Prediction Of Thyroid Disease In Early Stage Using Feature Selection And Classification Techniques

[1]Prof. Kanchan Umavane, [2]Mr. Jay Motiram Gulvi, [3]Mr. Pavan Vilas Khade, and [4]Ms. Samiksha V Gaikwad.

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]kanchanumavane2020@gmail.com, [2]jaygulvi7077@gmail.com, [3]pavanvilaskhade491@gmail.com, [4]samikshagaikwad161@gmail.com

Abstract - One of the most prevalent illnesses affecting women is thyroid disease. Thyroid illness frequently manifests as hypothyroidism. It is obvious that people with hypothyroidism are typically female. Because the majority of people are unaware of that illness, it is quickly developing into a serious illness. It is crucial to catch it early so that doctors can give patients better treatment and prevent it from becoming a severe problem. Machine learning disease prediction is a challenging job. In forecasting diseases, machine learning is crucial. Once more, unique feature selection methods have aided in the process of disease assumption and forecast.[1]

Keywords—Thyroid disease , Feature selection , Data mining , Recursive Feature Selection, Machine learning Classification.

I. INTRODUCTION

At the current state, the thyroid is one of the most critical diseases of all and it has quite the potential to be transformed into a common disease among the female mass. Experts estimate that 50 million individuals in Bangladesh have thyroid disease. Among them, women have a 10 times greater chance of developing thyroid disease.Though a vast majority of 50 million people are affected with thyroid disease, yet almost 30 million people among them are totally not aware of this condition. A study from the Bangladesh Endocrine Society(BES) depicts that around 20-30% of females are suffering from thyroid disease. The thyroid is a gland that is situated in the middle of the neck in our body. It is butterfly-shaped and small in size. It secretes several hormones that are mixed with blood and travel across the body to control various activities. The thyroi hormone is responsible for conserving metabolism, sleep, growth, sexual function, and mood. Depending on the secretion of thyroid hormone human can feel tired or restless and also may have weight loss. There are two main thyroid hormones: Triiodothyronine (T3) and Thyroxin (T4). These two hormones are mainly responsible for maintaining the energy in our bodies. Thyroid Stimulating Hormone(TSH) is produced by the pituitary gland that helps the thyroid gland to release T3 and T4.[1]

II. AIMS AND OBJECTIVE

a) Aim

The thyroid is a gland in our body that is located in the centre of the neck. It is tiny and butterfly-shaped. It releases a number of hormones that mix with blood and circulate throughout the body to regulate different bodily functions. The thyroi hormone regulates mood, growth, sleep, sexual function, and metabolism. May experience fatigue or restlessness and weight loss depending on the thyroid hormone's release. It is crucial to catch it early so that doctors can give patients better treatment and prevent it from becoming a significant problem. Machine learning illness prediction is a challenging task. In forecasting diseases, machine learning is crucial.[1]

b) Objective

An endocrine gland located in the neck, the thyroid's job is to create the hormones (FT3 and FT4) that are then released into the bloodstream. Heart rate, body temperature, and metabolism the body's process of utilising and consuming nutrients are just a few of the things that thyroid hormones control. Major disorders may develop in any situation where the thyroid gland functions either above or below normal levels (hyperthyroidism with high hormones vs. hypothyroidism with decreased hormones). Moreover, one or more swellings that develop inside the thyroid gland might cause it to become inflamed (thyroiditis) or to expand (nodules, multinodular goiter). Malignant tumours may develop in some of these modules. The management of thyroid problems is therefore a highly important subject.[1]

III. LITERATURE SURVEY

Paper1: Using Classification and Regression Trees, Early Diagnosis of Heart Disease

DOI : 10.5281/zenodo.7857229

The ultimate objective of this study is to develop a system for diagnosing heart sounds that will aid medical professionals in patient auscultation. the goal of reducing the number of unnecessary echocardiograms and of preventing the release of newborns that are in fact affected by a heart disease. In this study, 99.14% accuracy, 100% sensitivity, and 98.28% specificity were obtained on the dataset used for experiments.[6]

Paper 2: An Intelligent System for the Classification and Diagnosis of Thyroid Disease

This essay suggests a system for classifying and diagnosing thyroid illness, along with a description of the condition and suggestions for good health. For categorization, support vector machines are employed. The Particle Swarm Optimization uses SVM criteria for optimisation. A window is given to the user so that they can input information like the values of TSH, T3, T4, etc. While entering the numbers, there might be some values that are missing. The K-Nearest Neighbor algorithm is used to approximate the values that are absent from the user's input.[5]

Paper 3: Genetic Algorithms for the Diagnosis of Thyroid Disease Using PNN and SVM

