International Journal of Research Publication and Reviews

Early Thyroid Disease Diagnosis Prediction

Siddhi Borse, Mr. Pritesh Patil

AISSMS Institute of Information Technology, Pune -411011, Maharashtra, India

ABSTRACT

Based on classification A significant part of many medical services is played by machine learning. The most important and difficult responsibility in the medical industry is to correctly diagnose the patient's health conditions and to start treating the sickness as soon as possible. Let's use thyroid disease as an illustration. The typical and conventional approaches to thyroid diagnosis include a careful examination as well as numerous blood testing. The major objective is to accurately and promptly identify the disease in its earliest stages. The use of machine learning techniques in the medical profession is crucial for accurate decision-making, accurate disease diagnosis, and patient cost and time savings. This study's objective is to use categorization to predict thyroid illness. following predictive modeling. This study aims to predict thyroid disease using predictive modelling classification, followed by binary classification using decision trees ID3 and Naive Bayes algorithms. The Decision Tree algorithm is used to test for the presence of thyroid in the patient after retrieving the Thyroid Patient dataset with the correct attributes. Additionally, if thyroid is found, the Naive Bayes algorithm is used to determine the patient's thyroid stage.

1. Introduction

Diagnosing thyroid illness is not an easy task. It calls for numerous steps. The typical, conventional approach is a thorough physical examination and numerous blood samples for blood tests. A model that can identify thyroid disease at its very earliest stages is therefore essential.

About The Thyrod

located near the base of the human neck, this endocrine gland has the shape of a butterfly.The thyroid gland plays a crucial part in the growth and development of the human body as well as in maintaining and balancing the human metabolism. Blood circulation, body temperature regulation, muscle strength, and brain function are among the critical functions carried out by the thyroid gland [1]. Therefore, adequate thyroid hormone secretion results in a healthy human body. Any damage to or improper functioning of the gland may adversely impact the normal human body functioning. The health of people will suffer if there is either little or high hormone secretion[2].

Motivation

The majority of people in the world are afflicted by thyroid diseases and disorders, which are common hormonal issues. Thyroiditis and thyroid cancer are two conditions and illnesses related to the thyroid. One of the most noticeable pure endocrine glands, the thyroid is found in the front of the neck and surrounds the trachea.

An increasing thyroid prediction accuracy will enhance thyroid problem identification, which is a measure of the likelihood that a better result will be obtained when more data are used.

Aim and Objective(s) of the work

Aim:

The aim of project is to predict Thyroid Disease at Early Stages of diagnosis
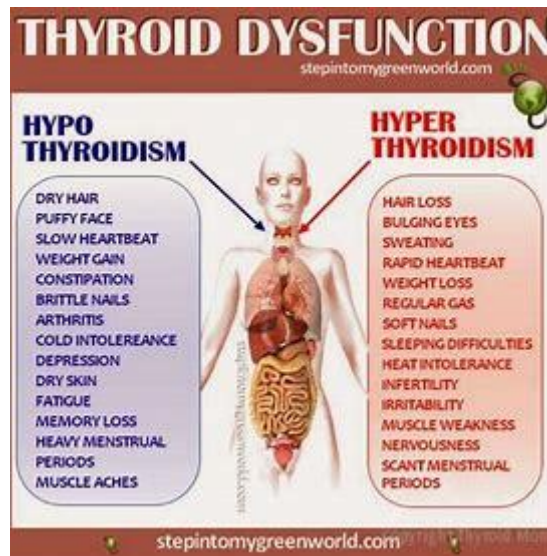
Project objectives:

Numerous issues that seem unconnected to one another, such as significant changes to your weight, energy level, digestion, or mood, might be brought on by thyroid issues.

Discover the symptoms of thyroid disorders so you can receive treatment if necessary. The thyroid is a tiny but potent gland with a butterfly form that is situated in the front of your neck.

2.CLASSIFICATION TASKS

A. Predictive Classification Modeling

The Classification Predictive Modeling operates by using the provided example input data to forecast the matching class label.

A strong train dataset is necessary for the development of a classification model in order for it to accurately learn behaviour and anticipate potential outcomes.

In Classification Predictive Modeling, there are four primary tasks.

Classification in Binary

Classification with Multiple Levels

Classification Using Multiple Labels

Uneven Classification

A classification using just two class tags is referred to as a binary classification. Decision Tree, Logistic Regression, and Nave Bayes are a few examples. More than two class tags are used in multi-class classification. When there are more classifiers, this type of categorization is highly helpful. Multi-Class Classification can also use the Binary Classification Algorithms. For this kind of categorization problem, Random Forest can serve as an illustration. In a multi-labeled classification, there are two or more class tags, and one or more of those class tags can also be used to predict the outcomes of other class tags. Multi-Label Decision Trees and Multi-Label Random Forests are a couple of examples. When there is an uneven distribution of the number of the example classes in a classification set, it is said to be imbalanced.

