*International Journal of Engineering Applied Sciences and Technology, 2021*
**Vol. 5, Issue 9, ISSN No. 2455-2143, Pages 173-177**
Published Online January 2021 in IJEAST (http://www.ijeast.com)

# THYROID DETECTION USING MACHINE LEARNING

Chandan R
Department of ECE
DSU, Bangalore, Karnataka, India

Chetan Vasan
Department of ECE
DSU, Bangalore, Karnataka, India

Chethan MS
Department of ECE
DSU, Bangalore, Karnataka, India

Devikarani H S
Department of ECE
DSU, Bangalore, Karnataka, India

***Abstract*—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 neighbours, ANN- Artificial Neural Network are used to predict the patient's risk of getting thyroid disease. Web app is created to get data from users to predict the type of disease.**

*Keywords*— **SVM, KNN, ANN, Logistic Regression, Flask.**

## I. INTRODUCTION

The evolvement computational biology is used in healthcare industry. It allows collection of stored patient data for the prediction of the disease. There are prediction algorithms which are available for the diagnosis of the disease at early stages. The medical information systems are rich of datasets but there are only few intelligent systems which can easily analysis the disease. Over a period of time, the machine learning algorithms have started playing a crucial role in resolving the complex and non-linear problems in the developing model. In any disease prediction models are used to override the features that can be selected from different datasets which can be used in classification in healthy patient as accurate as possible. If this is not done, misclassification can lead to a healthy patient getting unnecessary treatment.

The Thyroid gland is an endocrine gland present in the human neck beneath the Adam's apple which help in secretion of thyroid hormone that influence the rate of metabolism and protein synthesis. The thyroid hormones are useful in counting how briskly the heart beats and how fast we burn calories. The thyroid secretes two types of active hormones called levothyroxine (T4) and triiodothyronine (T3). These hormones help in regulating the body temperature. These also aid in energy-bearing and transmission in every part of the body and decisive in protein management. Iodine is considered as the main building block of the thyroid gland. It's prostrated in few specific problems. Undersupply of these hormones can lead to hyperthyroidism. There are many originations related to hyperthyroidism and underactive thyroids. There are various kinds of medications like thyroid surgery is liable to ionizing radiation, continual tenderness of the thyroid, deficiency of iodine and lack of enzyme to make thyroid hormones.

## II. DEFINITION OF PROBLEM

According to statistics, thyroid disorders are on the rise in India. Approximately 1 in 10 Indian adults suffer from thyroid problem. It has been estimated that around 42 million peoples suffer from thyroid disease. Predicting thyroid disorder by doctor is a tedious process which might lead to negative prediction, only experienced doctor can examine the case properly. To assist doctors machine learning can help them in diagnosis of disease and reduces their burden.

*International Journal of Engineering Applied Sciences and Technology, 2021*
**Vol. 5, Issue 9, ISSN No. 2455-2143, Pages 173-177**
Published Online January 2021 in IJEAST (http://www.ijeast.com)

### A. Objective

- The main objective is to develop a system which can predict the type of thyroid disease that patient is affected from.
- To predict thyroid disease with usage of minimum number of parameters.
- To predict all possible types of Thyroid diseases.

## III. LITERATURE SURVEY

[1] Ankita Tyagi and Rikitha Mehra "*Interactive Thyroid Disease Prediction System Using Machine Learning Techniques*" 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), 20-22 Dec, 2018, Solan, India.

In this work they use different classification algorithms- Decision Tree, Support Vector Machine, Artificial Neural Network, k-Nearer-Neighbor algorithm. Based on the data set obtained from UCI Repository, classification and prediction was performed and accuracy was obtained based on output produced. They have analyzed accuracy of algorithms used and comparison is made to find best technique with high accuracy.

Sunila Godara. [3] They have used Logistics Regression and SVM machine learning Technique to analyze Thyroid Dataset. Comparison was made between these two algorithms based on Precision, Recall, F measure, ROC, RMS error. Logistic Regression turned out has best classifier.

YongFeng Wang. [2] Thyroid Nodule is diagnosed for benign or malignant type using Ultrasound images of thyroid by image analysis - radiomics and deep learning based approaches. Comparison is made between these two approaches. The classification accuracy, sensitivity, and specificity of applying radiomics based method are 66.81%, 51.19% and 75.77%, respectively, while the evaluation indexes for the deep learning based method trained to the testing samples are 74.69%, 63.10% and 80.20%, respectively. Deep learning turned out as best approaches.

