

Thyroid Detection

Himanshi Sharma¹, Swati Panchal², Ashmeet Kaur Duggal³

¹Executive Software Developer Confidential Delhi, India

²Scholar Trainee, Wipro, Delhi, India

³Assistant Professor, SGBIMIT Delhi, India

Correspondence should be addressed to Himanshi Sharma; himanshisharma99588@gmail.com

Copyright © 2023 Made Himanshi Sharma et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- The thyroid disease is a common term related to medical state that prevents the thyroid gland from creating hormones in the correct quantity. The thyroid particularly creates hormones which keep the body working normally. When the thyroid gland makes so many hormones, the body uses up almost whole energy very rapidly. This type of state is known as hyperthyroidism, in which it uses so much of energy very rapidly, will make a person very much tired, it can increase the heart rate-causes a person to lose so much of weight and make a person feel jittery. On the other hand, the thyroid gland can also make very little thyroid hormone. This type of state in a person is known as hypothyroidism, when a person has very little thyroid hormone in their body, they can feel fatigued or might gain weight and even not be able to bear cool temperatures. This type of thyroid disease can be caused by varied conditions. This disease can also be inherited down to families- a child can inherit it from their parents. Thyroid detection is to study and analyze thyroid disease outbreaks disease analysis and prediction it can help in revealing or predicting whether a person is suffering from thyroid disease or not. The paper includes the use of different algorithms and the subsequent comparison of predictive models in terms of accuracy. The best one with higher accuracy will help in detecting thyroid disease much more accurately.

KEYWORDS- SVM, KNN, Logistic Regression, Thyroid Detection, machine learning.

I. INTRODUCTION

The secretor thyroid endocrine gland may be a tube gland and one in every of the foremost vital organs of a human body. This secretory organ secretes 2 hormones that facilitate in dominant the metabolism of the body. the 2 styles of thyroid disorders are glandular disease and glandular disorder. once the disorder happens within the body, they unleash sure variety of hormones into the body that imbalances the body's metabolism.

According to statistics, thyroid disorders are on the increase in India, just about one in ten Indian adults suffer from thyroid downside. It's been calculable that around forty-two million individuals suffer from thyroid unwellness. Predicting thyroid disorder by doctor may be a tedious method which could cause negative prediction, solely older doctor will examine the case properly. To assist to facilitate doctors, machine learning will help them in diagnosis and reduces this burden. It is to review

Thyroid unwellness irruption with the assistance of some basic visualizations techniques with the help of machine learning techniques. Machine learning may be a sub-category of computing and effectively automates the method of analytical model building and permits machines to adapt to new eventualities severally. The main objective of the Thyroid Detection is to produce the clarity whether or not an individual is caused by thyroid unwellness or not. And if the result's positive then it will predict the sort of thyroid unwellness that patient is affected from. It additionally helps in analyzing that factors that are affecting the disease, so it is prevented at earlier stage.

II. LITERATURE REVIEW

- Sunila Godara. In this work, there is the use of Logistic Regression and Support Vector Machines, for analyzing the dataset of thyroid. In this, there was comparison of mentioned two algorithms, which were based on Recall, Precision, F Score, ROC, and RMS error. The result of the comparison was that Logistic Regression came out to be the better classifier than SVM.
- Hitesh Gag. In the paper, neural network is used for segmentation and feature extraction from ultrasound images which is used in tumor prediction. All average values like accuracy score and other factors were more than 86%.
- Anita Tragic and Ricotta Mehta "Interactive Thyroid Disease Prediction System Using Machine Learning Techniques". In this work, they have used different classification algorithm techniques that are- Support Vector Machines, Decision Tree, K Nearest Neighbors, and Artificial Neural Network. The dataset has been taken from UCI repository. Based on this data the prediction and classification has been performed and the accuracy score was calculated based on the output produced.
- Yong Feng Wang. Thyroid Nodule is diagnosed for benign or malignant type with the help of images of thyroid ultrasound using image analysis, radiomics and deep learning approach has been considered in the paper. The comparison between the given two approaches has been made. The classification accuracy was 66.81%, sensitivity was 51.19%, and specificity was 75.77% for radiomics based method. And the evaluation indexes that were trained to the testing samples for the deep learning based method are 74.69%, 63.10% and 80.20%, respectively. Deep learning came out to be one of the best approaches.

