Mini Project Report on

Thyroid Prediction System

using machine Learning

Submitted by

Name of Student	Class	Roll No.
1. **Saakshi Pawar**	TE-3	29
2. **Shweta Upadhyay**	TE-3	40
3. **Kirti Gawade**	TE-3	7
4. **Dhwani Kutmutia**	TE-3	17

Under the guidance of

Mr. Satish Bangal

DEPARTMENT OF COMPUTER ENGINEERING

SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE

CHEMBUR, MUMBAI - 400088.

2021-2022

www.shahandanchor.com Tel: 022 25580854

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Mahavir Education Trust Chowk, W.T. Patil Marg, Chembur, Mumbai 400 088

Affiliated to University of Mumbai, Approved by D.T.E. & A.I.C.T.E. ISO

Awarded accreditation for Computer & Information Technology Engineering by NBA

(for 3 years w.e.f. 1st July, 2019)

Certificate

This is to certify that the report of the mini project entitled

Thyroid Prediction System

Is a bonafide work
of

Name of Student	Class	Roll No.
1. **Saakshi Pawar**	TE-3	29
2. **Shweta Upadhyay**	TE-3	40
3. **Kirti Gawade**	TE-3	7
4**. Dhwani Kutmutia**	TE-3	17

submitted to the

UNIVERSITY OF MUMBAI

during semester V

in

COMPUTER ENGINEERING DEPARTMENT

Mr. Satish Bangal **(Prof. Uday Bhave)**

Guide I/c Head of Department

Approval for Mini Project Report for T. E. Semester V

This mini project report entitled "**Thyroid Prediction System"** by **Saakshi Pawar, Shweta Upadhyay, Kirti Gawade** and **Dhwani Kutmutia** is approved for the partial fulfillment of the requirement for the completion of Semester V.

Name and Sign of Internal Examiner _____

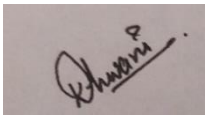
Name and Sign of External Examiner _____

Date:29-10-2021

Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of Student	Class	Roll No.	Signature
1. Saakshi Pawar	TE-3	29	
2. Shweta Upadhyay	TE-3	40	
3. Kirti Gawade	TE-3	7	
4. Dhwani Kutmutia	TE-3	17	

Date:29-10-2021

Place: Mumbai

Attendance Certificate

29 October 2021

To,
The Principal
Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai-88

Subject: Confirmation of Attendance

Respected Sir,

This is to certify that Third year (TE) students
1.Saakshi Pawar
2.Shweta Upadhyay
3.Kirti Gawade
4. Dhwani Kutmutia
have duly attended the sessions on the day allotted to them during the period from **July 2021** to **November 2021** for performing the Mini Project titled **Thyroid Prediction System.** .

They were punctual and regular in their attendance. Following is the detailed record of the student's attendance.

Attendance Record:

Date	**Saakshi Pawar**	**Shweta Upadhyay**	**Kirti Gawade**	**Dhwani Kutmutia**
	Present/Absent	**Present/Absent**	**Present/Absent**	**Present/Absent**
4/8/2021	Present	Present	Present	Present
11/8/2021	Present	Present	Present	Present
18//8/2021	Present	Present	Present	Present
21/8/2021	Present	Present	Present	Present
25/8/2021	Present	Present	Present	Present
8/9/2021	Present	Present	Present	Present
15/9/2021	Present	Present	Present	Present
22/9/2021	Present	Present	Present	Present
13/10/2021	Present	Present	Present	Present
22/10/2021	Present	Present	Present	Present

Signature and Name of Internal Guide
Mr. Satish Bangal

ABSTRACT

Thyroid gland is one of the most important organs in our body. The thyroid hormones are capable in controlling the metabolism. Machin learning plays a key role in the prediction of any kind of disease. Here, we have proposed to predict two classes of thyroid disorders using machine learning algorithms. The two classes are Hyperthyroidism and Hypothyroidism. Hyperthyroidism refers to excess secretion of thyroid hormones in the body whereas Hypothyroidism refers to less secretion of thyroid hormones in the body. For the prediction of thyroid disorders, we are going to using thyroid datasets.

.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Thyroid, a small gland located in anterior part of the neck is responsible for creating & secreting hormones that control Metabolism and protein synthesis. So thyroid is an essential part that serves many vital tasks in human body. Thyroid disease occurs when thyroid gland fails to produce right number of thyroid hormones. Most common thyroid diseases are hypothyroidism, hyperthyroidism, thyroiditis, etc. If your body makes too much thyroid hormone you can develop a condition called hyperthyroidism. In this condition, you feel tired and lose weight without trying and even your heart beats faster. If your body makes too little thyroid hormone then it is called hypothyroidism. In this, you might gain weight and you may be unable to tolerate low temperatures.

Machine Learning is an application of artificial intelligence that provides system the ability to learn and improve from experience without being programmed. It allows computer program to access data & use it to learn with the help of algorithms.

Thyroid diseases are increasing in magnitude & spreading all over the world. This disorder is primarily taking place at between age of 17-54. Thyroid detection is often a task because it presents symptoms that overlap with other medical conditions as well. Hence, early detection & accurate diagnosis can keep body balanced and save many lives. To achieve that, there is need to use ML algorithms to train systems and give highest accuracy.

CHAPTER 2

LITERATURE SURVEY

A brief literature review of 8 papers related to Thyroid Prediction System is done as mentioned below

Sr No	Paper Name / Year	Authors	Findings	Research Gaps
1	Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques Year: 2020	Priyanka Duggal, Shipra Shukla	Classification techniques like Naïve Bayes, Random Forest and Support Vector Machine were used to predict 4 types of thyroid disorders.	Data along with the attributes used for predicting should have been mentioned.
2	A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis. Year: 2020	Amulya R. Rao, B.S.Renuka	In the proposed system, Classification predictive modelling and thyroid dataset from the Kaggle machine learning website is used to predict the thyroid disease at an early stage.	There is a need to include the detailed use of classification techniques for better understanding.

3	Computer Aided Diagnosis of Thyroid Disease Using Machine Learning Algorithms Year: 2020	Md. Asfi-Ar-Raihan Asif, Mirza Muntasir Nishat, Fahim Faisal, Md. Fahim Shikder, Mahmudul Hasan Udoy, Rezuanur Rahman Dip and Ragib Ahsan	In this research, authors conducted an efficient data processing technique and investigated various machine learning algorithms for an early prediction of thyroid disease.	Algorithms are not explained in detail only an overview is given.
4	Application of Deep Learning in the Prediction of Benign and Malignant Thyroid Nodules on Ultrasound Images Year: 2020	Yinghui lu , yi yang, and wan chen	This paper discusses about the prediction of malignant thyroid nodules using ultrasound images with the help of deep learning algorithms. The efficiency of the detection model in this experiment for thyroid nodules is improved compared with the original model.	Size of sample images should be larger to increase accuracy in detection of malignant and benign thyroid nodules.
5	Interactive Thyroid Disease Prediction System Using Machine Learning Technique Year:2018	Anita Tyagi, Ritika Mehra, Aditya Saxena	The proposed system is implemented using Artificial Neural Network, support vector machine, K-NN and decision trees, which were used to predict the estimated risk on a patient's chance of obtaining thyroid disease.	There is a need to design an improved system by including the factors like age group, heredity, antibodies etc.

6	Disease Risk Prediction by Using Convolutional Neural Network Year: 2018	Sayali Ambekar,Rashmi Phalnikar	CNN-UDRP algorithm using structured data was used to predict the risk of heart disease. Naïve Bayes, KNN were also compared with CNN.	Algorithms are not explained in detail only overview is given.
7	Disease Prediction Using Machine Learning Techniques Year: 2021	Roop Chandrika Mallela, Reddy Lakshmi Bhavani & Dr. B. Ankayarkanni.	The project presented the technique of predicting the disease based on the symptoms of an individual patient. Chatbots which are prepared using ML which are used to predict diseases.	Result obtained from Naïve Bayes technique is not 100% accurate.
8	Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis Year: 2016	Prof. Dhomse Kanchan B. & Mr. Mahale Kishor M.	Diabetes prediction is showed using WEKA data mining tool by using and comparing SVM, Naive Bayes and Decision tree techniques (ML Techniques), the results were obtained based on time taken to build model, correctly classified instances, error and ROC area.	Algorithms of different techniques are mentioned but the main algorithm / flow chart of this project is not mentioned.

CHAPTER 3

PROBLEM STATEMENT

3.1 Problem Definition

More than half of the Indian population suffers from undiagnosed or misdiagnosed thyroid diseases also the symptoms of this disease often vary from person to person and are non-specific, so a correct diagnosis can easily be missed or misdiagnosed for irrelevant issues. so, our main aim is finding an accurate solution to this problem hence we are going to make a Thyroid Prediction System using Machine Learning to predict 2 types of thyroid disorders at an early stage with accuracy. This system will help people detect thyroid with the help of data set.

3.2 Objectives

The main objectives of our project are:

1. To build a platform for users to predict two classes of thyroid disorders with accurate result.
2. To build a user friendly and cost-free platform to use.
3. To build a platform to be able to predict from large datasets.

CHAPTER 4

PROJECT DESIGN

4.1 System Block Diagram

We have seen the aim of our project and have discussed the why and what's. Now it is time to check and see how the project works. Let us start with the block diagram first.

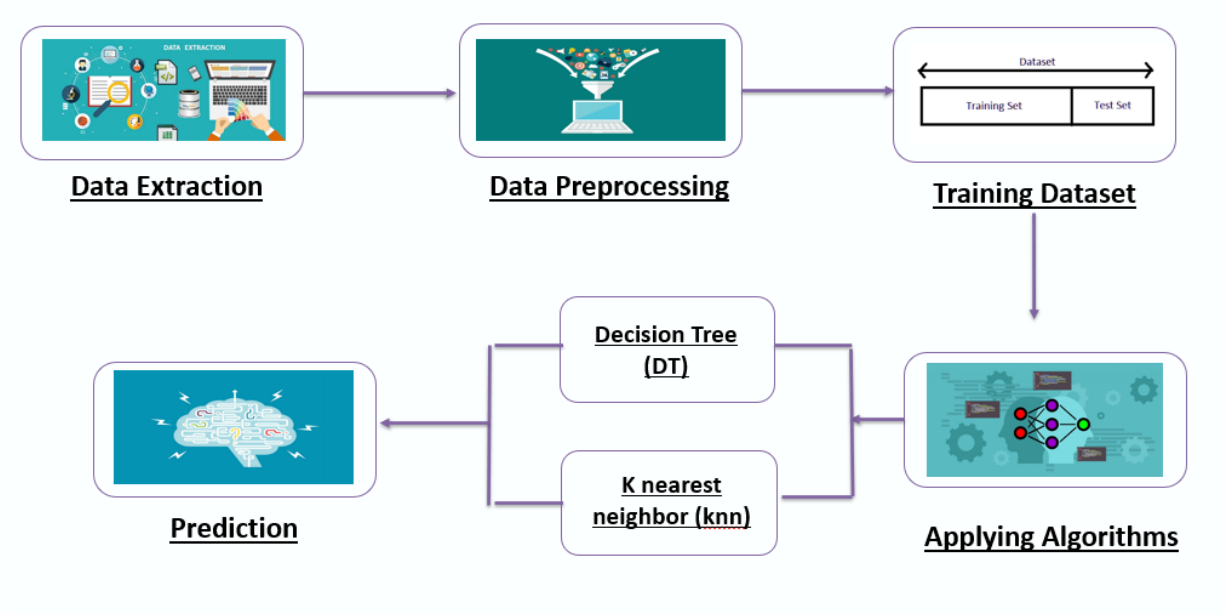


Fig 4.1.1 System Block Diagram

As observed from the block diagram, the first step is data extraction, this is where our project starts. We are going to collect a large amount of data set for thyroid disease to train our model. The more data we have, the more accurate result we will get. Now this data will have several attributes, several values, which brings us to our next step that is data preprocessing. where we will just check the names of the attributes, if they have constraints, any missing values, etc. once the data is processed properly the next step is to split the data into training set and testing set.

The next thing is algorithm selection. Selecting the right algorithm for the dataset is the most essential task in every Machine Learning model. We are applying 2 algorithms Decision Tree and KNN. We will then train the data with these models. The next step is selecting the model giving most accuracy. So, once we are done selecting that algorithm, we will use that algorithm in predicting whether the user has thyroid disorder or not.

4.2 ALGORITHM

4.2.1 DECISION TREE ALGORITHM

Decision tree induction is the learning of decision trees from class-labelled training tuples. A decision tree is a flowchart-like tree structure,

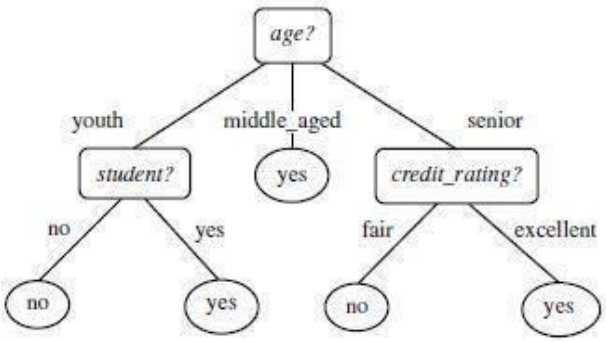


Fig 4.2.1.1 Decision Tree problem

- Decision tree induction is a non-parametric approach for building classification models.
- Finding an optimal decision tree is an NP-complete problem
- Techniques developed for constructing decision trees are computationally inexpensive, making it possible to construct models even when the training set size is very large.
- Decision trees, especially smaller-sized trees, are relatively easy to interpret.
- Decision tree provide an expressive representation for learning discrete- valued functions.

- Decision tree algorithms are quite robust to the presence of noise, especially when methods for avoiding overfitting.

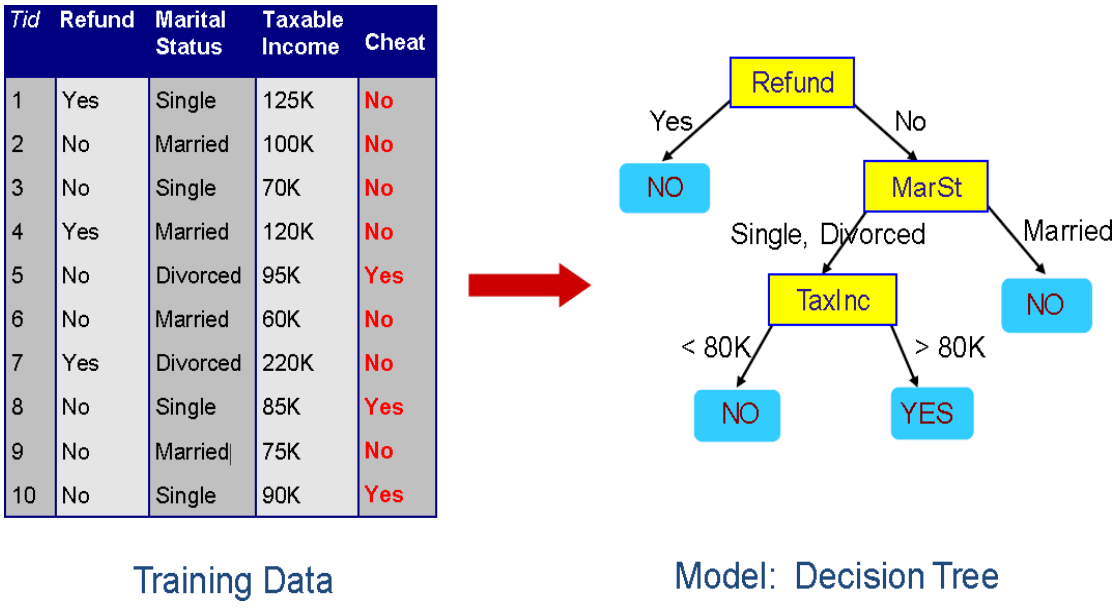


Fig 4.2.1.2 Decision Tree Example

- The presence of redundant attributes does not adversely affect the accuracy of decision tree.
- The construction of decision tree classifiers does not require any domain knowledge or parameter setting, and therefore I appropriate for exploratory knowledge discovery. Decision trees can handle high dimensional data.
- Their representation of acquired knowledge in tree form is intuitive and generally easy to assimilate by humans.
- The learning and classification steps of decision tree induction are simple and fast.
- In general, decision tree classifiers have good accuracy.
- Decision tree induction algorithm shave been used for classification in many application areas, such as medicine, manufacturing and production, financial analysis, astronomy, and molecular biology.

Algorithm: Generate_decision_tree. Generate a decision tree from the training tuples of data partition *D*.

Input:

- Data partition, *D*, which is a set of training tuples and their associated class labels;
- *attribute_list*, the set of candidate attributes;
- *Attribute_selection_method*, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a *splitting_attribute* and, possibly, either a *split point* or *splitting subset*.

Output: A decision tree.

Method:

```
(1)  create a node N;
(2)  if tuples in D are all of the same class, C then
(3)      return N as a leaf node labeled with the class C;
(4)  if attribute_list is empty then
(5)      return N as a leaf node labeled with the majority class in D; // majority voting
(6)  apply Attribute_selection_method(D, attribute_list) to find the "best" splitting_criterion;
(7)  label node N with splitting_criterion;
(8)  if splitting_attribute is discrete-valued and
         multiway splits allowed then // not restricted to binary trees
(9)      attribute_list ← attribute_list − splitting_attribute; // remove splitting_attribute
(10) for each outcome j of splitting_criterion
     // partition the tuples and grow subtrees for each partition
(11)     let Dj be the set of data tuples in D satisfying outcome j; // a partition
(12)     if Dj is empty then
(13)         attach a leaf labeled with the majority class in D to node N;
(14)     else attach the node returned by Generate_decision_tree(Dj, attribute_list) to node N;
     endfor
(15) return N;
```

Fig 4.2.1.3 Decision Tree Algorithm

4.2.2 KNN

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- The number of nearest neighbours to a new unknown variable that has to be predicted or classified is denoted by symbol 'K'

Steps to implement KNN Algorithm:

- Data Pre-processing step
- Fitting the K-NN algorithm to the Training set
- Predicting the test result
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result.

KNN Algorithm:

- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.

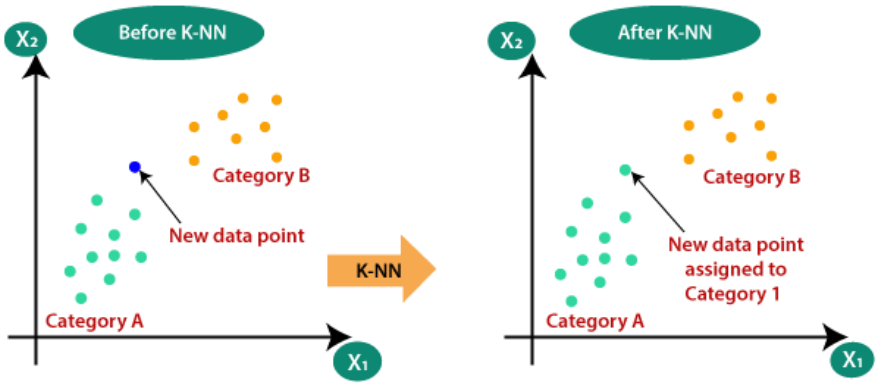


Fig 4.2.2.1 KNN Algorithm Steps

The following two properties would define KNN well −

- **Lazy learning algorithm** − KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.
- **Non-parametric learning algorithm** − KNN is also a non-parametric learning algorithm because it doesn't assume anything about the underlying data.

KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

CHAPTER 5

IMPLEMENTATION DETAILS

5.1 DESCRIPTION

The project Thyroid Prediction System is implemented using python (flask) in Back-End. JupyterNotebook is used for editing the code and running as well. The Front-end interface of this project is done using HTML, CSS. We are using 2 algorithms namely DT and KNN based on which we'll select the algorithm giving more accuracy. We found that DT gave more accuracy than KNN as seen in the below figure.

```
KNN:

Train Score:0.8747871495061869
 Test Score:0.8656377666817976
------------------------------------------------------------

Decision Tree:

Train Score:0.9125893972074015
 Test Score:0.899682251475261
------------------------------------------------------------
```

Fig 5.1.1 Accuracy Result of DT and KNN

After finalizing DT we will then test our data against it and use it for further predictions. We'll go to Home page where the user needs to the enter the values of TSH, TT4, T4U, T3, Age and needs to select the symptoms from given drop-down menu, for more accurate result the user needs to enter all the given symptoms, then thesystem will provide the accurate result. 2 classes of Thyroid disease ie Hyperthyroidism and Hypothyroidism can be predicted.

5.2 SNAPSHOTS

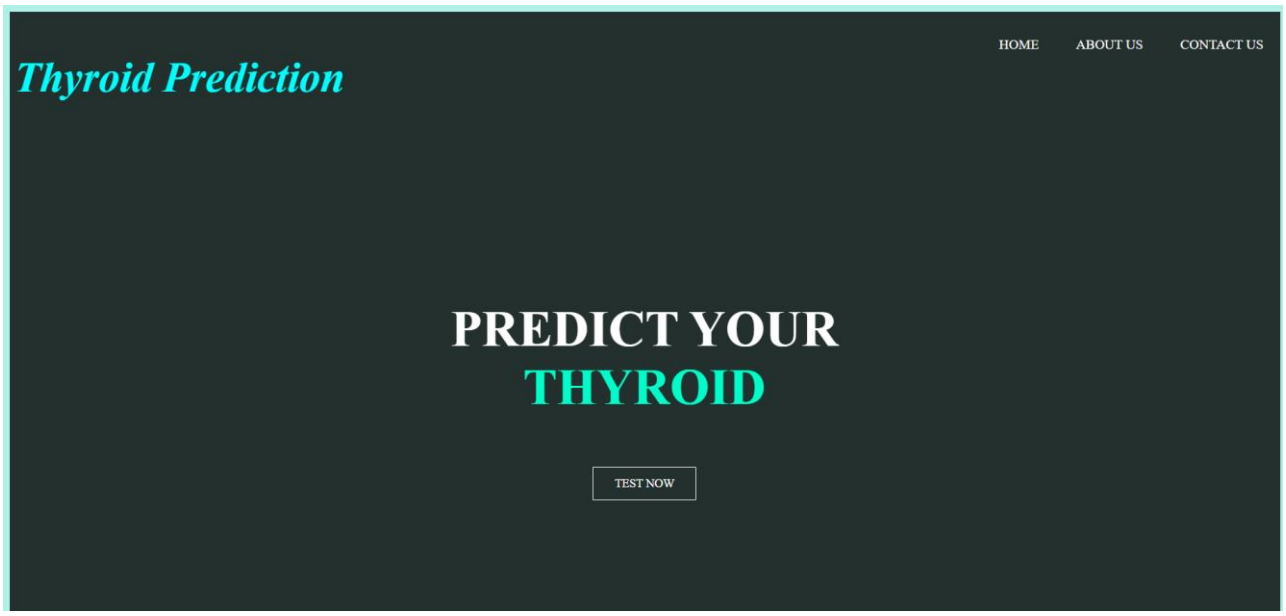


Fig. 5.2.1 Home Page

Fig. 5.2.2 Thyroid prediction

CHAPTER 6

RESULT AND ANALYSIS

Test case 1: Result is Negative

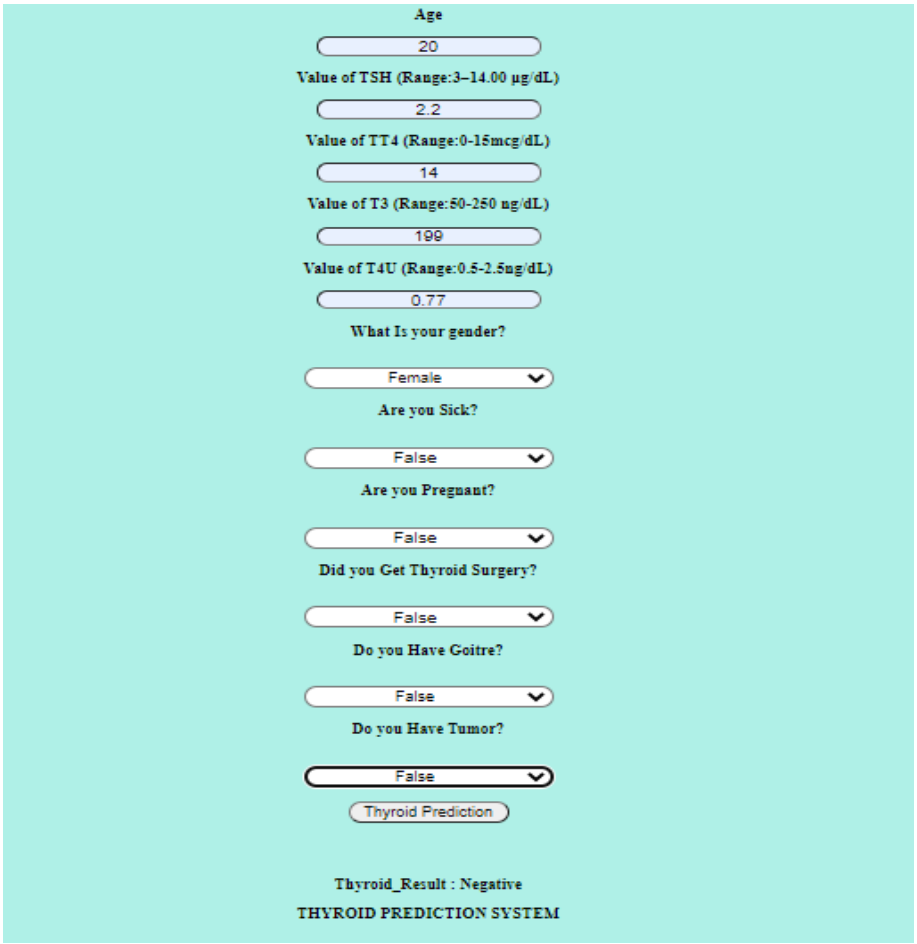


Fig 6.1 Thyroid Negative

Test Case 2: Result is Hyperthyroid

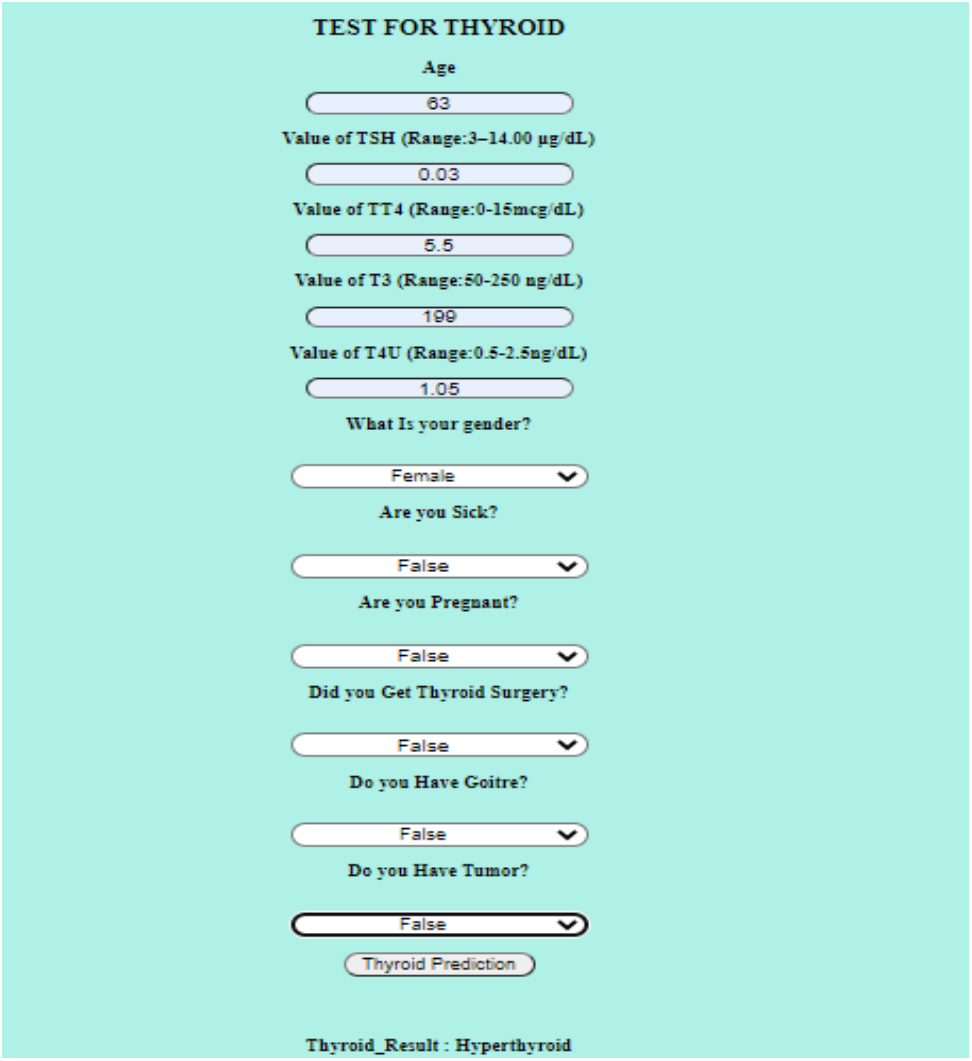


Fig 6.2 Hyperthyroid

Test Case 3: Result is Hypothyroid

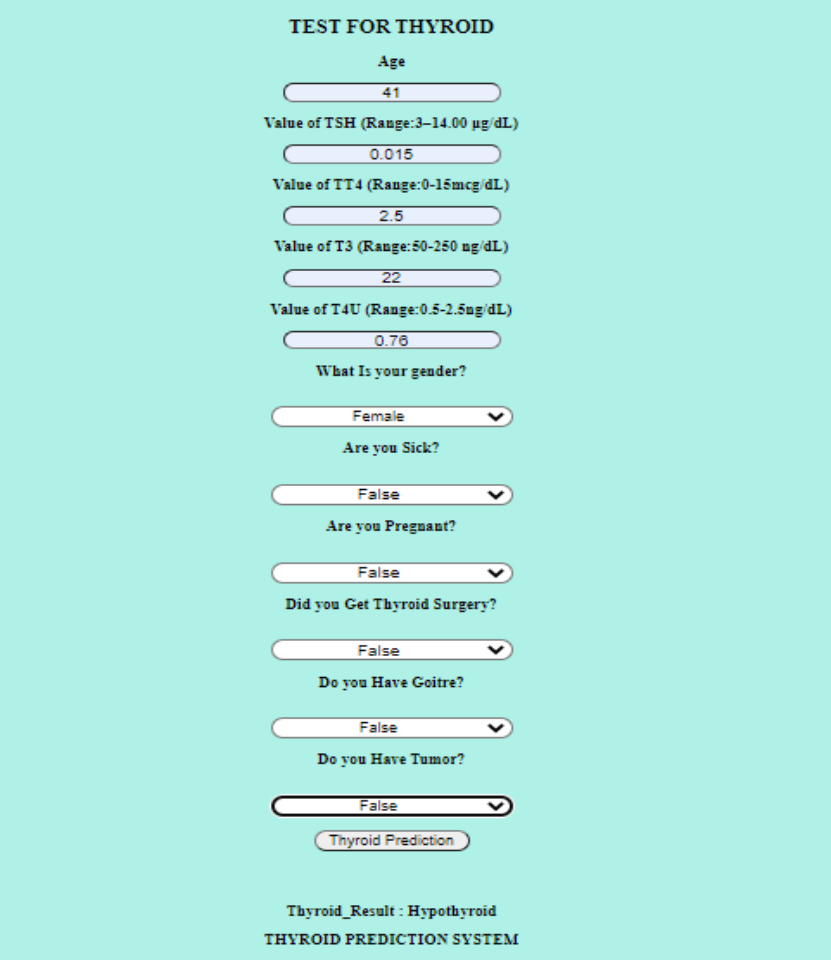


Fig 6.3 Hypothyroid

Test Case 4: Result is Sick

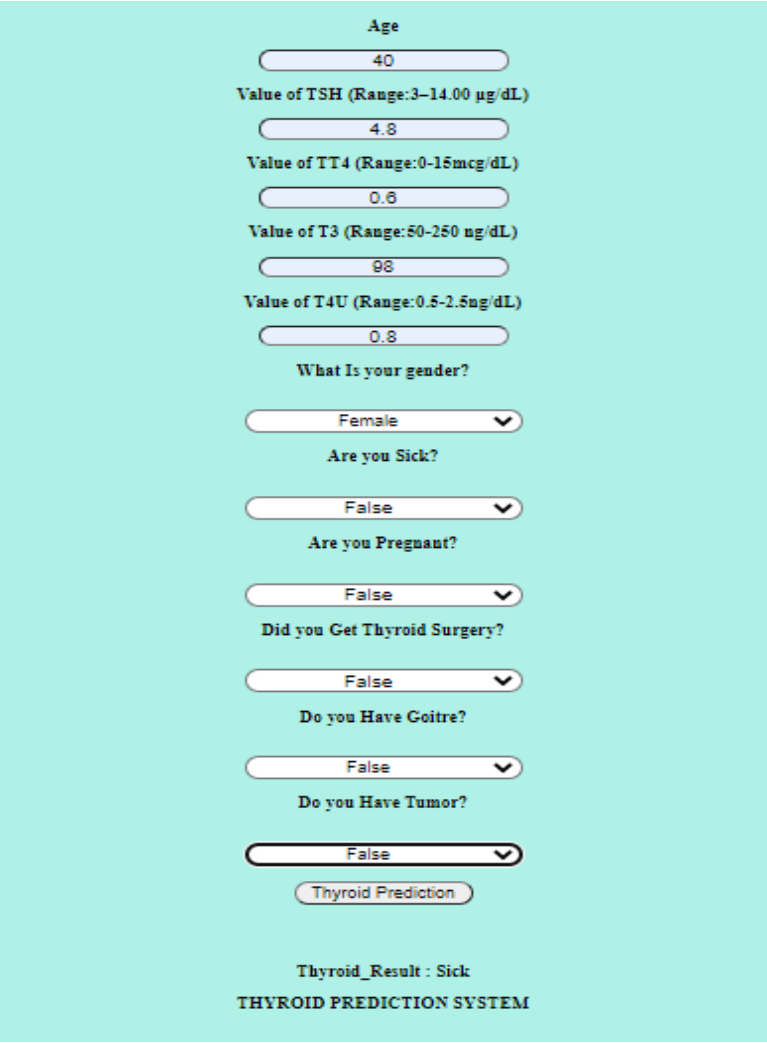


Fig 6.4 Sick

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 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 two different algorithms. The result of the accuracy of the decision tree algorithm was approx. 91 % which is the highest accuracy among the other algorithm which is K nearest neighbor whose accuracy was approx. 86%. And then we used decision tree to predict the result of our system as it gives more accurate result than knn. Hence, the system can predict 2 different thyroid class: hyperthyroid and hypothyroid and even the system can predict whether the person is sick or having thyroid as negative.

7.2 FUTURE SCOPE

The scope of our project involves implementation of a view doctors tab so that after the prediction the users can check the details of the doctors and contact accordingly. Also, a chatbot feature can be incorporated in the system to solve the queries of the users. For checking the overall predictions made in a day there can be a performance analysis tab as well. Lastly, a feedback feature for the user can be executed so that we can make the system more effective and convenient to use.

CHAPTER 8

REFERENCES

[1] Priyanka Duggal, Shipra Shukla, "Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques" accessed in 2020.

[2] Yinghui lu , yi yang , and wan chen, "Application of Deep Learning in the Prediction of Benign and Malignant Thyroid Nodules on Ultrasound Images" accessed in 2020

[3] Anita Tyagi, Ritika Mehra, Aditya Saxena, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique" accessed in 2018

[4] Roop Chandrika Mallela, Reddy Lakshmi Bhavani, Dr. B. Ankayarkanni, " Disease Prediction Using Machine Learning Techniques" accessed in 2021.

[5] Sayali Ambekar, Rashmi Phalnikar, " Disease Risk Prediction by Using Convolutional Neural Network" accessed in 2018.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be impossible without the mention of the people who made it possible, whose constant guidance and encouragement crowned our efforts with success.

We take this opportunity to express my profound gratitude to **Dr. Bhavesh Patel**, Principal SAKEC, for his constant support and encouragement.

We are grateful to **Ms. Rekha More**, Senior Assistant Professor, for her unfailing encouragementand suggestions, given to me in the course of my project work.

We would also like to thank **Mr. Uday Bhave,** Associate Professor and Head, Department of Computer Science and Engineering, for his constant support.

We express our gratitude to **Mr. Satish Bangal,** Assistant Professor, our project guide, for constantly monitoring the development of the project and setting up precise deadlines. His valuable suggestions were the motivating factors in completing the work.

Finally, a note of thanks to the teaching and non-teaching staff of Dept of Computer Engineering, for their cooperation extended to me, and my friends, who helped us directly or indirectly in the course of the project work.

SAAKSHI PAWAR

SHWETA UPADHYAY

KIRTI GAWADE

DHWANI KUTMUTIA