International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# Thyroid Disease Detection Using ML

**Gayatri Patil[1], Sonam Shinde[2], Swaliya Mulla[3], Priyanka Kadam[4] ,**

*[1]Student, [2]Student, [3]Student, [4]Student,*
*Computer Science And Engineering Department,*
*Padmabhooshan Vasantraodada Patil Institute of Technology, Budhgaon, Maharashtra*

---

**Abstract**- The purpose of this paper is to identify the diagnosis of thyroid disease and to classify types of thyroid disease. There are two types of thyroid diseases namely Hyperthyroid and Hypothyroid. Hypothyroid is a common variation of thyroid disease. Various Machine Learning Algorithms are being used to identify thyroid disease. We have mainly used classification algorithms named Decision Tree (DT), Random Forest (RF) and , KNN - K-nearest neighbours .By observing the results, we could predict that our Trained Dataset provides approximately more than 90% accuracy for Random Forest classification algorithm. We have attempted to reduce the number of disease detection parameters.

## 1. INTRODUCTION

Thyroid Detection System using Machine Learning is the system that is used to detect the Thyroid diseases from the Parameters which are given by the patients or any user. The system processes the parameters provided by the user as input and gives the output as the person has Thyroid or not.

The traditional method used by doctors to diagnose thyroid disease is time-consuming and it takes more efforts. 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.

With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care. By using Classification Algorithm we can detect the Thyroid. The types of Thyroid disorders are Primary Hypothyroid, Secondary Hypothyroid, Compensated Hypothyroid. When this disorder occurs in the body, they release certain type of hormones into the body which imbalances the body's metabolism. Thyroid related Blood test is used to detect this disease but it is often blurred and noise will be present. Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patients getting this disease. Machine Learning plays a very deciding role in the disease prediction.

The Thyroid gland is a vascular gland and one of the most important organs of a human body. This gland secretes two hormones which help in controlling the metabolism of the body. The two types of Thyroid disorders are Hyperthyroidism and Hypothyroidism. When this disorder occurs in the body, they release certain type of hormones into the body which imbalances the body's metabolism. Thyroid related Blood test is used to detect this disease but it is often blurred and noise will be present. Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patients getting this disease. Machine Learning plays a very deciding role in the disease prediction. Machine Learning algorithms, SVM - support vector machine, decision tree, logistic regression, KNN - K-nearest neighbors, ANN Artificial Neural Network are used to predict the patient's risk of getting thyroid disease.

## 2. LITERATURE REVIEW

Authors in [1] conducted a study aimed a smart and precise way to predict thyroid disease.The classification was carried out using Logistic regression Algorithm to train dataset and to predict thyroid disease with more accuracy. In this machine is trained to detect whether the person normal, hypothyroidism based on the user's input. when user enters data in web app the data will be processed in backend (model) and the result will be displayed on the screen.

Authors in [2] conducted a study aimed at diagnosing hyperthyroidism and hypothyroidism, the two most common thyroid disorders. The classification was carried out using two techniques, multinomial logistic regression models and neural networks. The research was conducted on 310 patients, and even in this case, the models took as input demographics and hormonal parameters. The results showed better performance of the neural network model (with an average accuracy of 96.3%) than multinomial logistic regression (with a mean accuracy of 91.4%) in all cases.

Authors in [3,4], also focus on machine learning techniques such as Support Vector Machine (SVM), Multiple Linear Regression, Nave Bayes, Decision Trees, to perform a comparative diagnosis of thyroid disease. Their results (precision is equal to 99.23%) show that decision trees have the best performance and can be used successfully as an aid in the detection of thyroid disease.

On the other hand, in the study [5], the authors used neural networks (MLP, PNN, GRNN, FTDNN, CFNN) to diagnose types of thyroid disease. More specifically they conducted a study on 244 subjects suffering from different pathologies to investigate the state of their thyroid, taking into account some hormonal parameters and the patient's age. The results of this research show that the neural network offers very precise answers, classifying correct thyroid pathologies based on hormonal parameters.

In the first group, we find the work of Izdihar and Bozkus [6] who used a dataset from the UCI repository to classify thyroid disease using the decision tree algorithm. In particular, they have developed a machine learning tool for the diagnosis of thyroid diseases, called MLTDD capable of making an intelligent forecast of thyroid gland diseases. This study shows an overall accuracy of 98.7% and 99.8% for testing.

## 3. METHODOLOGY

Developing a methodology for thyroid disease detection using machine learning involves a systematic approach to designing and implementing the project. Here's a proposed methodology to guide your work:

- **Problem Definition:**
  Clearly define the problem you want to address. Specify the scope, such as the types of thyroid diseases you aim to detect and the data sources you'll use.

- **Data Collection:**
  Collect a comprehensive dataset of patient records, which includes relevant information for thyroid disease diagnosis. Ensure you have both features (input data) and the target variable (diagnosis).

- **Data Preprocessing:**
  - Clean and preprocess the data:
  - Handle missing values, outliers, and errors in the dataset.
  - Normalize or standardize numerical features.
  - Encode categorical variables.
  - Ensure data quality and consistency.
- **Feature Engineering/Selection:**
  Identify the most informative features for thyroid disease detection. You may also create new features, like age groups or symptom severity scores, to improve the model's performance.

