

Thyroid Disease Prediction Using Machine Learning

Sowkya Innamuri
Computer Science Engineering
Narasaraopeta Engineering College
Narasaraopet, India
innamurisowkya@gmail.com

CH. Mounika Lakshmi
Computer Science Engineering
Narasaraopeta Engineering College
Narasaraopet, India
mounich234@gmail.com

D. Neha Sree
Computer Science Engineering
Narasaraopeta Engineering College
Narasaraopet, India
devathineha@gmail.com

Abstract - In today's world Machine Learning plays an important role in predicting diseases by taking relevant information and predicting result whether the patient having disease or not. The best example we have in our hands is thyroid disease, this specific disease needed to be test at early stages, so machine learning plays a crucial part in detecting it, we take patients details and perform some classification type algorithms and tell whether the person is having this thyroid or not. We performed many algorithms to get best accuracy and this will ensure us it performs well on any new data and give us good result.

Keywords - Different types of classification models like Random Forest Classification, KNN, Logistic Regression, Decision tree, Naïve Bayes.

I. INTRODUCTION

Thyroid diagnosis is not an easy task and it needs many procedures to be tested and many procedures involved. The most followed ways like taking blood samples and this is how they detect whether the person having thyroid or not [1].

So, there is a necessity for a model to predict disease at early stage. We are having good dataset and we can train it with different types of classification models and produce an accurate result.

There is research, which tells about thyroid had link with mental health disorders, like anxiety and depression.

Firstly, thyroid disease is a condition that affects the function of thyroid gland and it is a small butterfly-shaped gland located in the neck which produces hormones and disturbs metabolism, growth and development.

There are two types of thyroid disease called hypothyroidism and hyperthyroidism. It needs regular check-ups and should go through thyroid functioning have to be tested and then we can decide how many days will that take to recover.

Our model will have both train and test data. We ask patients to fill some details and values they get from their samples. Then we can decide easily from those data and we can skip further clinical examinations.

Classification models are really good at predicting things and in decision-making. They also solve May real world problems.

II. ABOUT THYROID

A. More about thyroid and its side-effects

This produces different types of hormones, in that two main hormones are thyroxine (T4) and triiodothyronine (T3). T4 is responsible in managing body metabolism, growth and development. It might affect different functioning organisms in our body.

As T3 is responsible for brain development and functioning. Another hormone is thyroid-stimulating hormone (TSH), it is main to produce thyroid hormones and produced by pituitary gland.

Based on these levels, we can decide patient is having underactive thyroid gland or overactive thyroid gland. Imbalance in thyroid hormone levels can cause a wide range of symptoms and health problems.

There are so many features like psych, lithium and goitre. Based on these features also we can talk more about disease. We need to ensure if the patient is female and need to ask if she is pregnant, we need to test them to avoid little baby thyroid affects.

These symptoms will become worse, if they are not treated at early stages. So, there is a need for proper prediction model which helps in treating patient's disease at early stages.

III. LITERATURE SURVEY

They analyzed correlation between all numerical features in the dataset such as Age, Gender, T3, TT4, T4U, TSH and FTI. They trained data under different types of classification models like KNN, Decision Tree, and Random Forest Classifier.

They proposed a dataset which contains 30 features that will help in predicting thyroid disease. The total rows were near to 4000 [2].

To know more about disease prediction, they observed different types of diseases like diabetes, heart disease and Breast cancer, how they worked on different models.

They learned about all types of classification models and studied about them. Applied all those models on both train

and test dataset. Finally checked accuracies and compared them to bring out best model.

They performed some preprocessing stages by knowing mean values of T3, TT4 and TSH. [3] Later they checked feature selection to eliminate unused columns. They performed prediction on thyroid disease data using Logistic Regression and found good accuracy.

Along with that they concluded Decision Tree Classifier [10] as good method when the number of classes increases in the thyroid model.

They have taken dataset from Kaggle and trained it well to perform well in giving good results and with good accuracy.

