High

Kumar Trinayan Bharadwaj

trinayanbharadwaj@gmail.com

High level design (HLD)

Thyroid Disease Detection

Table of Contents

**Abstract2**

**1 Introduction3**

1.1 Why this High-Level Design Document3

1.2 scope3

**2 General Description**4

2.1 Product Perspective4

2.2 Problem Statement4

2.3 Proposed Solution4

2.4 Further Improvements4

2.5 Technical Requirements4

2.6 Data Requirements5

2.7 Tools Used5

2.8 Constraints6

2.9 Assumptions6

**3 Design Details**6

3.1 Process Flow6

3.1.2 Model Training and Evaluation7

3.1.3 Deployment Process7

3.2 Event Log8

3.3 Performance8

3.2 Reusability8

3.2 Application Compatibility8

3.2 Resource Utilization9

3.2 Deployment9

**4 Conclusion**9

**Abstract**

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are one of the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of body's metabolism. The machine learning plays a decisive role in the process of disease prediction and this paper handles the analysis and classification models that are being used in the thyroid disease based on the information gathered from the dataset taken from UCI machine learning repository.

**1 Introduction**

**1.1 Why this High-level Design document?**

High-level design (HLD) explains the architecture that would be used to develop a system. The architecture diagram provides an overview of an entire system, identifying the main components that would be developed for the product and their interfaces. The HLD uses possibly nontechnical to mildly technical terms that should be understandable to the administrators of the system.

The HLD will -

* Present all of the design aspects and define them in detail.
* Describe the user interface being implemented.
* Describe the hardware and software interfaces.
* Describe the performance requirements.
* Include design features and the architecture of the project.
* List and describe the non-functional attributes like:
  + Security
  + Reliability
  + Maintainability
  + Portability
  + Reusability
  + Application compatibility
  + Resource utilization
  + Serviceability

**1.2 Scope**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. It uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

**2 General Description**

**2.1 Product perspective:**

Thyroid Disease detection is an AI based solution which will help us in detecting thyroid disease in humans. It uses four parameters TSH, FTI, TT4 and T4U values to predict the risk of thyroidal disease.

**2.2 Problem Statement:**

To create an AI solution which will help medical teams in detecting thyroid diseases and will help in:

* Reducing human error
* Efficiency in detecting disease
* 24\*7 working model

**2.3 Proposed Solution:**

The solution proposed here is an ensemble of two machine learning models each with 4 features mentioned above. This solution is built to address the above-mentioned cases. In the first case, this solution eliminates the risk of human error. Secondly, our solution with an accuracy of 84% is fit enough put into work. Thirdly, unlike humans our product doesn’t need rest. It can work all day long. Finally, based on the given prediction the medical team will take necessary steps.

**2.4 Further Improvements:**

The solution can be improved by adding more data from the other sources like hospitals. Also, further feature engineering could result in a better accuracy of our models.

**2.5 Technical Requirements:**

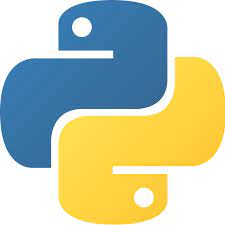
The primary requirement is a device through which our solution can be accessed. Our solution is deployed on Amazon web service (AWS). Thus, one will also require the web address as well.

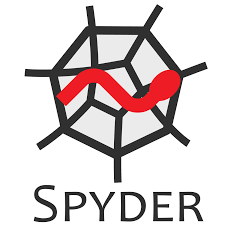
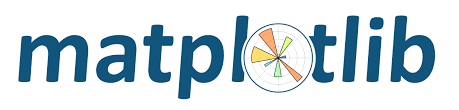
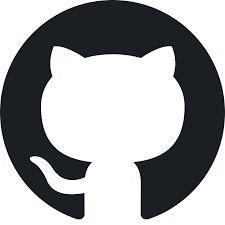
**2.6 Data Requirements:**

To run our solution, one needs only 4 values of the patience’s thyroid hormone value. TSH, FTI, TT4 and T4U are the thyroid hormones which needs to be measured in patience’s through blood test.