Thyroid chemicals are created by the thyroid gland to aid in controlling the body's metabolism. There are two types of thyroid hormone production abnormalities. Both hypothyroidism, which is associated with inadequate thyroid hormone production, and hyperthyroidism, which is associated with excessive thyroid hormone production. For the thyroid diagnosis, it is crucial to distinguish between these two illnesses. In order to classify data, support vector machines and stochastic neural networks are suggested. To deal with redundant and pointless features, these techniques mainly depend on potent classification algorithms.[8]

IV. EXISTING SYSTEM

At the current state, the thyroid is one of the most critical diseases of all and it has quite the potential to be transformed into a common disease among the female mass so machine learning algorithms such as decision trees, artificial neural networks, and support vector machines to predict thyroid disease based on patient data such as age, gender, and thyroid hormone levels. The study found that the support vector machine model had the highest accuracy in predicting thyroid disease. Predicting disease in machine learning is a difficult task It is not often known in advance which features will provide the best discrimination.[1]

V. COMPARATIVE STUDY

Sr. No.	Author	Project Title	Publication	Technology	Purpose
1	A. K. Aswathi, A. Antony	An Intelligent System for Thyroid Disease Classification and Diagnosis	IEEE, 2018	ICICCT	To collect more stable and uniform data for Thyroid Disease classification
2	A. Tyagi, R. Mehra, and A. Saxena	Interactive Thyroid Disease Prediction System Using Machine Learning Technique	IEEE, 2019	PDGC	The Main task is to detect disease diagnosis at the early stages with higher accuracy.
3	Q. Pan, , Y. Zhang, M. Zuo, L. Xiang, and D. Chen	Improved Ensemble Classification Method of Thyroid Disease Based on Random Forest	IEEE,2017	ITME	can enlarge the discrepancy of the base classifiers and improve the accuracy of the ensemble classifier.
4	K. Pavya, B. Srinivasan	Feature Selection Algorithms To Improve Thyroid Disease Diagnosis	IEEE,2017	ICIGEHT	To compare the utility of several supervised machine learning (ML) algorithms

Table no.1:Comparative analysis of existing system

VI. PROBLEM STATEMENT

The majority of people in the world are afflicted by thyroid diseases and disorders, which are common hormonal issues. Thyroiditis and thyroid cancer are two conditions and illnesses related to the thyroid. One of the most noticeable pure endocrine glands, the thyroid is situated at the front of the neck and surrounds the trachea (Beynon & Pinneri, 2016). It resembles a butterfly and has two long wings that stretch to either side of the human neck. It controls how our bodies regulate hormones. In addition to controlling blood pressure, body temperature, and heart rate, thyroid hormones also aid in the digestion of fat, protein, and carbs. On the other hand, a change in secretion can cause a range of physiological and mental issues, including brain.

VII. PROPOSED SYSTEM

This study developed an optimised SVM method for classifying hypothyroid disorders. Using classification machine learning techniques, including Random Forest , KNN (K Nearest Neighbor), SVM (Support Vector Machines), LR (Logistic Regression), and NN, they proposed a way of detecting the level of hypothyroid disorder and gives the better accuracy in result and also Identify the most important variables that contribute to the risk of accidents in the industry.[1]

VIII. ALGORITHM

The Algorithm for Prediction of thyroid Disease :

Step.1: Start

Step.2: Import necessary libraries

```
from django.shortcuts import render, HttpResponse
from .forms import UserRegistrationForm
from django.contrib import messages
from .models import UserRegistrationModel
from django.conf import settings
import os
import pandas as pd
```

Step.3: Load dataset

```
path = os.path.join(settings.MEDIA_ROOT , "full_dataset.csv")
df = pd.read_csv(path)
```

Step.4: Split the data into training and testing sets

```
X = df.iloc[:, :-1].values
y = df.iloc[:, -1].values
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.80, random_state=0)
```

Step.5: Train the logistic regression and SVM classifier

```
model = LogisticRegression()
model.fit(X_train, y_train)
model = SVC(kernel='rbf')
model.fit(X_train, y_train)
```

Step.6: Train a random forest and decision tree classifier

```
model = RandomForestClassifier()
model.fit(X_train, y_train)
model = DecisionTreeClassifier()
model.fit(X_train, y_train)
```

Step.7: Evaluate the logistic regression and SVM model

```
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
```

Step.8: Evaluate the random forest and decision tree classifier

```
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
```

Step.9: Print the results

```
print('Decision Tree Classifier Results:')
print('Accuracy:', accuracy)
print('Precision Score:', precision)
print('Random Forest Classifier Results:')
print('RF Accuracy:', accuracy)
print('RF Precision Score:', precision)
print('SVM Precision Score:', precision)
print('SVM Accuracy:', accuracy)
print('GB Accuracy:', accuracy)
print('GB Precision Score:', precision)
```

Step.10: Exit

IX. MATHEMATICAL MODEL

Random Forest : Random Forest are a supervised machine learning alorithan that is widely used in regression and classification problems.

$$RFfi_i = \frac{\sum_{j\in all\,trees} normfi_{ij}}{T}$$

Working : Select random K data points from the traing set then build the the decision trees with the selected data points then choose the number N for decision trees that can want to build , for new data points find the predictions of each decision tress.