3. DIAGRAMMATICAL DESIGN

Proposed machine learning classification model for thyroid disease diagnosis

Thyroid System

Thyroid Dysfunction

4.TECHNOLOGY

1.Collecting data The Kaggle Machine Learning website's Thyroid Dataset is used[6].

2. Cleaning and processing the data The qualities of the dataset are taken into account in order of importance. The traits that are most directly linked to thyroid disease are taken into account, while the others are ignored. The values of the attributes might be continuous or boolean (True/False).

3. Machine learning model Among the qualities taken into consideration, a train dataset is produced and provided as the input to the classification-based supervised ML model.

4. Learning To produce the findings, the suggested model uses a decision tree and the Naive Bayes method.

- The thyroid patient record's inputted values are calculated by the decision tree algorithm. The train dataset is used as the basis for the calculation. Therefore, the accuracy of the algorithm increases as the number of records in the train dataset increases.
- The algorithm produces yes or no values, i.e., whether the patient has thyroid illness or not.
- The Naive Bayes method is then used to determine which stage the patient is now in if the patient's output value is true. This provides the medical team with a significant advantage in the quick diagnosis of thyroid disease and the avoidance of unneeded lab testing.
- There are three stages of thyroid disease in this patient: minor, major, and critical. If the Decision Tree produces a thyroid true or positive result, the Naive Bayes technique is used. The Naive Bayes algorithm is a supervised learning method used in machine learning that is based on the Bayes' Theorem.

5.COMPARATIVE ANALYSIS

The outcomes of the suggested hypothyroidism detection model are shown in this section.

the performance data of naive Bayes, decision tree, random forest, multiclass classifier, and ANN for each class, i.e., negative, compensated, primary, and secondary in the dataset, in terms of precision, accuracy, recall, F1-score, and ROC.

The outcomes demonstrate that the random forest model has the highest precision for negative, compensated, and primary classification, respectively, with precision values of 0.998, 0.97, and 0.957, respectively. Similarly, the random forest outperforms other models included in the study when it comes to recall, F1-score, and ROC.

6. CONCLUSION

In this section, the results of the suggested hypothyroidism detection model are displayed. The naive Bayes, decision tree, random forest, multiclass classifier, and ANN performance data in terms of precision, accuracy, recall, F1-score, and ROC for each class, including negative, compensated, primary, and secondary in the dataset. The results show that the random forest model, with precision values of 0.998, 0.97, and 0.957, respectively, has the highest precision for negative, compensated, and main classification. Similar to recall, F1-score, and ROC, the random forest performs better than other models used in the study.

REFERENCES

[1] "Prediction of Thyroid Disease Using Data Mining Techniques" by Bibi Amina Begum and Dr. Parkavi The 2019 ICACCS, the 5th International Conference on Advanced Computing & Communication Systems

[2] Ritika Mehra, Ankith Tyagi, and Aditya Saxena "Interactive Machine Learning Technique Thyroid Disease Prediction System" PDGC-2018, the 5th IEEE International Conference on Parallel, Distributed, and Grid Computing, will be held in Solan, India, from December 20 to 22.

[3] "An Intelligent System for Thyroid Disease Classification and Diagnosis" by Aswathi A. K. and Anil Antony Second International Conference on Creative Communication and Computational Technologies Proceedings (ICICCT 2018) Part Number: CFP18BAC-ART; ISBN: 978-1-5386-1974-2; IEEE Xplore Compliant

[4] "A Empirical study on Disease Diagnosis using Data Mining Techniques," by M Deepika and Dr. K. Kalaiselvi IEEE Xplore Compliant Part Number: CFP18BAC-ART; ISBN: 978-1-5386-1974-2; Proceedings of the Second International Conference on Inventive Communication and Computational Technologies (ICICCT 2018).

[5] "Classification of Thyroid Disease Using Data Mining Techniques," by Sumathi A, Nithya G, and Meganathan S. Volume 119, Number 12, 2018, 13881–13890, International Journal of Pure and Applied Mathematics

[6] Hanung Adi Nugroho, Md. Dendi Maysanjaya, and Noor Akhmad Setiawan A Comparison of Classification Techniques for Thyroid Disease Diagnosis 2015 International Conference on Applications of Intelligent Technology