

# **NAAN MUDHALVAN PROJECT REPORT**

**On**

**“Thyroid disease classification using Machine Learning”**

Submitted in partial fulfilment of the requirements for the award of the award  
of the degree of

**Bachelor of Science**

**In**

**Computer Science**

Submitted by

**Team Leader**

Pon Prabha. S -20201371506123

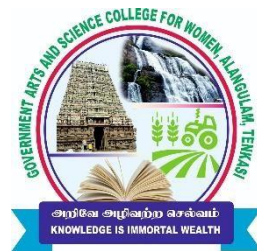
**Team Members**

Muthuselvi. M - 20201371506121

Sneha. S - 202013715061238

Suganthi. C - 20201371506142

Suthi varshini. M – 202013715044



Department of Computer Science,

Government Arts and science college for women, Alangulam,

Tenkasi – 627 851

April-2023

# **Table of Contents**

## **1. INTRODUCTION**

1.1 Overview

1.2 Purpose

## **2. Problem Definition & Design Thinking**

2.1 Empathy Map

2.2 Ideation & brainstorming map screenshot

## **3 RESULT**

Final findings (Output) of the project along with screenshots

## **4. ADVANTAGES**

## **5 APPLICATION**

## **6 CONCLUSION**

## **7 FUTURE SCOPE**

## **8 FUTURE WORK**

## **9 APPENDIX**

# **1. INTRODUCTION**

## **1.1 Overview**

Thyroid disease is increasingly common worldwide, affecting health condition in multiple countries. This disorder is characterized by affecting multiple functionalities of the human body, directly damaging the living condition of people who suffer it. In the case of India, it is estimated that there are approximately more than 42 million people who suffer a similar situation. For the national case, it is calculated the approximately 4% of the population suffers from some disorder hormonal, and it estimated that approximately worldwide about 200 million people have a thyroid disorder. Machine learning is not new to thyroid research. Artificial Neural Network(ANN) and decision trees have been used in Thyroid detection and diagnosis for nearly 20 years.

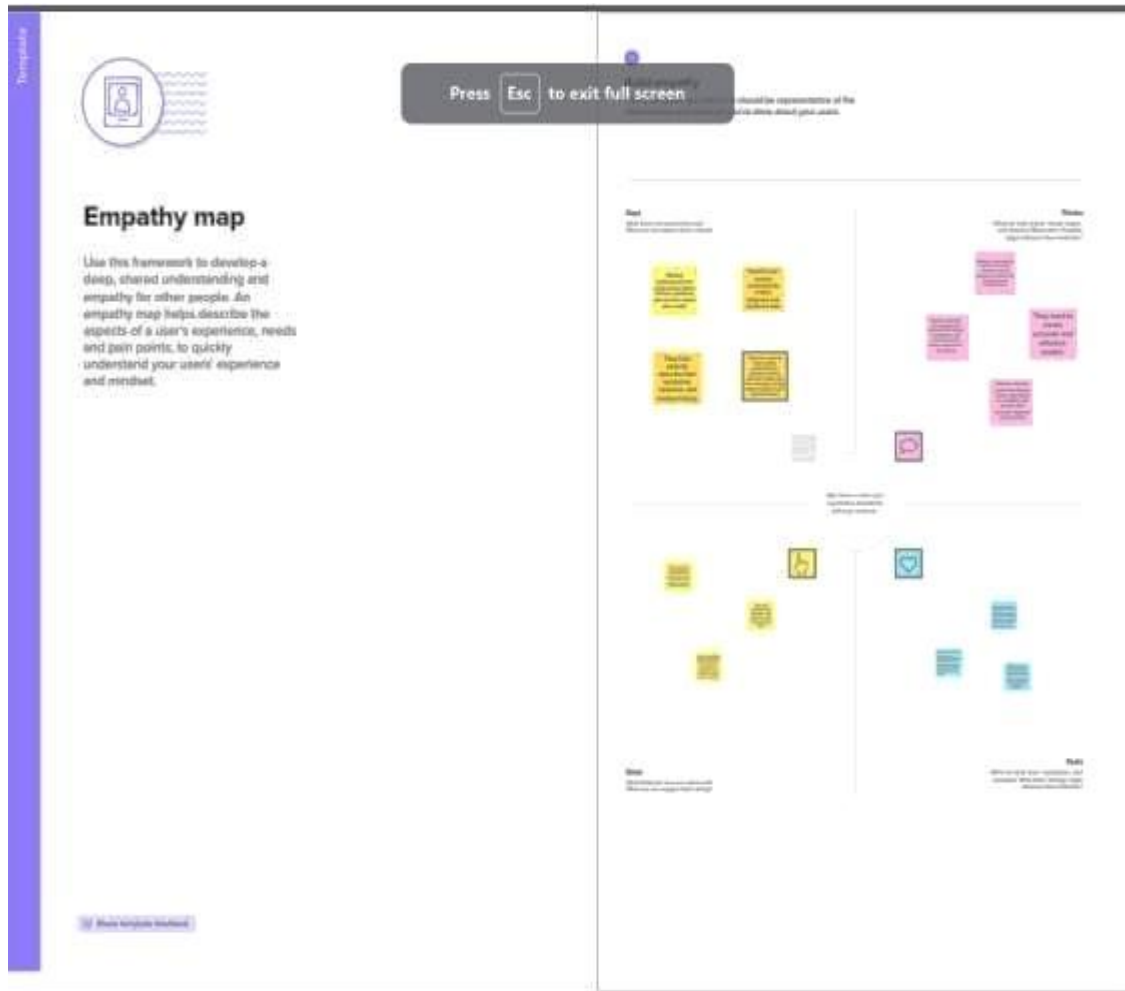
Machine learning is a tool to utilized for the creation and evaluation of algorithm that facilitate prediction, pattern recognition, and classification. ML is based on four steps. Collecting data, picking the model, training the model, testing the model. The relation between BC and ML is not recent, it had been used for decades to classify tumours and other malignancies, predict sequences of genes responsible of cancer and determine the prognostic.

