



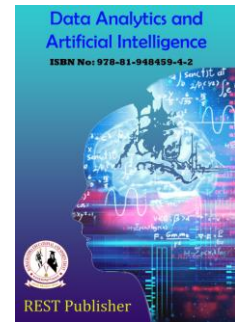
Data Analytics and Artificial Intelligence

Vol: 3(2), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: <http://restpublisher.com/book-series/daai/>

DOI: <https://doi.org/10.46632/daai/3/2/29>



Thyroid Detection using Machine learning

SavitaAdhav, AmitGoge, VipulJadhao, AbhishekMarkad, SurajJadhav

Information Technology JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India.

*Corresponding Author Email: iamaabhi82@gmail.com

Abstract. *Thyroid is a gland which is present in front of the neck, wrapped around the windpipe (trachea). Thyroid shape is like Butterfly that's smaller in the middle and 2 side wings which are around the throat. As Thyroid is a gland it is important in our body when thyroid produces too much thyroid hormones then that condition is called hyperthyroidism and if thyroid gland produces less thyroid hormones then that condition is called hypothyroidism. Machine learning is one of the most important tools to classify the diseases nowadays whether a person has a disease or not like Cancer detection, kidney disease detection or Diabetes etc. Our project is to predict whether a person has Thyroid disease or not. Whether he has hypothyroidism or hyperthyroidism, or he is not suffering from thyroid disease.*

Keywords: *Thyroid detection, machine learning, Logistic regression, SVM, Naves Bayes, Decision Tree, Linear regression, ANN, K-means.*

1. INTRODUCTION

The evolvement of technology in biology is used in healthcare industry. Due to the super computers and high computational power, it is easy to detect and solve problems which were very hard to detect in previous era. As medical information is rich in datasets but very few algorithms can give more accuracy. As detecting a disease directly impact person health it is very important to use algorithms or systems which have high accuracy rate otherwise person would suffer from different disease due to medical treatment. As in recent years machine learning has shown good results as healthcare has vast amount of data to analysis of dataset and removing values which are not necessary for disease detection. As it is also important what kind of data to remove it and what kind of data should be keep in dataset. If the data is unnecessary, it can hamper accuracy of dataset.

Thyroid disease can be classified mainly into 2 categories that is Hyperthyroidism and Hypothyroidism.

Hyperthyroidism: -Hyperthyroidism is caused by an increase in thyroid hormone levels. Hyperthyroidism is a disorder in which the thyroid gland releases so many thyroid hormones. Neck enlargement, nervousness, menstrual cycles shortening, dry skin, elevated temperature sensitivity, hair thinning, weight loss, increased heart rate, high blood pressure, heavy sweating, irregular stomach movements, and hands shaking are some of the signs.

Hypothyroidism: -Hypo means deficient or less in medical terms. Inflammation and thyroid gland injury are the two primary causes of hypothyroidism. Hypothyroidism is a condition in which the thyroid gland is underactive Hypothyroidism is caused by a decline in thyroid hormone production. Neck swelling, dry skin, hand numbness, hair issues, heavy menstrual cycles, obesity, low heart rate, increased temperature sensitivity, and intestinal problems are some of the symptoms.

2. LITERATURE SURVEY

Thyroid Disease Classification Using Machine Learning Algorithm [1] Khalid Salman and Emrullahsonac The first model all the characteristics consisting of 16 inputs and one output were taken, and the result of the accuracy of the random forest algorithm was 98.93, which is the highest accuracy among the other algorithms. In the second embodiment, the following characteristics were omitted based on a previous study. The removed attributes were 1- query_thyroxine 2- query_hypothyroid 3- query_hyperthyroid. Here we have included the increased accuracy of some algorithms, as well as the retention of the accuracy of others. It was observed that the accuracy of Naive Bayes algorithm increased the accuracy by 90.67. The highest precision of the MLP algorithm was 96.4 accuracy. Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques RajasekharChaganti, FurqanRustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez and Imran Ashraf.[2] They have tried to reduce input number of features and they have reduced input from 32 to 16 and they observe that after 16 input accuracy of the model

is almost same after giving more than 16 inputs. Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study [3]SaimaSharleen Islam¹, Md. Samiul Haque¹, M. SaefUllah Miah², Talha Bin Sarwar² and RamdhanNugraha. This study presents an experimental study for different machine learning algorithms: neural networking classifiers, tree-based classifiers, statistical classifiers to predict thyroid risk. Among all the algorithms, the ANN classifier outperforms others with an accuracy of 0.9587. The CatBoost and XGBoost classifiers come second and third with the accuracy of 0.9538 and 0.9533, respectively. Increasing the Prediction Accuracy for Thyroid Disease [4]A Step Towards Better Health for Society Ritesh, Jha.Vandana Bhattacharjee · Abhijit Mustaf. Proposed feature reduction technique by this researcher they got a accuracy of 98.70 and they have used Neural network using decision tree for feature reduction. Thyroid Disease Treatment prediction with machine learning approaches [5]LerinaAversanoa,Mario Luca Bernardia, Marta Cimitileb, Martina Iammarinoa, Paolo EmidioMacchiac, Immacolata Cristina Nettorec, Chiara Verdon. The results obtained demonstrate a good performance of the EXTC model, compared to the other approaches used, with an F-score equal to 84%. On the other hand, the main limitation of this study concerns the quality of the dataset, as, as already mentioned, this was constructed starting from real data belonging to patients being treated at a hospital.