**2.7 Tools Used:**

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask are used to build the model.

**  **

**   **

* Jupyter notebook and Spyder are used as IDE.
* Python is used as our programming language.
* HTML is used to design our frontend.
* AWS is used for deployment of the model.
* Pandas, NumPy and Scikit-learn are used for data processing.
* Flask API is used to integrate our backend with our frontend.
* GitHub is used to store our code.
* Matplotlib and seaborn are used to visualize our data.

**2.8 Constraints:**

The Thyroid disease detection solution should be user friendly, as automated as possible and users should not be required to know any of the workings.

**2.9 Assumptions:**

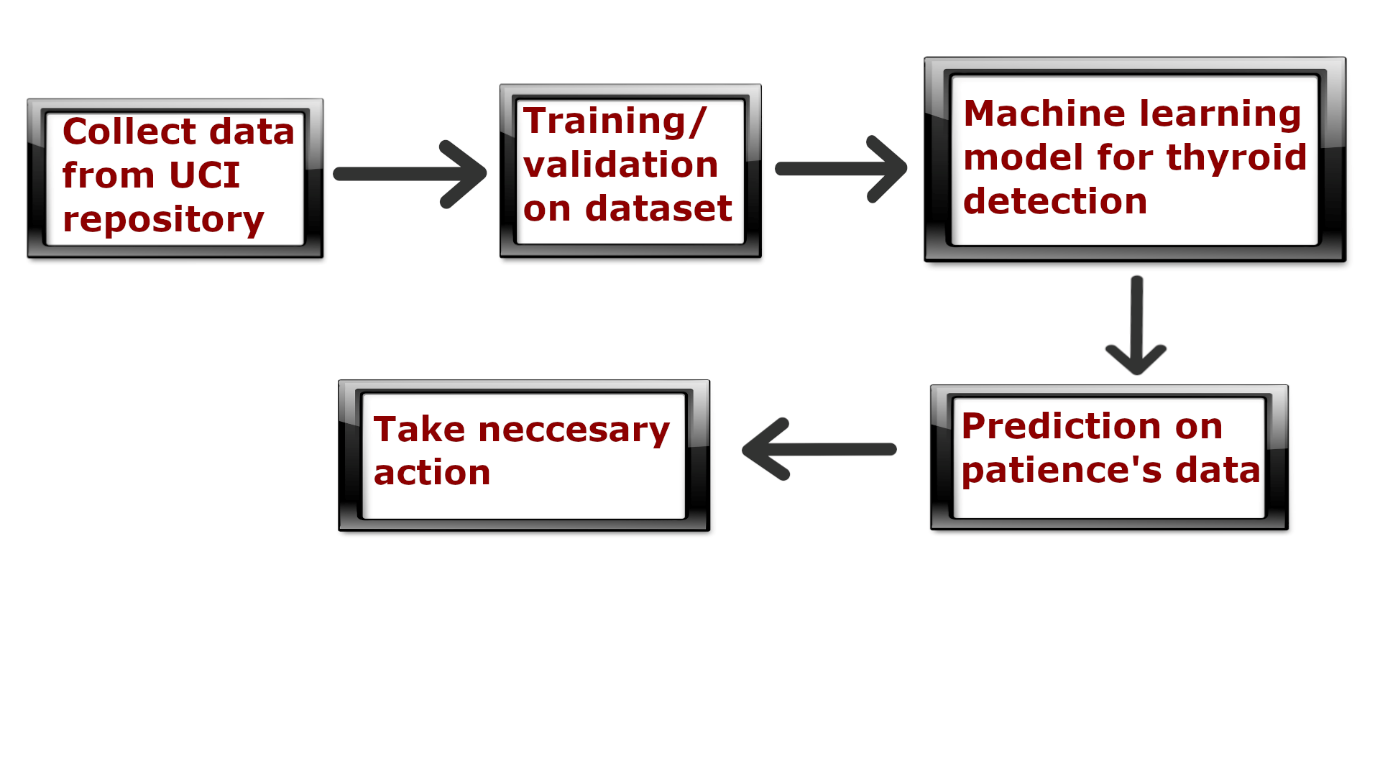
Machine learning based solution is used for detecting thyroid disease so it is assumed that there is a linear correlation between the 4 features. Also, correct values of the hormones is expected in order to get a valid prediction.