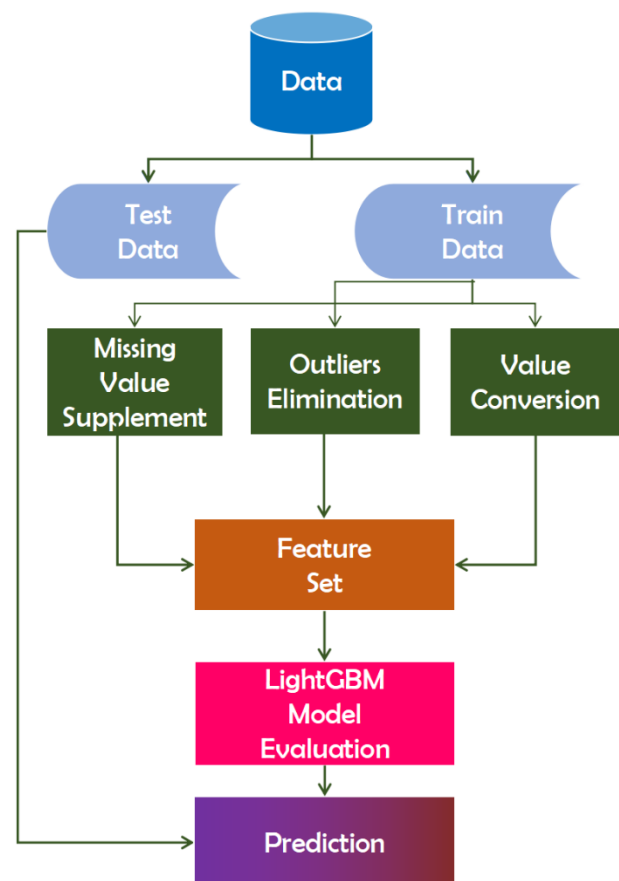
IV.DATASET DESCRIPTION

I have chosen my dataset from Kaggle website [13]. The dataset contains patient information like their age, gender, blood samples details. We store all those details in database. From the dataset we take only few attributes. Those attributes which is more responsible for causing thyroid disease and remaining are neglected simply removed for better accuracy. Mostly my dataset features are Boolean values like t for True and f for False and M for Male and F for female. Most features are object and remaining are numeric.

When I applied model to give me best features that affect patient record to give whether the person is having thyroid or not. They are Age, Gender, T3, T3 measured, Referral Source and FTI.

SLNO.	Attribute Name	Value Type
1	Age	Continuous
2	Gender	M, F
3	T3	Continuous
4	T3 Measured	F, T
5	Referral Source	SHVC, Other, SVI, STMW
6	FTI	Continuous

This table will make you understand about data that contain in dataset.



This diagram depicts about the process we done throughout.

B. About Models in Machine Learning

Classification model is a type of machine learning model that is used to classify data into different classes and it is a supervised learning and it will produce unseen to known category. [11]

There are different types of classification models like Binary Classification which predict outcomes as yes or no, true or false, positive or negative.

Another one is multi-label classification this predicts multiple categories and labels for the input given by user.

Another two are Imbalanced and Hierarchical used for complex outcomes [5]. As our prediction uses Binary Classification model.

In Classification model, there are different models like Logistic Regression it is a statically model that draw relationship between input and output classes. Decision Tree Classifier is another best model and it is tree-based structure and produce decision-based outcomes.

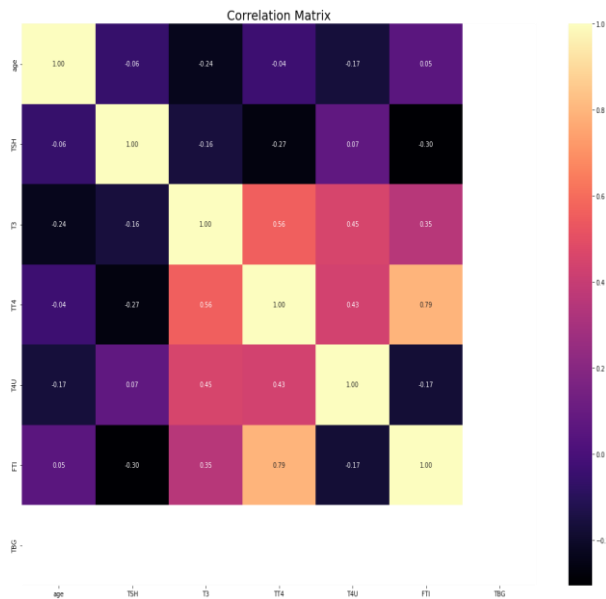
Random Forest this model combines multiple decision trees and it produce best accuracy. Naïve Bayes model is a probabilistic that uses theorem (Bayes') to predict particular data point for given class.