3.ALGORITHMS

In The problem of domain is classification of thyroid patient.so our algorithms for that are varies classification algorithms. K-means, random forest, decision tree, svm and some deep learning algorithms have given best results to classify thyroid patient. K-means is a clustering algorithm in which each observation is partitioned into a single cluster with no information about how confident we are in this assignment. When a new patient's data is provided to model it check for nearest cluster and then accordingly classify the patient. Decision Tree- it in machine learning algorithm in which at each layer classification is done on the basis of data.in training phase, a tree is built which classify the new instance. Random Forest - A random forest consists of multiple random decision trees. Two types of randomness are built into the trees. First, at each tree node, a subset of features is randomly selected to generate the best split. Second, each tree is built on a random sample from the original data. it is generally used to increase the accuracy. Svm-support vector machine classifies the data using support vectors. it can also work for nonlinear data.

work flow

1. Creating Dataset
2. Data preprocessing
3. Classify dataset in train and test
4. Model creation using train data
5. Model evaluation using test data
6. Predication of new patient based on model

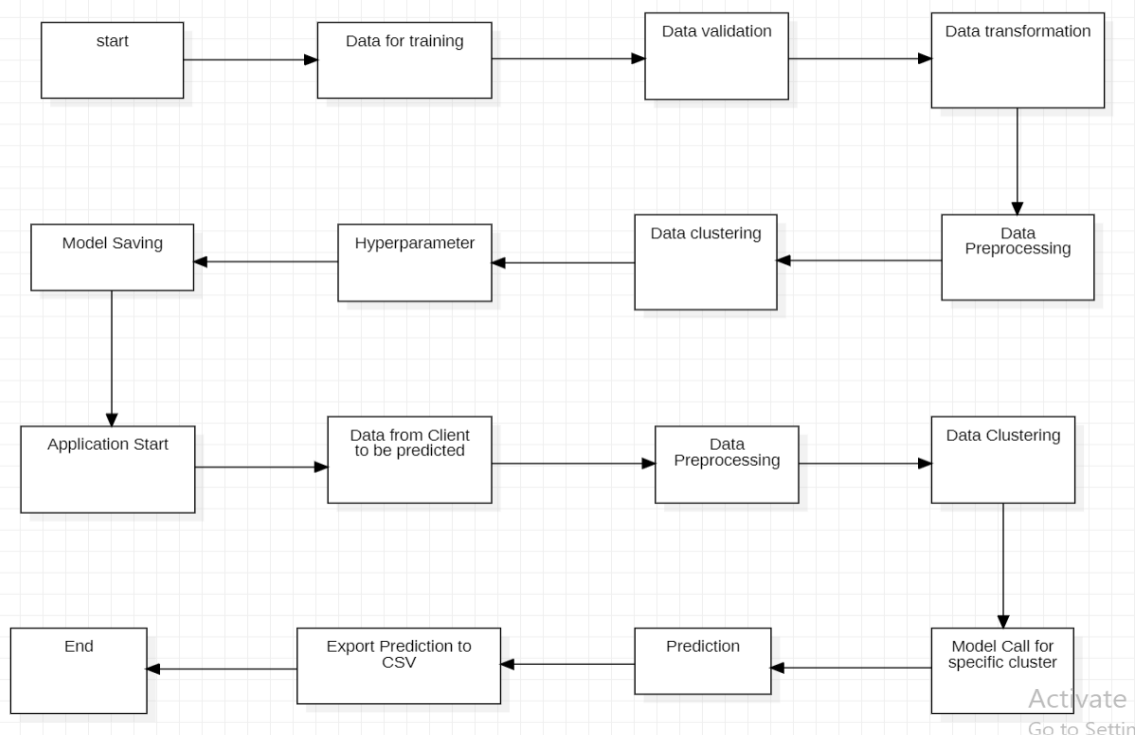


FIGURE 1. Architecture

Mathematical Model

Our System S take the input in from user, then process on it, in third step provide the output in form of whether the user have hyperthyroidism or hypothyroidism.

1. Taking input from user (I)
2. Processing on input (P)
3. Output (O)

$$S = \Sigma \{I, P, O\}$$

In processing, our machine learning Model actually works on input. In second step our model pre-process the input then gives prediction.

This how our system gives output.

System S –

$$S = \Sigma \{I, P, M, O\}$$

1. Taking input (I)
2. Pre-processing input data (P)
3. Model prediction (Classification) (M)
4. Output (O)

4. CONCLUSION

Thyroid disease cases are increasing in big numbers now days. Thyroid disease detection has emerged as an important medical problem and requires efficient automatic prediction models. Our study deals with the classification of thyroid disease between hyperthyroidism and hypothyroidism. This disease was classified using algorithms. Machine learning has showed us good results using several algorithms. We have made use of logistic regression algorithm to train our dataset and to predict thyroid disease with more accuracy. Here the machine is trained to detect whether the person normal, hyper-hypothyroidism 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.

REFERENCES

- [1]. Khalid Salman and Emrullahsonac (2021) Thyroid Disease Classification Using Machine Learning Algorithm
- [2]. RajasekharChaganti, FurqanRustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez and Imran Ashraf (2021) Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques.
- [3]. SaimaSharleen Islam¹, Md. Samiul Haque¹, M. SaefUllah Miah², Talha Bin Sarwar² and RamdhanNugraha (2021) Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study.
- [4]. Ritesh, Jha. Vandana Bhattacharjee · Abhijit Mustaf (2020) Increasing the Prediction Accuracy for Thyroid Disease: A Step Towards Better Health for Society.
- [5]. Mario Luca Bernardia, Marta Cimitileb, Martina Iammarinoa, Paolo EmidioMacchiac, Immacolata Cristina Nettorec, Chiara Verdon (2020) Thyroid Disease Treatment prediction with machine learning approaches LerinaAversanoa.