(IJRST) 2022, Vol. No. 12, Issue No. IV, Oct-Dec **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

Detection of Thyroid using Different Machine Learning Approach[1]

Bhavay Khatri

Heera Lal Public School, Delhi

DOI:10.37648/ijrst.v12i04.003

Received: 19 August 2022; Accepted: 28 October 2022; Published: 17 November 2022

ABSTRACT

The purpose of this paper is to identify the diagnosis of thyroid disease and to classify the two possible types of thyroid disease (hypothyroidism and hyperthyroidism). Numerous machine-learning algorithms are currently being used to identify thyroid disease. Nonetheless, our goal is to implement the machine learning algorithm, which will allow for faster and more accurate diagnosis of thyroid disease and type. The project is being implemented in Python, and the platform from which the dataset was derived is Kaggle. The Kaggle dataset was trained using a variety of machine-learning techniques. In addition, we have attempted to reduce the number of disease detection parameters.

INTRODUCTION

According to statistics, thyroid disease is prevalent in today's society. One adult has thyroid disease every ten. This thyroid condition affects over forty million people, according to estimates. The conventional method is time-consuming for diagnosing thyroid disease. Due to the less precise nature of using a traditional process, it has been observed that in some instances, a traditional procedure can result in incorrect results predictions. This system can accurately predict the outcome by employing the Random Forest machine learning algorithm, which provides greater precision than other algorithms for the same parameters.

The traditional method used by doctors to diagnose thyroid disease is time-consuming and laborious, necessitating effort and patience. Even traditional methods can sometimes produce false results. Using machine learning techniques, this procedure can be made simple and precise. By taking a few parameters, this system will automate the diagnosis of thyroid disease and provide a quicker, more accurate result.

The patient's health suffers due to the traditional diagnosis process's lack of precision and occasionally inaccurate predictions. This system can improve this and make doctors' jobs easier. It will help them diagnose early so that more treatment can be done. This procedure will be made much simpler and quicker by this system. Additionally, it will be more effective and accurate.

This study aims to specify the kind of thyroid disease that will determine the outcome, simplify the time-consuming prediction process for doctors, and increase prediction accuracy. Anyone who becomes excessively weak or vice versa will be able to receive the diagnosis at any time because the system will be available as an app or a website. In contrast to the traditional offline procedure, this service will be provided online. Since the system won't cost anything, the thyroid diagnosis won't cost anything. Based on the diagnosis, the user can see a doctor for additional medication and treatment.

SUGGESTED SYSTEM

To prepare the dataset for a variety of machine learning algorithms, including Logistic Regression, Random Forest, Gradient Boosting, and others, we use a dataset on thyroid disease from the Kaggle website for this method.

Several algorithms can be used with the dataset; we choose the best one and get a more accurate result in a shorter time.

[1] *How to cite the article:* Khatri B., Detection of Thyroid using Different Machine Learning Approach, IJRST, Oct-Dec 2022, Vol 12, Issue 4, 11-13, DOI: http://doi.org/10.37648/ijrst.v12i04.003

(IJRST) 2022, Vol. No. 12, Issue No. IV, Oct-Dec **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

We finally select the most accurate machine learning algorithm utilized in our model after carrying out the various algorithms and evaluating their outcomes.

A. The most recent version of a Garvan Institute archive of thyroid diagnoses, which includes 9292 records from 1984 to the beginning of 1987, can be found in this directory. 1) Diagnoses each record (record identification) using 29 attribute values.

B. Pre-processing During this step, we prepare the row dataset so that it can be used to construct and train machine learning models.

C. Feature Selection In this step, the row dataset's key features (columns) are chosen. In this study, we use 17 features, 15 of which serve as the model's input and 2 as output for analyzing the result.

D. Dividing the Dataset Now, we divide the dataset into training and testing data in proportions of 70% and 30%, respectively. The dataset is now prepared for use with various machine-learning algorithms.

E. The Gradient Boosting Algorithm The prediction speed and accuracy of the gradient boosting algorithm stand out, particularly when working with large and complex datasets. Errors are a big part of any machine-learning algorithm, we already know. The most common kinds of error are bias error and variance error. The gradient boost algorithm aids in reducing the model's bias error. This algorithm has delivered the best results when it comes to machine learning solutions for businesses and Kaggle competitions.

