IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 3, March 2021

DOI: 10.17148/IARJSET.2021.8318

Prediction of Thyroid Disease Using Machine Learning Techniques

Srinidhi C R[1], Dr. Prabha R[2]

M.Tech, Department of CSE, Dr. A.I.T, Bengaluru, India[1]

Associate Professor, Department of CSE, Dr. A.I.T, Bengaluru, India[2]

Abstract: The most important technique in Machine Learning is Classification. There are Various diseases that can be diagnosed early and can be treated at early Stage. Example, Thyroid diseases. Random Forest Technique is used to identify the type of thyroid using Anaconda and Python Programming Language to implement the Random Forest Algorithm.

Keywords: Machine learning techniques, Python, Random Forest Algorithm, Disease Prediction.

1. INTRODUCTION

In healthcare and medical science, the applications based on Data mining are very beneficial and important. The large amount of data gathered from health care organization has no organizational value unless transformed into most useful information and knowledge, which could be helpful in cost controlling, increasing the profits, and high-quality maintenance of patient healthcare. Classification Algorithms is one of the most important applications in the field of data mining, which can be useful for decision making in many real-world problems.

Decrease production in the thyroid hormones causes Hypothyroidism. The medical term hypo means deficient or less. The causes for hypothyroidism are inflammation and thyroid gland damage. The Symptoms includes obesity, low heart rate, increase in cold sensitiveness, neck swelling, dry skin, hands numbness, hair problem, heavy menstrual periods and digestive problems. And these Symptoms may worsen over period if not treated.

2. LITERATURE SURVEY

Bibi Amina Begum has proposed a method for detecting thyroid disease using Data Mining Techniques. Data Mining plays a crucial role in the medical field for making decisions and diagnosis of disease. Classification techniques are applied on the dataset to predict thyroid gland disorder. Linear regression is used to predict which hormone will affect male and female. Disease diagnosis helps to give better treatment to the patients. Algorithms such as K-Nearest Neighbour, Naïve Bayesian, SVM and ID3 are used in this study. [1]

Shaik Raziya have compared various machine learning algorithms for thyroid prediction. SVM, Multiple Linear Regression, Naïve Bayesian and Decision Tree are the machine learning algorithms used for the comparison. The data sets were collected from the UCI machine learning repository. All of the above mentioned algorithms were executed and the resulting accuracy was compared. The Decision Tree algorithm shows greater accuracy followed by SVM and Multiple Linear Regression. Whereas Naive Bayes stands at the last with least accuracy. [2]

The diagnosis of thyroid gland disorders is one of the application for important classification problem. This study majorly focuses on thyroid gland medical diseases caused by underactive or overactive thyroid glands. The dataset used for the study was taken from UCI repository. Classification of this thyroid disease dataset was a considerable task using decision tree algorithm. The overall prediction accuracy is 100% for training and in range between 98.7% and 99.8% for testing.[5]

Further another system was proposed with KNN and SVM for thyroid disease diagnosis. This system had two phases. In first phase, it is checked whether dataset contains any missing value if it has missing value then KNN imputation method is used to fill the missing values in input. And after that the dataset is sent into SVM to detect thyroid disease. In second phase, if dataset contains no missing value it is sent directly into SVM without sending data into KNN algorithm. This system is one which can reduce thyroid diagnosis cost for patients.[4]

3. DESIGN AND METHODOLOGY

A system architecture diagram would be used to show the relationship between different components. Usually they are created for systems which include hardware and software and these are represented in the diagram to show the interaction between them.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 3, March 2021

DOI: 10.17148/IARJSET.2021.8318

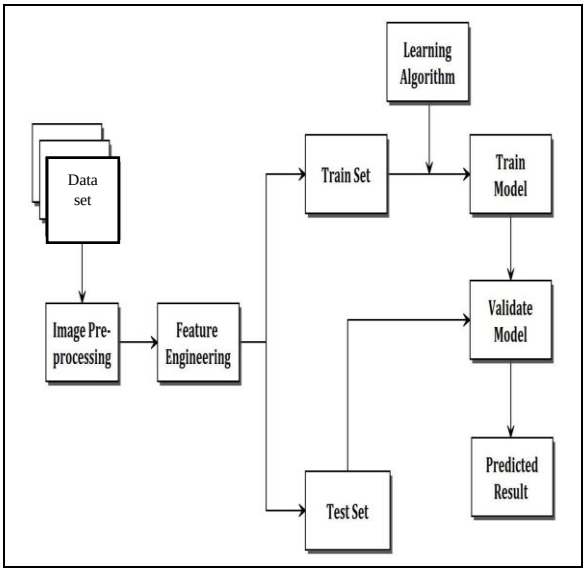

Figure 1: System Design

a) Random Forest Algorithm

1. Select k data sets randomly from given training data.
2. For selected data sets build a decision trees for these subsets which is chosen.
3. Choose number N for decision trees that you want to build.
4. Repeat steps 1 and 2.
5. For new data sets, calculate each decision tree prediction and add new data set to class which has majority.