- K. Polat, S. Sahan and S. Gunes,(2007) “A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis,”. In this classification problem they have used K-nn algorithm. There is an advantage of using this particular algorithm as in training, data never gets lost.
- Chandan R and Chetan Vasan, they created a web application which is used to detect the thyroid disease, a person is suffering from. In this, KNN, SVM and many more algorithms have been used. They have used ANN particularly for the prediction of the type of disease

III. IMPLEMENTATION

A. Data Description

The dataset that has been used, is “hypothyroid.csv” [1]. Data will contain different classes of thyroid and 30 columns of different values.

Column names: [age, sex, on_thyroxine, query_on_thyroxine, on_antithyroid_medication, sick, pregnant, thyroid_surgery, I131_treatment, query_hypothyroid, query_hyperthyroid, lithium, goitre, tumor, hypopituitary, psych, TSH_measured, TSH, T3_measured, T3, TT4_measured, TT4, T4U_measured, T4U, FTI_measured, FTI, TBG_measured, TBG, referral_source,class]

"Class" column will have four unique values “negative, compensated_hypothyroid, primary_hypothyroid, secondary_hypothyroid ”.

B. Proposed Algorithm

The machine learning algorithms used- KNN (K-nearest neighbors) algorithm, Random Forest, Decision Tree, Support Vector Machines(SVM) algorithm, Logistic Regression and amongst these algorithms whichever algorithm performs well and have a higher accuracy, it will be used to predict the risk of getting thyroid disease.

Brief description of the machine learning algorithms:

- KNN- KNN stands for K-Nearest Neighbour. This type of machine learning algorithm can be used for both type of machine learning that is supervised and unsupervised machine learning. This algorithm finds the similarities between the available cases and new data point that can be near about each other.
- Random Forest Classifier- Random Forest is very famous machine learning algorithm. It belongs to the supervised learning technique. This algorithm can be used as classification and regression machine learning problems Random Forest is a classifier that contains numerous decision trees on various subsets of the given dataset and it takes the mean, to improve the accuracy of prediction with respect to the dataset.
- Decision Tree Classifier- Decision Tree can be used for regression problems as well as classification problems. It resembles to tree, in which, inner nodes have the features, branches have the decision rules and each and every leaf node contains the output.
- Support Vector Classifier- Support Vector Machines or SVM is the Supervised Machine Learning algorithm, that can be used as Classification and regression problems. Primarily, it is used basically for Classification problems. The main goal of the SVM is to

create the best fit line or decision boundary which can differentiate n-dimensional space into classes, so that it can put the new data point in the correct category in future.

- Logistic Regression- This algorithm is used for classification type of problems.
- This algorithm used to predict the output of a categorical dependent variable. Hence, the output must show categorical or discrete value. It can be Yes or No, 0 or 1, true or False, etc.,

C. Accuracy Score

In this paper, the algorithm that will be used is “Support Vector Classifier” algorithm because the accuracy score of this algorithm is higher than other algorithms.

Table 1: Comparison of accuracy score

	Algorithms	Accuracy Score
1	Decision Tree Classifier	88.344371
2	Support Vector Machine	94.172185
3	Random Forest Classifier	93.907285
4	K neighbors Classifier	93.377483
5	Logistic Regression	93.907285
	Highest Accuracy “Support Vector Machine”	94.172185

We can observe by looking at TABLE1 that amongst all the algorithms Support Vector Classifier performs well and has a higher accuracy score which is around 94.17%,

D. Prediction

```
svc.predict([[67, 1, 1, 0,
             0, 1, 0, 0,
             1,1, 1, 1,
             1, 1, 1, 1, 1, 1, 1, 11,
             1]])
```

```
array([2])
```

Figure 1: Prediction

The Fig1 shows the prediction of the details that we have entered, gives us output 2 that means, the type of thyroid disease came out to be “primary_hypothyroid”.