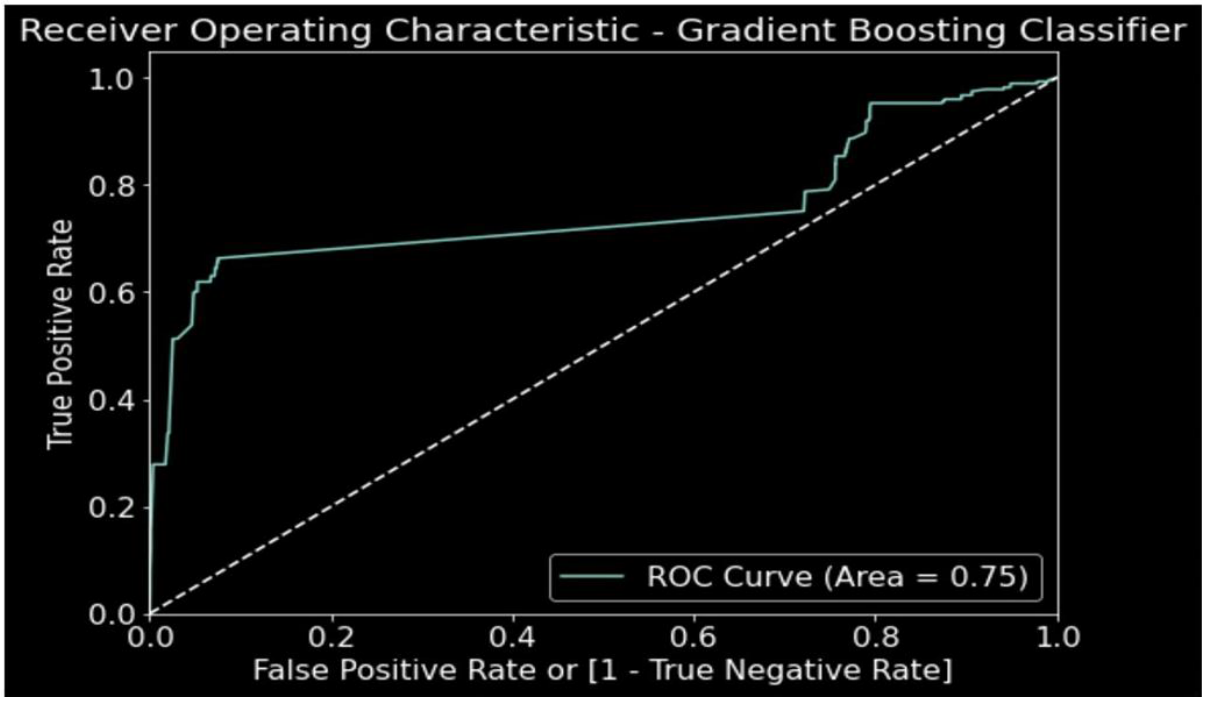

Fig 1. Gradient Boosting Algorithm

F. Predictive Model Outcome In this step, we evaluate the various algorithms and select the most precise ones for the final implementation.

RESULT ANALYSIS

After comparing and analysing logistic regression, Random Forest Algorithm Tuned 1, Tuned 2, Gradient Boosting Algorithms, and Random Forest Algorithm Tuned 1, Tuned 2. Gradient Boosting Algorithms, it was discovered that the Tuned 1 Random Forest Algorithm achieved an accuracy of 89.9 per cent across all three parts of the experiment. On the other hand, logistic regression had the second-best accuracy, with 78 Algo Random Forest.

The benefits and robustness of the new dataset, which enables physicians to obtain more precise and accurate results in a shorter time, are evident when the results are analysed.

Accuracy	AUC
Logistic Regression	0.874378 0.602132
Random Forest	0.894676 0.420194

Random Forest Tuned 1	0.893527	0.591608
Random Forest Tuned 2	0.888548	0.545813
Gradient Boosting	0.411720	0.524825

Fig 2. Result Analysis of algorithms

CONCLUSION:

We learned a lot from various sources to complete this project. To complete this project, we used a variety of methods. Using machine learning algorithms makes it easier to identify and detect objects with high precision. Due to the overlapping of their symptoms with those of other conditions, thyroid disease is difficult to diagnose; however, our model, described in this study, provides a more effective and accurate result based on a few inputs.

REFERENCES

1. Ankita Tyagi and Ritika Mehra. (2018). "Interactive Thyroid Disease Prediction System using Machine Leaming Techniques" published on ResearchGate.
2. Yong Feng Wang, (2020). "Comparison Study of Radiomics and Deep-Leaming Based Methods for Thyroid Nodules Classification using Ultrasound Images" published on IEEE Access.
3. Sunila Godara, (2018). "Prediction of Thyroid Disease Using Machine Leaming Techniques" published on IJEE.
4. Hitesh Garg, (2013). "Segmentation of Thyroid Gland in Ultrasound image using Neural Network" published on IEEE.
5. Jahantigh, F.F.: Kidney diseases diagnosis by using fuzzy logic. In: 2015 International Conference on Industrial Engineering and Operations Management.