V. PROPOSED WORK

We decided to develop a model that will predict thyroid disease in patients easily by filling the needed values.

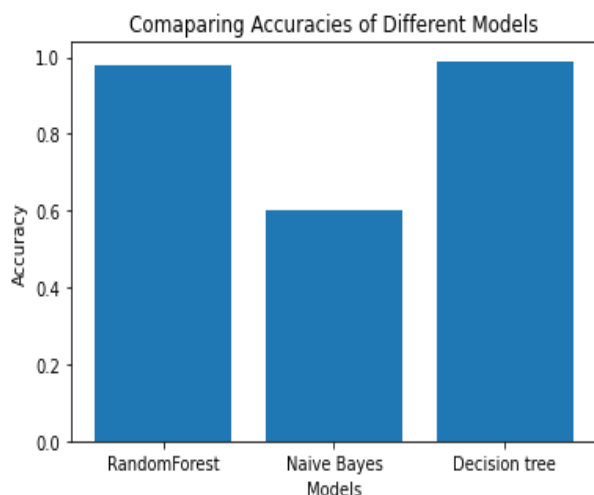
We took our data from Kaggle website. That dataset contains 30 features and we applied different types of preprocessing techniques to clean data and remove outliers.

We draw correlation matrix to understand better about relationship between our features.



After outlier detection, we split data into train and test. Next, we check balancing data using SMOTE type over sampling or under sampling, using these we balance data.

On the both trained and test data, applied different types of classification models and found out which model produce the best accuracy.



From this we can find out perfect model and ready out to find patients thyroid disease.

VI. CONCLUSION

Thyroid disease prediction involves various steps like image processing, blood test and analyzing the samples. All this can be done well in machine learning. All machine learning models ensure to give promising results.

Based on different datasets everything varies and clean, quality dataset produce good accuracy and good results.

REFERENCES

- [1] Prediction of thyroid Disease Using Data Mining Techniques” 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019
- [2] “Interactive Thyroid Disease Prediction System Using Machine Learning Technique” 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), 20-22 Dec, 2018, Solan, India
- [3] A K and Anil Antony “An Intelligent System for Thyroid Disease Classification and Diagnosis” Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2
- [4] “A Empirical study on Disease Diagnosis using Data Mining Techniques.” Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2
- [5] S “Classification of Thyroid Disease using Data Mining Techniques” International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 13881-13890
- [6] “A Comparison of Classification Methods on Diagnosis of Thyroid Diseases” 2015 International Seminar on Intelligent Technology and Its Applications
- [7] “Thyroid Data Prediction using Data Classification Algorithm” IJRST –International Journal for Innovative Research in Science & Technology| Volume 4 | Issue 2 | July 2017
- [8] “A Study of Data Mining Techniques to Detect Thyroid Disease” International Journal of Innovative Research in Science, Engineering and Technology (Vol. 6, Special Issue 11, September 2017)
- [9] Diagnosis of Thyroid Disease Using Data Mining Techniques: A Study” International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov - 2016
- [10] Prediction of Thyroid Disease Using Machine learning Techniques” International Journal of Electronics Engineering (ISSN: 0973-7383) Volume 10 • Issue 2 pp. 787-793 June 2018
- [11] Transforming clinical data into actionable prognosis models: machine learning Framework and field deployable app to predict outcome of Ebola Patients, PLoS Negl. Trop. Dis. 10 (3) (2016) e0004549.
- [12] <https://machinelearningmastery.com/types-of-classification-in-machine-learning>
- [13] <https://www.kaggle.com/kumar012/hypothyroid>