- **Data Splitting:**
  Divide the dataset into:

  Training set: Used to train machine learning models.

  Validation set: Used to fine-tune hyperparameters and prevent overfitting.

  Test set: Reserved for final model evaluation.

- **Model Selection:**
  Choose appropriate machine learning algorithms for binary or multi-class classification (e.g., logistic regression, decision trees, random forests, support vector machines, or neural networks).

- **Model Training:**
  Train the selected model(s) using the training dataset.

  Hyperparameter Tuning:

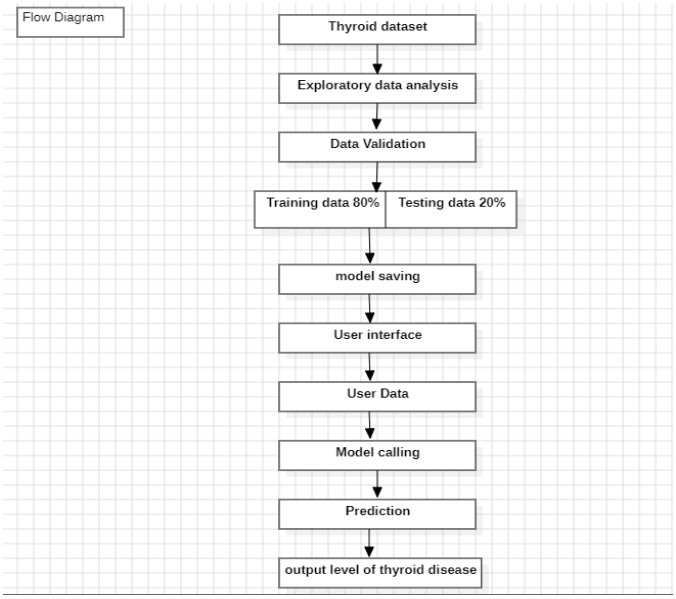
  Optimize the model's hyperparameters using techniques like grid search or random search.

- **Model Evaluation:**
  Evaluate the model's performance on the validation and test datasets using relevant evaluation metrics (e.g., accuracy, precision, recall, F1 score, AUC-ROC). This step helps you select the best-performing model.

- **Interpretability:**
  If applicable, assess the model's interpretability by analysing feature importance and providing explanations for predictions.

- **Model Validation:**
  Validate the model's performance on external datasets (if available) to ensure its generalizability.

- **User Interface (UI):**
  Develop a user-friendly interface for healthcare professionals to input patient data and receive model predictions.




**Flow Diagram**

## 4. CONCLUSION

Thyroid Disease Detection System using Machine Learning is a project idea that aims a smart and precise way to predict thyroid disease. We have made use of K-Nearest-Neighbors, Decision Tree, Random Forest algorithms to train our dataset and to predict thyroid disease with more accuracy. Here the machine is trained to detect whether the person no thyroid, Primary Hypothyroid, Secondary Hypothyroid, or Compensated Hypothyroid based on the user's input. So when user enters data in web app the data will be processed in backend (model) and the result will be displayed on the screen. Our objective was to give society an efficient and precise way of machine learning which can be used in applications aiming to perform disease detection.

## 5. FUTURE SCOPE

Dataset is unbalanced, as the three classes are not equally represented. This phase mitigates this problem by oversampling the examples in the minority class, randomly. On the other hand, the main limitation of this study concerns the quality of the dataset.

In the future, to better generalize our findings it is necessary to further expand the set of data and attributes considered. With more data the training process is likely to produce more effective classifiers also allowing a more reliable estimate of the exhibited performance. Finally, another aspect that could be investigated concerns the presence of any secondary thyroid disease linked to the patient, to understand if there is a particular additional thyroid disease that can affect hypothyroidism. In fact, it often happens that patients are suffering from more than one thyroid disease at the same time.

Further development can be do by using image processing of ultrasonic scanning of thyroid images to predict thyroid nodules and cancer, which cannot be recognized in blood test report. By combining both the results, thyroid disease prediction can cover all thyroid related diseases

## 6. REFERENCES

1] Mr.M.Ram Chandra, Nunchu Ayushi Yadav,Gadey Yeshwanth Reddy, S.Sai Abhinav, “THYROID DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES”,Vol 13, Issue 06, JUNE/ 2022.
2] Shiva Borzouei, Hossein Mahjub, NegarAsaad Sajadi, and Maryam Farhadian. Diagnosing thyroid disorders: Comparison of logistic regression and neural network models. Journal of Family Medicine and Primary Care, 9:1470, 06 2020.
3] Ankita Tyagi and Ritika Mehra. Interactive thyroid disease prediction system using machine learning technique, 5th IEEE International Conference on Parallel, Distributed and Grid Computing Dec 2019.
4] S. Razia, P. SwathiPrathyusha, N. Krishna, and N. Sumana. A comparative study of machine learning algorithms on thyroid disease prediction. International journal of engineering and technology, 7:315, 2018.
5] Mahdiyar O. Obeidavi M. R, Rafiee A. Diagnosing thyroid disease by neural networks. 2017.
6] Izdihar Al-muwaffaq and Zeki Bozkus. Mltdd : Use of machine learning techniques for diagnosis of thyroid gland disorder. 2016.