## **1.2 Purpose**

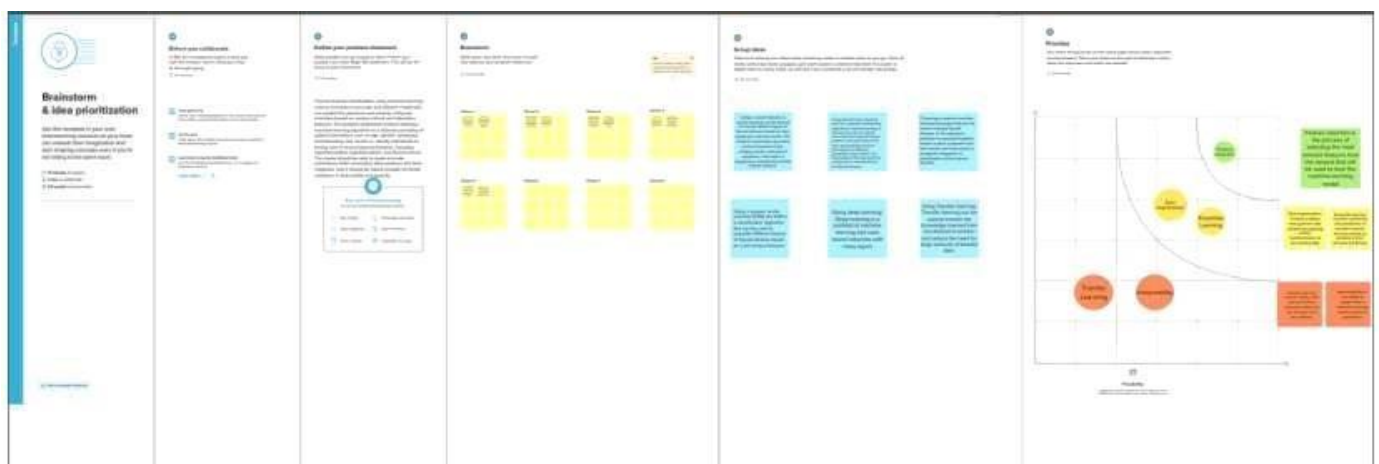
With the vast amount of data and information difficult to deal with, especially in the health system, machine learning algorithm and data mining techniques have an important role in dealing with data. In our study, we used machine learning algorithms with thyroid disease. The goal of the study is to categorize thyroid disease into three categories: hyperthyroidism, hypothyroidism and normal. Thyroid Disease Classification uses dataset extracted from the UC Irvine Machine Learning Repository. The hypothyroid dataset is used for the research and development department for experimental purposes. The dataset contains 3090 instances. In this 139 data comes under hypothyroid and 2941 data is negative cases. The classification of dataset is used to give better treatment, decision making, diagnose disease.