SVM : In the Support vector machine or SVM is one of the most popular supervised learning algorithams, which is used for classification as well as regression problems.the goal of SVM is to create the best line or best decision boundry.

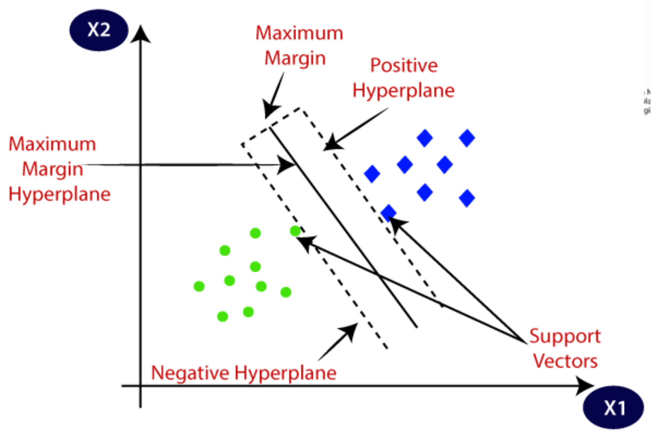


Fig.1: Support Vector Machine

equation of a hyperplane is w.x+b=0. If the value of w.x+b>0 then person can say it is a positive point otherwise it is a negative point.

Logistic Regression: Logistic regression is a supervised machine learning alorithams mainly used for classification tasks.

g(z) = 1 / (1 + e^(-z))

Dicision Tree : Decision tree is the most powerful and popular tool for classification and prediction.

Naïve Bayes : Naïve Bayes classifiers are a collection of classification algorithams based on Bayes Theorem. Bayes' Theorem finds the likelihood of an event happening given the likelihood of another event that has just happened. Bayes' hypothesis is expressed numerically as the following condition:

P(A | B) = P(B | A)P(A) / P(B)

X. SYSTEM ARCHITECTURE

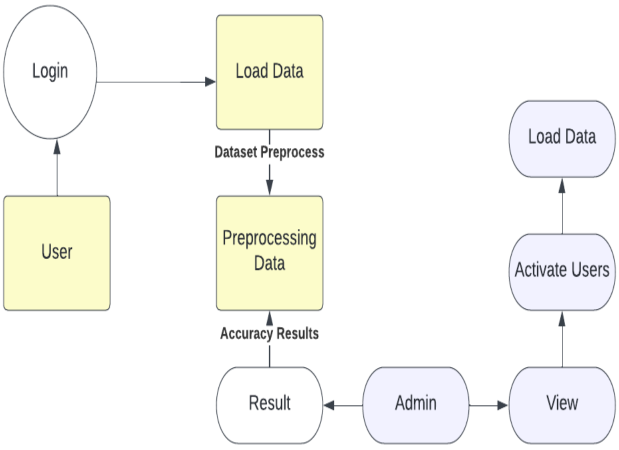


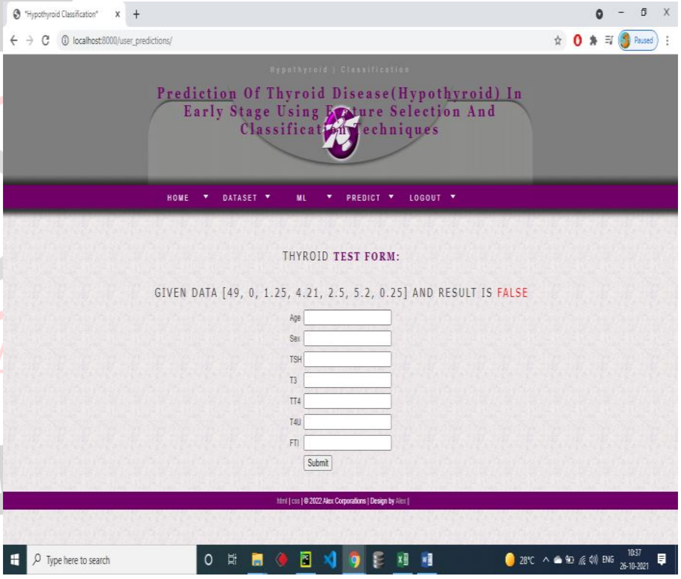
Fig.2: System Architecture

Explanation : the dataset is taken from the repository will contain all the features, classification and there respective values of the patient, the architecture are work like first user login in the system then load and process the data then user get accurate result. From the dataset the correlated features that plays vital role in detection of disease will be selected. The selected dataset is pass to the all classification model to predict the result and the performace measure will be calculated for each models. Accuracy, precision, F1 score, recall, ROC will be calculated and will be used for analysis. The model with the overall highest score will be the best classifier.