4. RESULTS

The Machine Learning Technology has become very easy to predict relation and patterns of various data's. This paper mainly involves in predicting the type of thyroid diseases. The model is built using training data set which have the data cleaning and data transformation. This model has 98% Accuracy in analysis the data set with the help of data visualization.

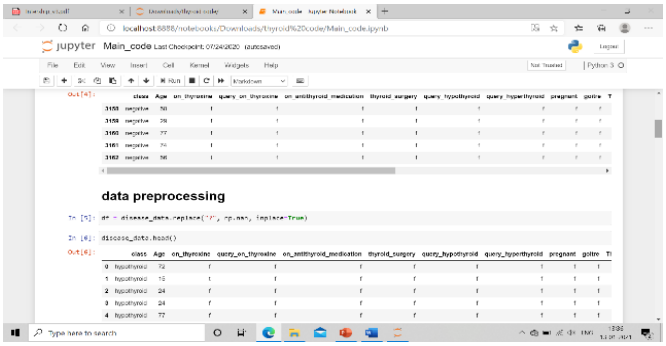

Figure 2: Data of the Model

In the above diagram we have different parameters like age, Class, on_thyroxine, thyroid surgery, Pregnant, Goitre and outcome for prediction of thyroid for different Scenarios.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 3, March 2021

DOI: 10.17148/IARJSET.2021.8318

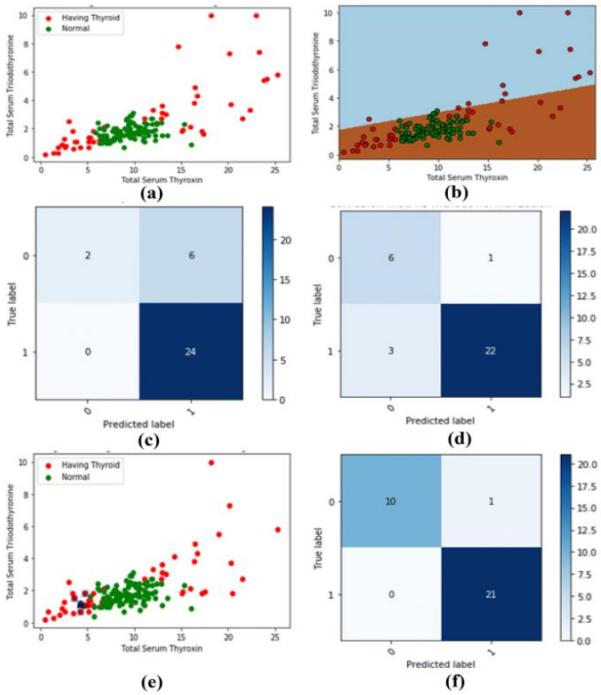

Figure 3: Graph

A dot chart or dot plot is a statistical chart consisting of data points plotted on a fairly simple scale. The numeric data is normally distributed to the left or right and it can also consider in the table. Age, Class, thyroid surgery, Pregnant and some more features.

5. CONCLUSION

In this work, we have used machine learning algorithms to predict thyroid disease. In this system, we have used data mining classification algorithms and regression algorithms. So, both regression and classification are combined to produce accurate diagnosis results. The logistic regression is more efficient and accurate compared to other classification techniques. But other recent techniques can be combined in future to give still more accurate results of thyroid diagnosis.

REFERENCES

[1] Bibi Amina Begum and Dr.Parkavi A "Prediction of thyroid disease using data mining techniques",5th International Conference on Advanced Computing & Communication Systems (ICACCS) 2019. (references)

[2] Shaik Razia, A Comparative study of machine learning algorithms on thyroid disease prediction, International Journal of Engineering & Technology, 7 (2.8) (2018) 315-319

[3] Thyroid: https://en.wikipedia.org/wiki/Thyroid

[4] Aswathi A K and Anil Antony "An Intelligent System for thyroid disease classification and diagnosis" 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2. (references)

[5] Shaik Razia, Swathi Prathyusha, Vamsi Krishna, Sathya Sumana "A Comparative study of Machine Learning Algorithm on thyroid disease prediction", International Journal of Engineering & Technology, 7 (2.8) (2018) 315-319. (references)

[6] Haria Viral, More Suraksha, Patel Bijal and Patil Harshali "Thyroid Prediction System using Machine Learning Techniques", International Journal of Scientific Research and Reviews (IJSRR) 2018, 7(4), 674-681. (references)