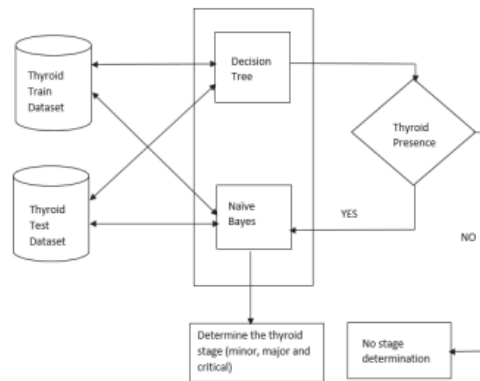


Fig. 3. Implementation of the Naive Bayes and Decision tree works

From the attributes mentioned the dataset is divided into training and testing set. This training is then passed on to supervised learning algorithm. A unsupervised learning algorithm is a machine learning technique which uses a set of learners such that it can get a previous knowledge about the set of data. Along with the above mentioned attributes the authors also took up in to consideration of many other factors such as medicines which are allergic to a particular patient, whether the patient has undergone any surgery in recent days and test related to thyroid where taken in the previous couple of days. By these informatio it saves much time of doctors and also saves the cost expenditure that will burdern the patients. Within simple terms we can conclude that it will ease the process of detection.

All the above mentioned details of the patients are stored in a cloud server which is dedicated for this process. This cloud server can be used up by any health care organization such that it can be used by anybody at any time. Both the method that is decision tree and naïve bayes both shows a significant role in classification of approximately 94.6%. The reason behind the accuracy is that cause of the size of the dataset. Simply we could conclude that if more number of dataset the accuracy will be

high. Thyroid is a vital organ in our body. Normal level of the products of thyroid gland is important for the normal functioning of the body. Any derangement from normal level can affect the metabolic activities in our body. Thyroid being an endocrine gland is usually affected in many of the familial syndromes. The data we have about the family can be used to predict the probability of an individual having a thyroid disease



Fig. 4. Mutiphase classification and prediction problem.

Further papers when we refer, apart from the normal machine learning algorithm the authors also tried to exploit the method of artificial neural network [3]. Neural network is pragmatic approach where the data is trained contain values which are absolute, discrete and including functions which are vector valued. Neurons are the basic building block of a neural network which is basically used to manipulate a particular problem in the way mentioned. In artificial neural network the most commonly or rather widely used technique is Backpropagation. A neural network consists of 3 layers that is the input layer, hidden layer and the output layer. The input is fed into the input layer. The output from the input layer fed into the hidden layers which then processes and pass on to the last layer which is the output layer.

K-nearest neighbor is a simple method of supervised learning under the machine learning algorithm. It simply stores all the data which is available and when a new data point is introduced, it is classified on the basis of the similarity present. It is a type of algorithm which is used in both the cases that is classification and regression. K-NN is also known as lazy learners as it stores the training data and knowledge or rather training is done immediately. Here K is a positive number such that the closeness between the points is the basic influence of the classification. Here the distance formula which is used is the Euclidean distance formula and in some cases they also used Manhattan Distance formula. Using the above the methods that is the K-Nearest neighbor, artificial neural network and the decision tree the corresponding accuracy obtained was 93.8%, 95.6% and 75.76%. In any of the experiment in the machine learning domain the confusion matrix plays a very significant

role [4].

Table 1
Confusion Matrix

	Class 1 Predicted class	Class 2 Predicted
Class 1 Actual	True Positive (TP)	False Negative (FN)
Class 2 Actual	False Positive (FP)	True Negative (TN)

Confusion matrix also known as the error matrix is a table which shows the actual condition which includes the positive and negative values with the calculated or rather concluded value which also include the positive and negative instances. On the basis of these parameters we calculate the performance of this particular model by calculating its accuracy, sensitivity, precision, specificity and sensitivity.