Hitesh Garg. [4] Feed Forward 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%.

## IV. SYSTEM ANALYSIS AND DESIGN

### A. METHODOLOGY

For predicting Thyroid disease analyzing blood report is required to analyze and predict disease. Thyroid blood test data set analysis will be conducted using various supervised machine learning classifier techniques. Based on the accuracy of different algorithm, best accuracy algorithm will be chosen to fetch the result.

For first part, thyroid data set is taken from UCI repository. The dataset of hyperthyroidism and hypothyroidism is used where hyper and hypo are the two labels. These data set need to be checked before feeding it to training. There may be presence of null data or unnecessary data, this should undergo data cleaning to remove such data.

Cleaned data is used as training data and test data, which is fed as input to the algorithm.

The algorithm extracts the features from different dataset to classify the data according to the labels. To check the accuracy of the prediction, test data is fed to the algorithm.

Based on the feature extracted, probability will be generated for test data by comparing the features of both. Highest probability value will be classified to that particular label whether it is hyperthyroidism or hypothyroidism.

Web app is developed using python Flask (back end) and web application is designed using HTML5 ( front end) and CSS, where the chosen ML model will be linked with web app and HTML. The users blood test data will be entered in web app front end and the back end will process the data using model and result will be designed on screen.

*International Journal of Engineering Applied Sciences and Technology, 2021*
**Vol. 5, Issue 9, ISSN No. 2455-2143, Pages 173-177**
Published Online January 2021 in IJEAST (http://www.ijeast.com)

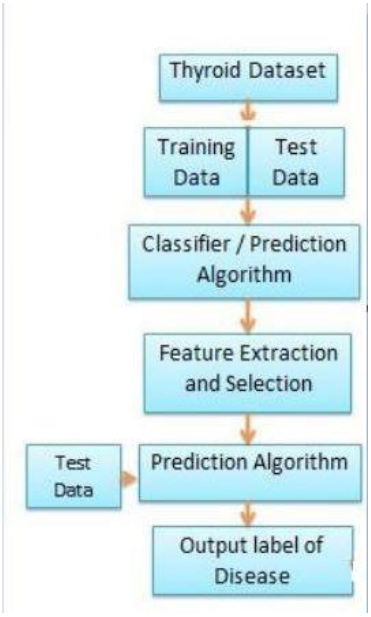


Fig 1: work flowchart

B. Data set

- A new_thyroid dataset is accessible in UCI machine learning repository and this set has 215 samples and 5 features.
- 5 lab tests are used to try to predict whether a patient's thyroid to the class is normal, hypothyroidism or hyperthyroidism. The diagnosis (the class level) was based on a complete medical record.
- This dataset has 150 instance of normal class, 35 instance of hyperthyroidism class and 35 instance of hypothyroidism class.
- Class attribute, T3 resin uptake test, total T4 , total T3 , TSH , difference of TSH value after injection of 200 micrograms of thyrotropin releasing hormone (as shown in fig 2).

```
[ ] data.shape
    (215, 6)
[ ] data.head(5)
```

| | class | T3_resin_uptake | total_thyroxin(T4) | total_T3 | TSH | diff_TSH |
|---|---|---|---|---|---|---|
| 0 | 1 | 107 | 10.1 | 2.2 | 0.9 | 2.7 |
| 1 | 1 | 113 | 9.9 | 3.1 | 2.0 | 5.9 |
| 2 | 1 | 127 | 12.9 | 2.4 | 1.4 | 0.6 |
| 3 | 1 | 109 | 5.3 | 1.6 | 1.4 | 1.5 |
| 4 | 1 | 105 | 7.3 | 1.5 | 1.5 | -0.1 |

Fig 2: dataset header

## V. EVALUATION AND TESTING

### A. Result

- Heat map

A heat map (or heatmap) is a data visualization technique that shows magnitude of a phenomenon as color in two dimensions. The variation in color may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space. Correlation between parameters of our dataset is interpreted and pictorial view is obtained as shown in fig 3.