E. Confusion Matrix

A confusion matrix is a matrix of NxN used to calculate the interpretation of the model specifically of classification type of model. In this N represent the target classes. The confusion matrix shows the comparison between the actual and the predicted values. It basically shows the whole overview of the model that means the error and the true values, the model is predicting with respect to the actual values.

```
con=confusion_matrix(y_test,y_pred5)
con
array([[698,  0,  4],
       [ 32,  0,  0],
       [  8,  0, 13]], dtype=int64)
```

Figure 2: Confusion Matrix

In confusion matrix there are 4 values:

- TP - TP represents True Positive values that means the model correctly predicts the classes that are positive.
- TN- TN represents True Negative values that means the model correctly predicts the classes that are negative.
- FP- FP represents False Positive values that means, the model incorrectly predicts the classes that are positive.
- FN- FN represents False Negative that means, the model incorrectly predicts the classes that are negative.

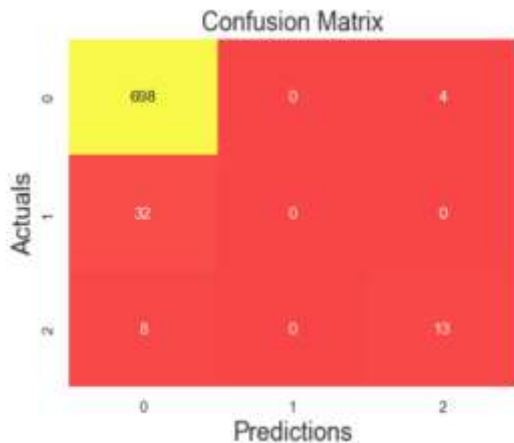


Figure 3: Confusion matrix Visualization

IV. CONCLUSION

The “Thyroid Detection” has been designed to make it easier for the user or patients to not only detect or predict whether they are suffering from thyroid or not but also it will be providing us the type of thyroid you are affected with in the case of positive results. This will be achieved by performing data validations, data pre-processing, EDA(Exploratory Data Analysis). Then by using Support Vector Machines algorithm, the user will be able to predict the results with the help of details provided by the user.

For prediction purpose, user need to enter certain details such as age, sex, Are you on thyroxine treatment or not? Are you pregnant or not? Are you on antithyroid medication or not? etc. And then with the help of entered details the prediction will happen.

In this paper, the various python packages and modules and libraries have been used like flask, pymysql, numpy, pandas, sklearn, etc. which will help us achieve the prediction.

ACKNOWLEDGMENT

We thank Krish Naik who is a co-founder of INEURON, for providing the best knowledge in terms of machine learning and deep learning and for sharing lots of machine learning and deep learning projects that has helped in this work.

REFERENCES

- [1] Dataset collected from Kaggle
<https://www.kaggle.com/datasets/nguyenthilua/hypothyroid.csv>
- [2] Sunila Godara.They have used Logistics Regression and SVM machine learning algorithm Technique to analyze Thyroid Dataset.
https://www.researchgate.net/publication/311496931_A_Novel_Weighted_Class_based_Clustering_for_Medical_Diagnostic_Interface
- [3] Hitesh Garg. Neural Network is used for feature extraction and segmentation from Ultrasound images to predict the tumors. The accuracy and other factors were measured and all the average values were above 86%.
https://www.researchgate.net/publication/271557279_Segmentation_of_thyroid_gland_in_ultrasound_image_using_neural_network
- [4] Ankita Tyagi and Rikitha Mehra “Interactive Thyroid Disease Prediction System Using Machine Learning Techniques”.
https://www.researchgate.net/publication/331533687_Interactive_Thyroid_Disease_Prediction_System_Using_Machine_Learning_Technique
- [5] YongFeng Wang,(2020). “Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images” published on IEEEAccess.
<https://ieeexplore.ieee.org/abstract/document/9034133>
- [6] K. Polat, S. Sahan and S. Gunes (2007) “A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis,” Expert Systems with Applications,(vol. 32, pp. 1141-1147).
<https://www.sciencedirect.com/science/article/abs/pii/S001048250600076X>
- [7] Chandan R and Chetan Vasan, they developed the web application on detecting the type of thyroid disease a person is suffering from.
https://www.researchgate.net/publication/352379472_THYROID_DETECTION_USING_MACHINE_LEARNING
- [8] Tiwari, V., Diwan, T.D., Miri, R.: Design and implementation of an efficient relative model in cancer disease recognition. IJARCSSE (2013)
<https://www.ijert.org/design-and-implementation-of-an-efficient-relative-model-in-cancer-disease-recognition>
- [9] Khaleel, M.A.: A survey of data mining techniques on medical data for finding frequent diseases. IJARCSSE (2013)
<https://www.semanticscholar.org/paper/A-Survey-of-Data-Mining-Techniques-on-Medical-Data-Khaleel-Pradham/cf7e4693021ac09e3ea960bf44b70e45654b7501>
- [10] Parthiban, G., Srivatsa, S.K.: Applying machine learning methods in diagnosing heart disease for diabetic patients. IJAIS, 25–30 (2012)
<https://www.semanticscholar.org/paper/Applying-Machine-Learning-Methods-in-Diagnosing-for-Parthiban-Srivatsa/3362c5df179e709aa4ad411b10ce852cef3e30a>
- [11] Vembandasamy, K., Sasipriya, R., Deepa, E.: Heart diseases detection using naive bayes algorithm. IJISSET (2015)
https://www.researchgate.net/publication/341831889_Heart_Diseases_Prediction_using_Deep_Learning_Neural_Network_Model
- [12] Iyer, A., Jeyalatha, S., Sumbaly, R.: Diagnosis of diabetes using classification mining techniques. IJDKP, 1–14 (2015)
<https://scirp.org/reference/ReferencesPapers.aspx?ReferenceID=1965043>
- [13] Tan, K.C., Teoh, E.J., Yu, Q., Goh, K.C.: A hybrid evolutionary algorithm for attribute selection in data mining. IJDKP, 8616–8630 (2009)
https://www.researchgate.net/publication/220215347_A_hybrid_evolutionary_algorithm_for_attribute_selection_in_data_mining
- [14] Jian, A.: Machine learning techniques for medical diagnosis. ICSTAT (2015)
- [15] Alic, B.: Machine learning techniques for classification of diabetes and cardiovascular diseases. Mediterr.Ean Conf. Embed. Comput. (2017)