XI. ADVANTAGES

- Helps solve complex real- world problems with several constraints.
- Tackle problems like having little or almost no labeled data availability.
- Opens the door to the eventual development of Artificial General Intelligence.
- Without any human oversight, it instantly recognises the crucial characteristics.
- Less computational power or resources like RAM, CPU, GPU or TPU, etc.
- Less quantity of data, can achieve more accuracy.

XII. DESIGN DETAIL


"Hypothyroid Classification"
localhost:8000/user_predictions/
Paused
Hypothyroid | Classification
HOME
DATASET
ML
PREDICT
LOGOUT
Type here to search
28°C
ENG
10:37
26-10-2021

Prediction Of Thyroid Disease(Hypothyroid) In Early Stage Using Feature Selection And Classification Techniques

THYROID TEST FORM:

GIVEN DATA [49, 0, 1.25, 4.21, 2.5, 5.2, 0.25] AND RESULT IS FALSE

Age
Sex
TSH
T3
TT4
T4U
FTI

Submit

html | css | © 2022 Alex Corporations | Design by Alex |

Fig 3: Result

The result analysis is done on the basis of scores obtained by classification models of performance measures (Accuracy, precision, F1 score, recall, ROC). Each models have different score each patients have different age , gender , TSH, T3, FT1 values so according to that result will be show. The model with the overall highest score will be the best classifier.

XIII. CONCLUSION

Thus we have tried to implement the paper “Prediction ofThyroid Disease(Hypothyroid) in early stage using Feature selection and Classification Techniques”, Md Riajuliislam Khandakar Zahidur Rahim, Antara Mahmad, IEEE 2021 and the conclusion as follows :

DOI : 10.5281/zenodo.7857229

The felature selection technique RFE helps user to get better accuracy with all other classifiers. In our findings, we have seen that RFE significantly helps us to predict hypothyroid in the primary stage by using a real-time dataset. It is very difficult for us to collect data in this current pandemic situation. As a result, we have collected only 519 data. So, considering the situation and the constraint we couldn't study on a larger dataset. In our study, we have seen that there have not been done any work in thyroid based on Bangladesh before. We have a limitation of data to work with. So, in the future, we want to work with a larger dataset and we hope that more people from our country will show interest to work on this disease that will help us to find a better solution able to predict disease in the primary stage with better accuracy.

REFERENCE

[1] Md Riajuliislam; Khandakar Zahidur Rahim; Antara Mahmud," Prediction OfThyroid Disease(Hypothyroid) In EarlyStage Using Feature Selection And Classification Techniques" 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) INSPEC Accession Number: 20633592 DOI: 10.1109/ICICT4SD50815.2021.9397052.

[2]Prof Vishal R. Shinde,"Analysis and prediction of cardiovascular disease using machine learning classifiers" in IJREAM, ISSN : 2454-9150, Volume 08, Issue 01, APR 2022 Special Issue (indexed in scope databasehttps://sdbindex.com/documents/00000217/00001-75837.pdf).

[3] Prof Vishal R. Shinde ,"Machine learning algorithm for stroke disease classification" in IJREAM, ISSN : 2454-9150, Volume 08, Issue 01, APR 2022 Special Issue.

[4] A. Begum, and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques", 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 342-345, 06 June, 2019.

[5] A. K. Aswathi, and A. Antony, "An Intelligent System for Thyroid Disease Classification and Diagnosis", 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018), pp. 1261-1264, 27 September, 2018.

[6] A. M. Amiri, and G. Armano, "Early Diagnosis of Heart Disease Using Classification And Regression Trees", In The 2013 International Joint Conference on Neural Networks, pp. 1-4, 09 January, 2014.

[7] M. R. N. Kousarrizi, F.Seiti, and M. Teshnehlab, "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification", International Journal of Electrical & Computer Sciences IJECS-IJENS, pp. 13-19, February, 2012.

[8] Fatemesh Saiti, Afsanesh Alavi Naini, Mahdi Aliyari Shroorehdeli, Mohammad Teshnehalb "Thyroid Disease Diagnosis Based on Gentic Algotitham Using PNN and SVM" 3rd International Conference on Bioinformatics and Biomadical (ICBBE) 2009.5163689.

DOI : 10.5281/zenodo.7857229