Fig 3: Heat Map

KNN algorithm has been implemented and 93.84% accuracy score is obtained as shown in fig 4.

```
[ ] from sklearn.metrics import confusion_matrix, accuracy_score
    cm = confusion_matrix(y_test, y_pred)
    print(cm)
    accuracy_score(y_test, y_pred)

    [[49  0  0]
     [ 2  9  0]
     [ 2  0  3]]
    0.9384615384615385
```

Fig 4: KNN algorithm accuracy

SVM algorithm has been implemented and 95.38% accuracy score is obtained as shown in fig 5.

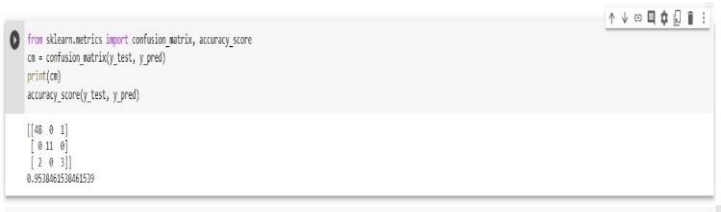


Fig 5: SVM algorithm accuracy

ANN algorithm has been implemented and 75.38% accuracy score is obtained as shown in fig 6.

*International Journal of Engineering Applied Sciences and Technology, 2021*
**Vol. 5, Issue 9, ISSN No. 2455-2143, Pages 173-177**
Published Online January 2021 in IJEAST (http://www.ijeast.com)

```
[ ] y_pred = ann.predict(X_test)

[ ] from sklearn.metrics import confusion_matrix, accuracy_score
    cm = confusion_matrix(y_test, y_pred)
    print(cm)
    accuracy_score(y_test, y_pred)

    [[49  0  0]
     [11  0  0]
     [ 5  0  0]]
    0.7538461538461538
```

Fig 6: ANN algorithm accuracy

Decision Tree algorithm has been implemented and 92.3% accuracy score is obtained as shown in fig 7.

```
[ ] from sklearn.metrics import confusion_matrix, accuracy_score
    cm = confusion_matrix(y_test, y_pred)
    print(cm)
    accuracy_score(y_test, y_pred)

    [[46  1  2]
     [ 1 10  0]
     [ 1  0  4]]
    0.9230769230769231
```

Fig 7: Decision Tree algorithm accuracy

Logistic Regression has been implemented and 96.92% accuracy score is obtained as shown in fig 8.

```
[ ] from sklearn.linear_model import LogisticRegression
    classifier = LogisticRegression(random_state = 0, max_iter=500)
    classifier.fit(X_train, y_train.ravel())

    LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                       intercept_scaling=1, l1_ratio=None, max_iter=500,
                       multi_class='auto', n_jobs=None, penalty='l2',
                       random_state=0, solver='lbfgs', tol=0.0001, verbose=0,
                       warm_start=False)

[ ] y_pred = classifier.predict(X_test)

[ ] classifier.score(X_test,y_test)

    0.9692307692307692
```

Fig 8: logistic regression algorithm accuracy

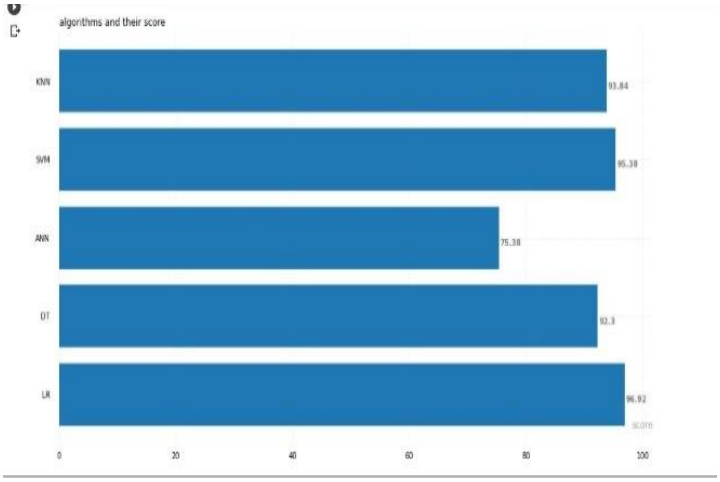


Fig 9: bar graph of score

Since accuracy obtained from logistic regression model (96.92%) was highest, this model will be considered for our prediction model. To save this model for interfacing with web app, joblib library is used and file "thyroid_prediction_lr.pkl" will be created.