## 2. Problem Definition & Design Thinking

### 2.1 Empathy Map

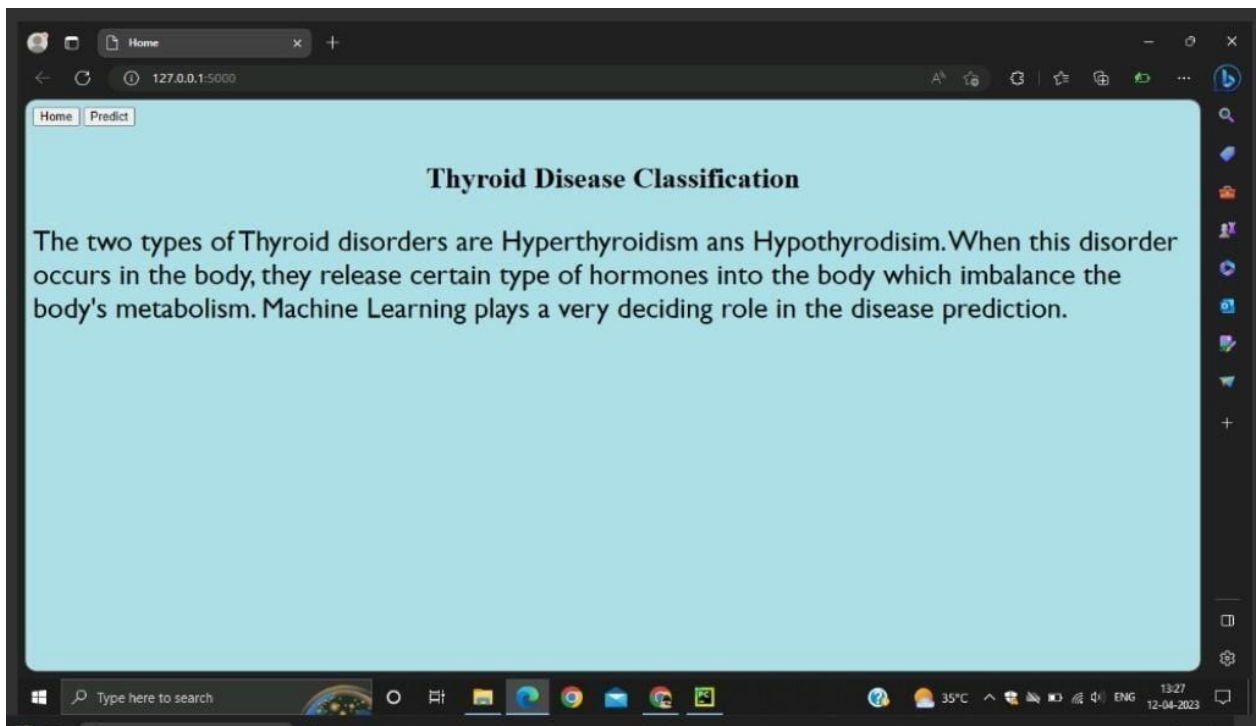


### 2.2 Ideation & Brainstorming map screenshot

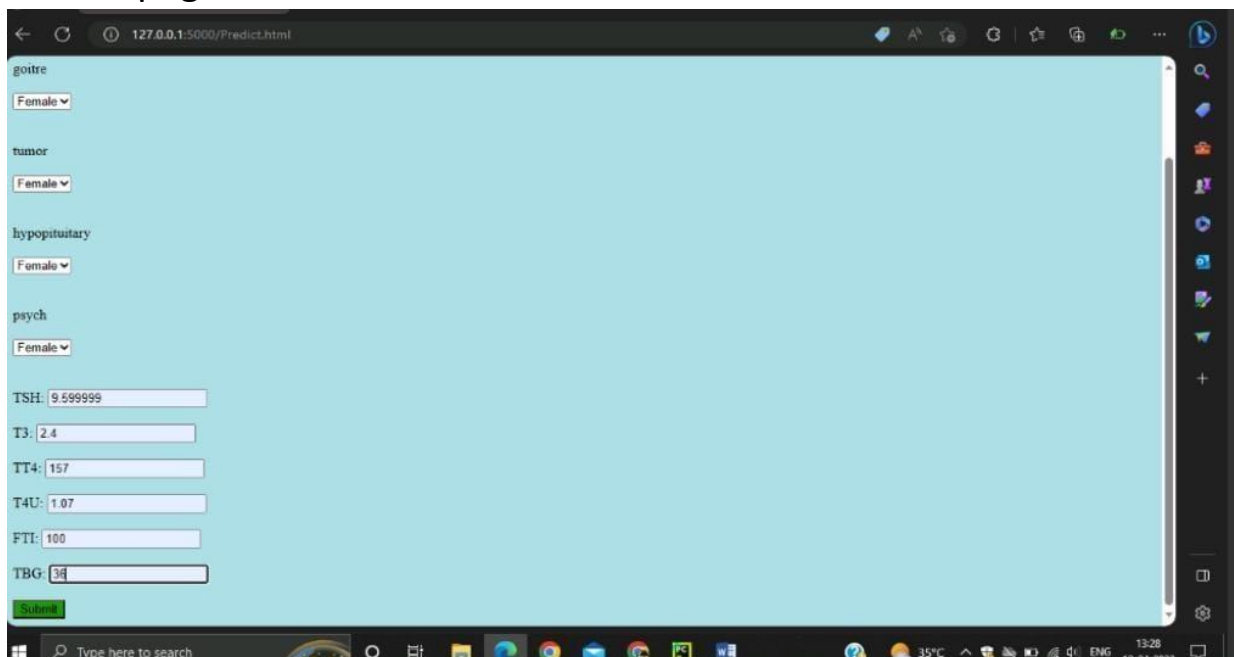


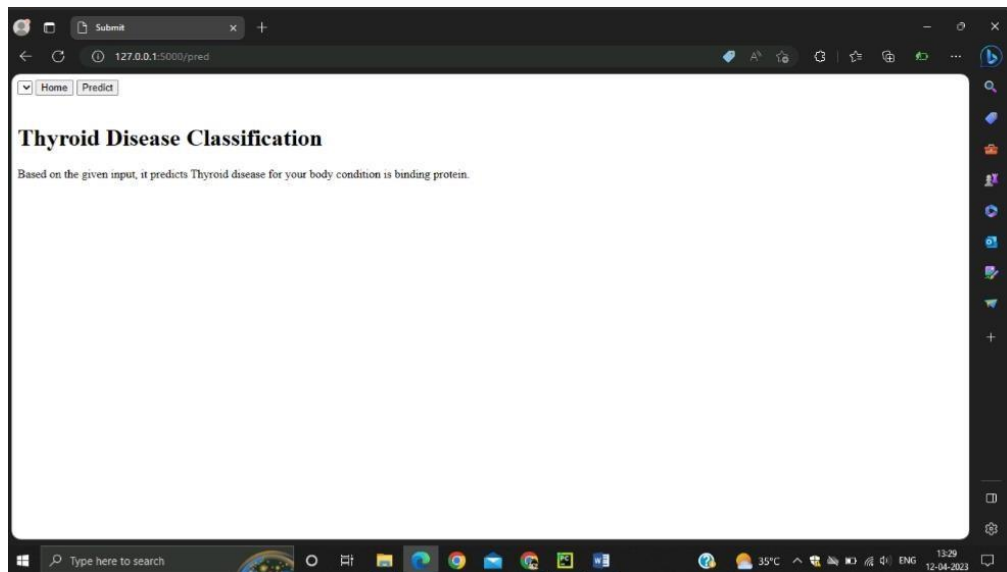
### 3. RESULT

Home page will look this



Predict page





## 4.ADVANTAGE

This study examines multiple machine learning, including support vector machine(SVM), randomforest , XGBoost, in order to enhance thyroid prediction accuracy and thereby identify thyroid problems. Thus, if abnormal thyroid hormone levels are recognized early enough, patients may be prescribed the appropriate medicine therapy. To summarize, the major contribution is as follows

- To estimate the likelihood of a better outlook when more data is used, as increasing thyroid prediction accuracy will enhance thyroid problem identification.
- To enhance the model's performance, a variety of pre-processing techniques like identifying and handling the missing values, encoding the categorical data etc. are applied.
- To evaluate the effectiveness of the employed machine learning algorithms, the accuracy, precision, recall, and F1-scores are examined.

## 5. APPLICATION

With the vast amount of data and information difficult to deal with, especially in the health system, machine learning algorithms and data mining techniques have an important role in dealing with data. In our study, we used machine learning algorithms with thyroid disease. The

goal of this study is to categorize thyroid disease into three categories: hyperthyroidism, hypothyroidism and normal, so we worked on this study using data from Iraqi people, some of whom have an overactive thyroid gland and others who have hypothyroidism, so we used all of the algorithms. Support vector machines, random forest classification of thyroid disease.

## **6.CONCLUSION**

Thyroid disease is one of the diseases that afflict the world's population, and the number of cases of this disease is increasing. Because of medical reports that show serious imbalances in thyroid diseases, our study deals with the classification of thyroid disease between hyperthyroidism and hypothyroidism. This disease was classified using algorithms. Machine learning showed us good results using several algorithms and was built in the form of two models. In the first model, all the characteristics consisting of 16 inputs and one output were taken, and the result of the accuracy of the Support vector algorithm, 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\_hypothyroid. Here we have included the increased accuracy of some algorithms, as well as the retention of the accuracy of others.