The ratio of TP+TN to the total that is TP+TN+FP+FN is called as the classification accuracy. A value ranging from 0 to 1 is taken up classification accuracy. Precision is the ration of the true positive conditions to the sum of the true positive and false positive. Similarly Recall is the ration of TP to TP+FN. If the value of precision =1 indicates that whatever may be the value of recall the classifier is the most suitable classifier for a particular instance. The Genetic makeup of an individual and the gene pool of a family can help us a lot in predicting the thyroid diseases. [5] For example MEN syndromes are a syndrome which runs in families and affects multiple endocrine organs. By studying the DNA of individuals and assessing it to axon level can help us not only predicting the pattern of disease but also the age of onset of disease in an individual.

In the further study of our project we come across a new term called as PNN. Probabilistic Neural network is a type of supervised neural network that is very commonly implemented the case of recognition of patterns, mapping of the non-linear data and similar application where it requires to the deep interpretation of the network which can be easily understood by the probability density. This method is very close to Bayes classification. The architecture of PNN is very similar to FNN, just a small change that it include 4 layers in which 1 layer us the pattern layer.

The anatomical location of thyroid makes it easily accessible for ultrasound examination. The glandular structure can easily be assessed [6]. Benign and malignant disorders have different appearances in ultrasound. Thyroid segmentation and volume reconstruction plays an important role in visualization of the gland when it is diseased as most of the diseases bring about a change in shape and size of the gland.

4. Conclusion

Medical domain has always advanced with the application of machine learning, deep learning or any other data science methods. From the entire paper we understand that thyroid is very important gland in our body. It is necessary to closely observe the changes that happen in this gland such that it helps in the smooth functioning of our body. Many machine learning techniques which include K-Nearest Neighbor, Probabilistic Neural Network, Decision Tree, Artificial Neural Network, and

Naïve Bayes., are being properly studied such that the accuracy of each algorithm can be improved. The use of such innovation in the medical domain with the help machine learning can save both time and money. Early detection of the diseases can lead to the betterment of one's life.

References

- [1] A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis Amulya.R.Rao Electronics and Communication Department Sri Jayachamarajendra and B.S.Renuka.
- [2] Ttd: Thyroid Disease Type Diagnostics Jamil Ahmed And M. Abdul Rehman Soomrani
- [3] Interactive Thyroid Disease Prediction System Using Machine Learning Technique Ankita Tyagi Ritika Mehra Aditya Saxena
- [4] Application Of Two-Class Neural Network-Based Classification Model To Predict The Onset Of Thyroid Disease Tabrej Khan
- [5] Thyroid Disease Diagnosis Based On Genetic Algorithms Using Pnn And Svm Fatemeh Saiti¹ , Afsaneh Alavi Naini² Tehran, Mahdi Aliyari Shoorehdeli¹ , Mohammad Teshnehlab¹
- [6] Thyroid Ultrasound Texture Classification Using Autoregressive Features In Conjunction With Machine Learning Approaches Prabal Poudel¹ , Alfredo Illanes¹ ,Elmer J. G. Ataide¹ , Nazila Esmaeili¹ , Sathish Balakrishnan¹ And Michael Friebe¹ (Senior Member, Ieee)
- [7] D. Selvathi, V. S. Sharnitha, "Thyroid Classification And Segmentation In Ultrasound Images Using Machine Learning Algorithms". In Signal Process Ing, Communication, Computing And Networking Technologies (Icscn), International Conference On Ieee. Pp. 836-841, July, 2011.
- [8] N. H. Mahmood, A. H. Rusli, "Segmentation And Area Measurement For Thyroid Ultrasound Image". International Journal Of Scientific And Engineering Research. vol. 2, No. 12, 2011
- [9] A. Keles,&Keles. (2008) "Estdd: Expert System For Thyroid Diseases Diagnosis" Expert Systems With Applications 34(1), Pp. 242–246.
- [10] K. Polat, S. Sahan, & 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 .(2007) Vol: (32), P.P 1141–1147.