Web app is created to interface user and the trained model. Python Flask coding is used to create web app and html is used to design web page.

Fig 10 shows the data entered into the parameter box of web app created, which will be used at back end by model to predict thyroid disease.

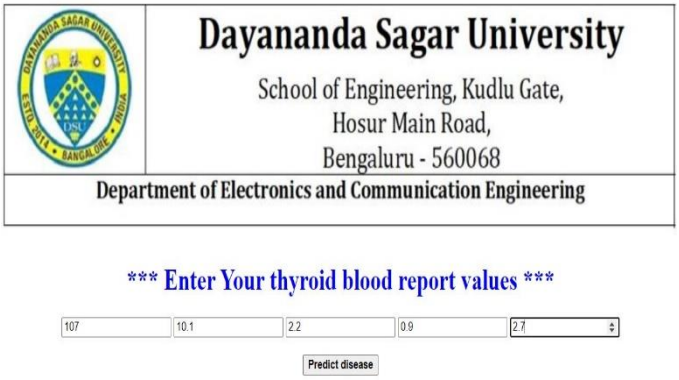


Fig 10: result example

For the data given through the web page as shown in fig 10, after processing the data by the model, result will be displayed in the same web page as shown in fig 11.

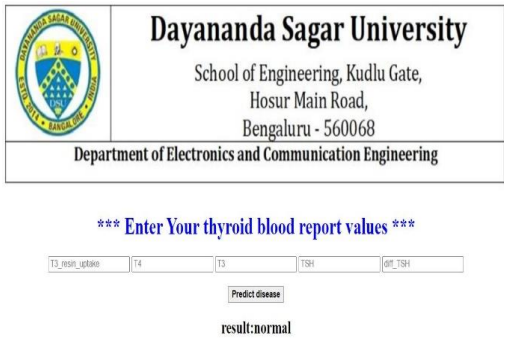


Fig 11: example result2

## VI. CONCLUSION

Thyroid Detection using Machine Learning is a project idea that aims a smart and precise way to predict thyroid disease. 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.

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.

## VII. REFERENCE

[1]Ankita Tyagi and Ritika Mehra. (2018).“Interactive Thyroid Disease Prediction System using Machine Learning Techniques” published on ResearchGate.

[2] YongFeng Wang,(2020). “Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images” published on IEEEAccess.

[3] Sunila Godara,(2018). “Prediction of Thyroid Disease Using Machine Learning Techniques” published on IJEE.

[4] Hitesh Garg,(2013). “Segmentation of Thyroid Gland in Ultrasound image using Neural Network” published on IEEE.

[5] L. Ozyılmaz and T. Yıldırım,(2002). “Diagnosis of thyroid disease using artificial neural network methods,” in: Proceedings of ICONIP’02 9th international conference on neural information processing (Singapore: Orchid Country Club, pp. 2033–2036).

[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).

[7] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab,(2009) “Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM,” in 3rd International Conference on Bioinformatics and BiomedicalEngineering. ICBBE 2009.

[8] G. Zhang, L.V. Berardi,(2007) “An investigation of neural networks in thyroid function diagnosis,” Health Care Management Science,1998, (pp. 29-37.)

[9] V. Vapnik,(2012).Estimation of Dependences Based on Empirical Data, Springer, New York.

[10] Obermeyer Z,(2016). Emanuel EJ. Predicting the future— big data, machine learning, and clinical medicine. N Engl ; (375:12161219).

[11] Breiman L.(2001) StatisticalModeling: the two cultures.Stat Sci. ;16:199-231..

[12] Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. (2017) Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol. ; 9:245-250

[13] S. Godara and R. Singh,(2016) "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", Indian Journal of Science and Technology, (Vol. 910).

[14] Sunila, Rishipal Singh and Sanjeev Kumar.(2016) "A Novel Weighted Class based Clustering for Medical Diagnostic Interface." Indian Journal of Science and Technology (Vol 9).

- https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease
- https://www.tutorialspoint.com/python_data_science/index.html
- https://www.tutorialspoint.com/flask/index.htm