https://www.researchgate.net/publication/318503093_Machine_Learning_Techniques_for_Classification_of_Diabetes_and_Cardiovascular_Diseases

- [16] Vijayarani, S., Dhayanand, S.: Liver disease prediction using SVM and naive bayes algorithms. *Int. J. Sci. Eng. Technol. Res. (IJSETR)* 4, 816–820 (2015)
https://www.researchgate.net/publication/339551659_Liver_Disease_Prediction_using_SVM_and_Naive_Bayes_Algorithms
- [17] Karlik, B.: Hepatitis disease diagnosis using back propagation and the naive bayes classifiers. *J. Sci. Technol.* 1, 49–62 (2011)
https://www.researchgate.net/publication/265946801_Hepatitis_Disease_Diagnosis_Using_Backpropagation_and_the_Naive_Bayes_Classifiers
- [18] Razia, S., Narasingarao, M.R.: A Neuro computing framework for thyroid disease diagnosis using machine learning techniques. 95(9), 1996–2005. ISSN: 1992–8645 www.jatit.org. E-ISSN: 1817-3195
<https://www.jatit.org/volumes/Vol95No9/11Vol95No9.pdf>
- [19] Razia, S. and Rao, M.N.: Machine learning techniques for thyroid disease diagnosis - a review. (INDJST) *Indian J. Sci. Technol.* 9(28), Article number 93705 (2016). ISSN: 09746846
https://www.researchgate.net/publication/339020889_Machine_Learning_Techniques_for_Thyroid_Disease_Diagnosis_A_Systematic_Review
- [20] Razia, S., Narasingarao, M.R.: Development and analysis of support vector machine techniques for early prediction of breast cancer and thyroid. (JARDCS) *J. Adv. Res. Dyn. Control. Syst.* 9(6), 869–878 (2017). ISSN: 1943-023X
https://www.researchgate.net/publication/319094774_Development_and_analysis_of_support_vector_machine_techniques_for_early_prediction_of_breast_cancer_and_thyroid
- [21] Fathima, A., Manimegalai, D.: Predictive analysis for the arbovirus-dengue using SVM classification. *Int. J. Eng. Technol.* 521–527 (2012)
<https://www.semanticscholar.org/paper/Predictive-Analysis-for-the-Arbovirus-Dengue-using-Fathima-Manimegalai/e6d9121158ae0377ef1dcb6c62f391ce0c112bd2>