## **7.FUTURE SCOPE**

In the future we will test these algorithms on different datasets with a large number of instances so that we can confirm the conclusion we have made in this paper. It would be better to use real life datasets from different fields of science to exhaustively test these algorithms and compare their performance.

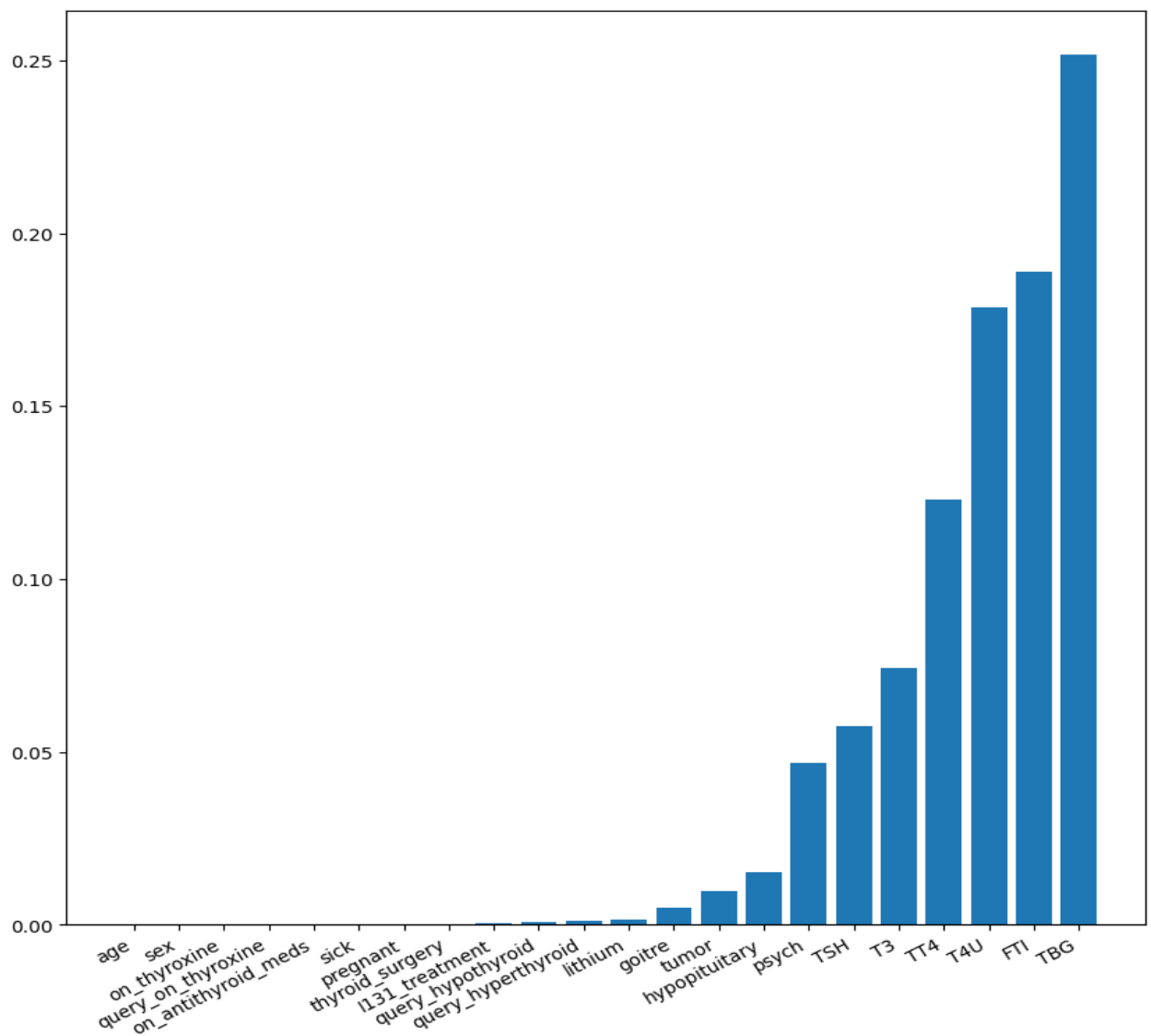
## **FUTURE WORK**

1. To make the model more practical, different features are also including as like reducing the inputs, applying more algorithms to make model are accurate.

2. A much larger dataset will help in getting a more accurate result.
3. Provide a Wonderful and Interactive Interface to the client for easy use.
4. Develop a new algorithm based on the existing algorithms which give the more accurate result in minimum time with more effectiveness.
5. A mobile application or desktop application could be developed to make the user experience easier.

## Appendix

Image of plotting importing feature:





### Correlation using Heatmap:

