**THYROID DISEASE CLASSIFICATION USING MACHINE LEARNING**

Submitted in partial fulfillment of requirement for the award of the Degree

**Bachelor of Computer Science**

In the faculty of Computer Science of Bharathiar University, Coimbatore

Submitted by

TEAM ID:NM2023TMID17826

ANUKAVIYA.K

SIVADHARSHINI.R

KAVITHA.M

ABINAYA M

Under the guidance of

**Mrs.V.PUNITHAVATHI MCA,M.Phil,DOA.,**

Guest lecturer , Department of Computer Science



**DEPARTMENT OF COMPUTER SCIENCE**

**L.R.G GOVERNMENT ARTS COLLEGE FOR WOMEN**

**(Affiliated To Bharathiar University)**

**TIRUPUR-4**

**APRIL-2023**

# LRG GOVERNMENT ARTS COLLEGE

**NAAN MUDHALVAN PROJECT WORK**

**(AFFILIATED TO BHARATHIAR UNIVERSITY) TIRUPUR-641602**

## TITLE : Thyroid Disease Classification Using ML

This is to certify that this is a bonafide record of work done by the above

students of III B.Sc (CS) Degree **NAAN MUDHALVAN PROJECT** during the year ……….

Submitted for the Naan Mudhalvan project work held

### on………….20

**CLASS TUTOR HEAD OF DEPARTMENT**



|  |  |
| --- | --- |
| **CHAPTER NO** | **CONTENTS** |
| 1 | INTRODUCTION |
| 2 | PROBLEM SELECTION |
| 3 | IDEATION |
| 4 | REQUIREMENT ANALYSIS |
| 5 | INPUT DESIGN  5.1. INPUT DESIGN DESCRIPTION |
| 6 | OUTPUT DESIGN  6.1.OUTPUT DESIGN DESCRIPTION |
| 7 | DESCRIPTION OF MODULES |
| 8 | PROJECT PALNNING PHASE |
| 9 | PROJECT DEVELOPMENT PHASE |
| 10 | CONCLUSION |
| 11 | APPENDICES  A.TECHNICAL FLOW  B.SAMPLE INPUT  C.SAMPLE OUTPUT  D.SAMPLE CODING |





The thyroid gland is a small organ that’s located in the front of the neck,

wrapped around the windpipe (trachea). It’s shaped like a butterfly, smaller in the middle with two wide wings that extend around the side of your throat. The thyroid is a gland. The two main types of thyroid disease are hypothyroidism and hyperthyroidism. Both conditions can be caused by other diseases that impact the way the thyroid gland works.

Thyroid disease can affect anyone — men, women, infants, teenagers and the elderly. It can be present at birth (typically hypothyroidism) and it can develop as you age (often after menopause in women).

Symptoms of an overactive thyroid (hyperthyroidism) can include

Experiencing anxiety, irritability and nervousness, Having trouble sleeping, Losing weight, Having an enlarged thyroid gland or a [goiter,](https://my.clevelandclinic.org/health/diseases/12625-goiter) Having muscle weakness and tremors, Experiencing irregular menstrual periods or having your menstrual cycle stop, Feeling sensitive to heat and Having vision problems or eye irritation.

Symptoms of an underactive thyroid (hypothyroidism) can include Feeling tired ([fatigue)](https://my.clevelandclinic.org/health/symptoms/21206-fatigue), Gaining weight, Experiencing forgetfulness, Having frequent and [heavy menstrual periods,](https://my.clevelandclinic.org/health/diseases/17734-menorrhagia-heavy-menstrual-bleeding) Having dry and coarse hair, Having a hoarse voice and Experiencing an intolerance to cold temperatures.

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 disease prediction. Machine Learning algorithms (SVM, Random Forest Classifier, K-NN, logistic regression and DT Classifier are used to predict the patient's risk of getting thyroid disease. The web app is created to get data from users to predict the type of disease.



Predicting thyroid disease using machine learning is an interesting and important project that can have significant implications in the field of healthcare. Here are some steps you can follow to select a suitable project:

1. Research existing work: Before starting any project, it's important to research existing work in the field. Look for similar projects that have been done before and see what approaches were used. This will help you understand the current state-of-the-art and identify any gaps in the literature that your project can address.
2. Define the problem: Once you have a good understanding of the current research, define the problem you want to solve. For example, you could focus on predicting the risk of developing thyroid disease in a particular population or predicting the progression of the disease in patients who have already been diagnosed.
3. Gather data: In order to train a machine learning model, you will need a dataset. Look for publicly available datasets or consider collecting your own data. Make sure that the data you use is of high quality and is representative of the problem you are trying to solve.
4. Train and evaluate the model: Once you have chosen an algorithm, train the model on your dataset and evaluate its performance using appropriate metrics such as accuracy, precision, recall, and F1 score. If the model's performance is not satisfactory, consider tweaking the algorithm or using a different one.
5. Deploy the model: Once you have a model that performs well, deploy it in a real-world setting. This could involve integrating it into an electronic health.

|  |  |
| --- | --- |
| i. | Collect Data: Collect data from various sources, including medical records, patient information, and clinical studies. Ensure that the data is in a structured format and that it contains all the necessary features for building a model. |
| ii. | Preprocess the Data: Once you have the data, preprocess it by cleaning, normalizing, and transforming it into a format that can be used by a machine learning model. |
| iii. | Feature Selection: Choose the relevant features that will be used to predict the thyroid disease. These features can include patient demographics, lab test results, and medical history. |
| iv. | Choose a Model: Choose an appropriate machine learning algorithm that will be used for prediction. There are various algorithms available, including decision trees, random forests, and support vector machines. |

|  |  |
| --- | --- |
| v. | Train the Model: Use the preprocessed data and the chosen algorithm to train the model. Split the data into training and testing sets to evaluate the model's accuracy. |
| vi. | Evaluate the Model: Evaluate the model's performance using different metrics such as accuracy, precision, recall, and F1score. |
| vii. | Deploy the Model: Finally, deploy the model into a production environment where it can be used to predict the thyroid disease of new patients. |



**Data collection:**

Gather a large dataset of patient information, including demographics, medical history, symptoms, lab test results, and imaging data (if available). The dataset should include both positive and negative cases of thyroid disease to ensure the model is trained on a balanced dataset.

**Data preprocessing:**

Clean the data, remove any missing or irrelevant information, and encode categorical variables.

**Feature select:**

Identify the most important features that contribute to the prediction of thyroid disease.

**Algorithm select:**

Select the appropriate machine learning algorithm for the task, such as logistic regression, decision trees, random forests, or support vector machines.

**Model training:**

Split the dataset into training and validation sets, train the model on the training set, and evaluate the performance on the validation set. Adjust the model hyperparameters to optimize performance.

**Model evaluation:**

Evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. Use cross-validation to ensure the model's generalizability to new data.

**Model deployment:**

Deploy the trained model in a user-friendly interface, such as a web application or mobile app, to enable healthcare professionals to make accurate predictions of thyroid disease in their patients.

**Maintenance and updates:**

Regularly update the model with new data and retrain as necessary to ensure the accuracy of the predictions.



The input design is the process of entering data to the system. The input design goal is to enter to the computer as accurate as possible. Here inputs are designed effectively so that errors made by the operation are minimized.

The inputs to the system have been designed in such a way that manual forms and the inputs are coordinated where the data elements are common to the sources document and to the input. The input is acceptable and understandable by the users who are using it.

Input design is the process of converting user-originated inputs to a computer-based format input data are collected and organized into group of similar data. Once identified, appropriate input media are selected for processing.

Input design means the physical and performance requirements of a device that are used as a basis for device design. Input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc



### HOME

In the home page a simple concept of thyroid is displayed and it shows a predict option to get to the predict page.

### PREDICT

In this page it asks for input of certain data that is necessary to predict the thyroid type it also shows a submit button which redirects it to the submit page.



A design output is a drawing or specification or manufacturing instruction. Design outputs describe all the components, parts, and pieces that go into your device. Design outputs describe all assemblies and subassemblies of product.

Output design is the process of converting data into hard copy that is understood by all. The various outputs have been in such a way that they represent the same format that the office and management used to.

Computer output is the most important and direct source of information to the user. Efficient, intelligible output design should improve the systems relationships with the user and help in decision making. A major form of out is the hardcopy from the printer. Output requirements are designed during system analysis.



SUBMIT

In this page after enter necessary data in predict page then click the submit button. It display the result of that entered data like (Normal, hypothyroid and hyperthyroid).



Modules are unit of code written in access basic language.

* HOME
* PREDICT
* SUBMIT

HOME

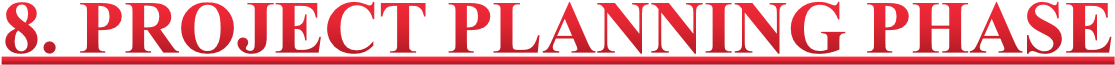
* In the home page a simple concept of thyroid is displayed and it shows a predict option to get to the predict page.

PREDICT

* In this page it asks for input of certain data that is necessary to predict the thyroid type.

SUBMIT

* In this page the result of the predicted thyroid type is displayed like whether the user is normal, hyperthyroid or hypothyroid.



**Define the problem:**

Define the problem you want to solve. For example, you maywant to build a model that can accurately predict the risk of thyroid disease in patients based on certain clinical and demographic features.

**Gather data**:

Identify the data sources that can be used to build the machine learning model. This may involve gathering data from electronic health records, medical imaging, or patient surveys.

**Choose a machine learning algorithm:**

There are several machine learning algorithms that can be used to build a predictive model. You will need to choose an algorithm that is appropriate for your data and problem.

**Train the model:**

Once you have chosen an algorithm, you will need to train the model on the data. This involves dividing the data into training and validation sets, and using the training set to optimize the model's parameters.

**Evaluate the model**:

After training the model, you will need to evaluate its performance on the validation set. This will give you an idea of how well the model will perform on new data.

**Deploy the model:**

Once you have a model that performs well on the validation set, you can deploy it in a clinical setting. This may involve integrating it into an electronic health record system, or creating a standalone application.



**Data collection and preparation:**

This involves gathering relevant data from various sources and preparing it for use in machine learning algorithms. In the case of thyroid disease prediction, this might include medical records, lab results, and patient demographics.

**Feature selection and engineering:**

Once the data has been collected, the next step is to select the most relevant features (i.e., variables or attributes) that are likely to be predictive of thyroid disease. This might involve domain expertise from medical professionals, as well as statistical techniques such as correlation analysis or principal component analysis.

**Model selection and training:**

With the features selected, the next step is to choose an appropriate machine learning algorithm to use for prediction. This might include traditional models such as logistic regression or more advanced methods such as deep learning

**Model evaluation and tuning:**

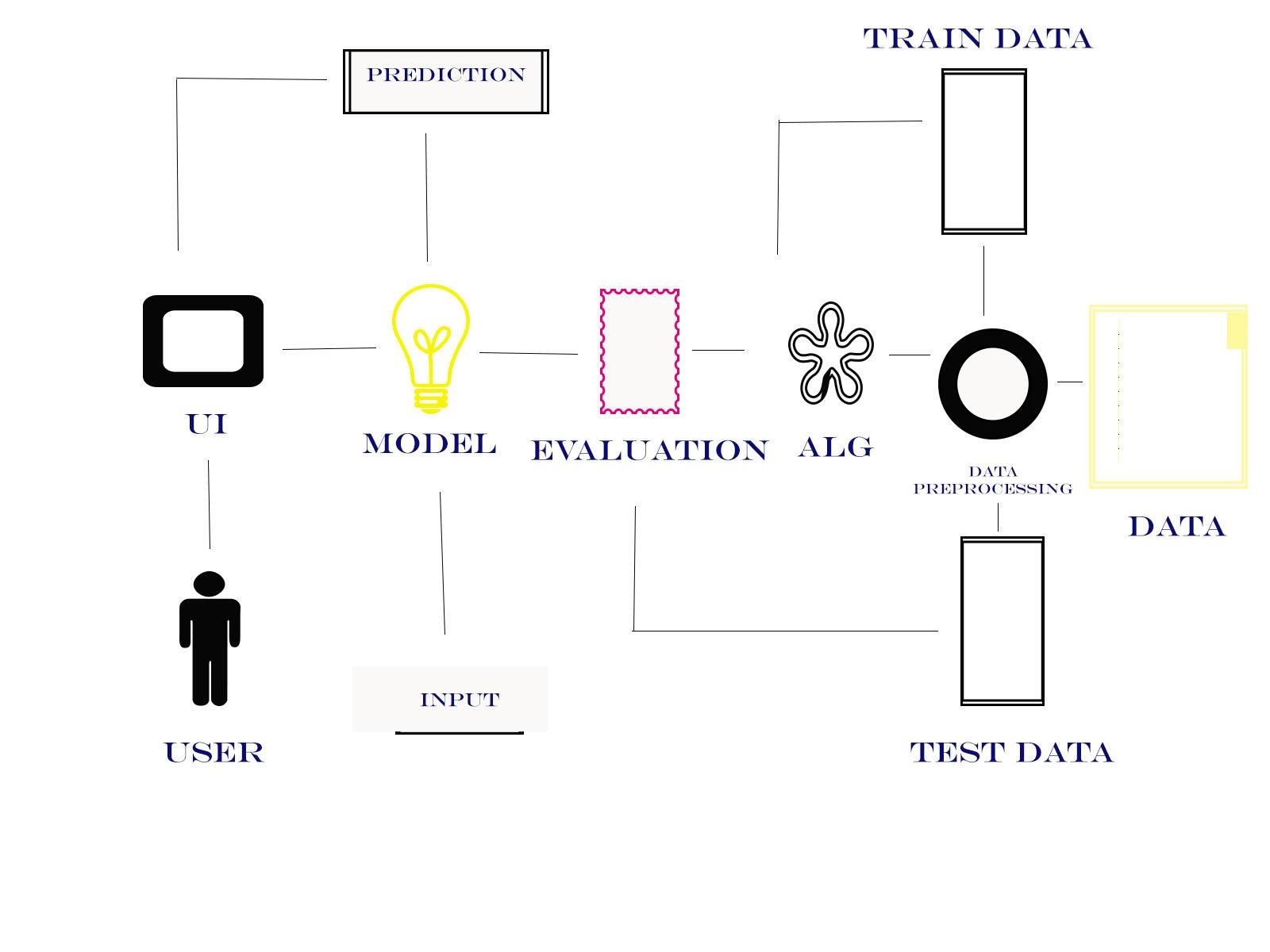
Once the model has been trained, it is evaluated using various metrics such as accuracy, precision, and recall **Deployment and monitoring:**

Finally, the trained and validated model can be deployed for use in clinical settings. Ongoing monitoring and evaluation of the model's performance are critical to ensure that it remains accurate and effective over time.



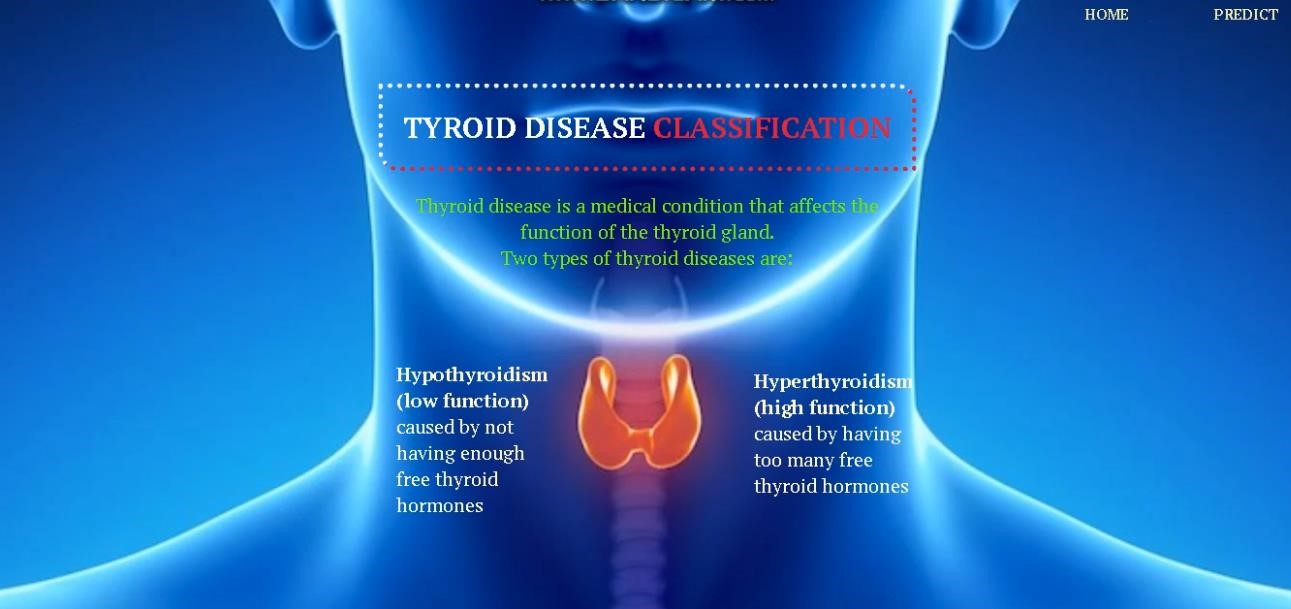
Machine Learning plays a very deciding role in disease prediction. Machine Learning algorithms such as SVM, Random Forest Classifier, K-NN, logistic regression and DT Classifier are used to predict the patient's risk of getting thyroid disease. The web app is created to get data from users to predict the type of disease. It predicts with 84% accuracy score.

# APPENDICES

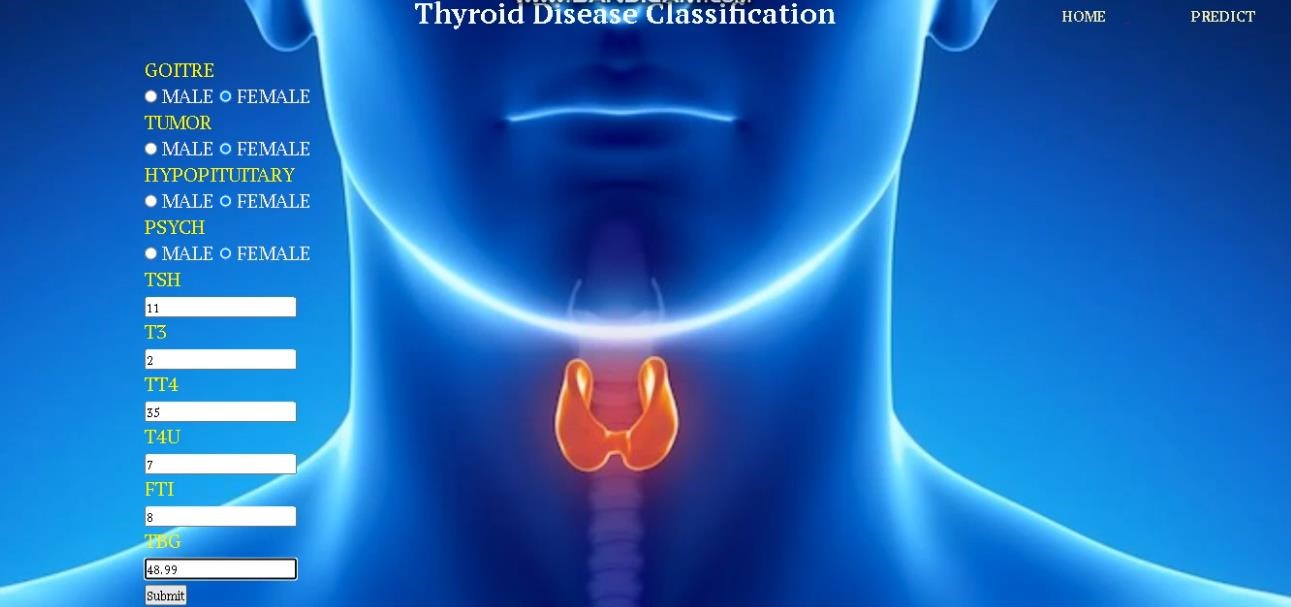


**HOME**





**PREDICT**



**SUBMIT**