**3 Design Details**

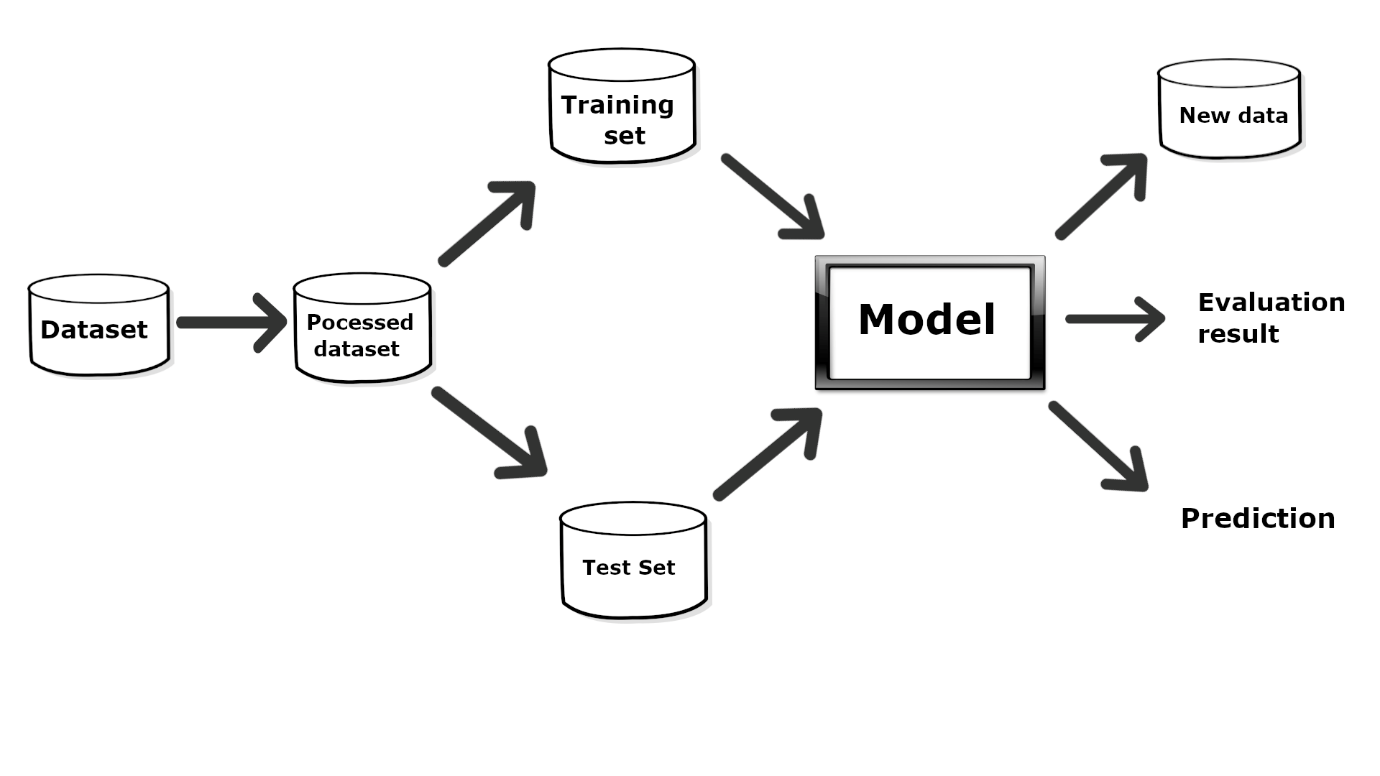
**3.1 Process Flow**

For identifying thyroidal diseases, we will use a machine learning based model. Below is the process flow diagram shown below.

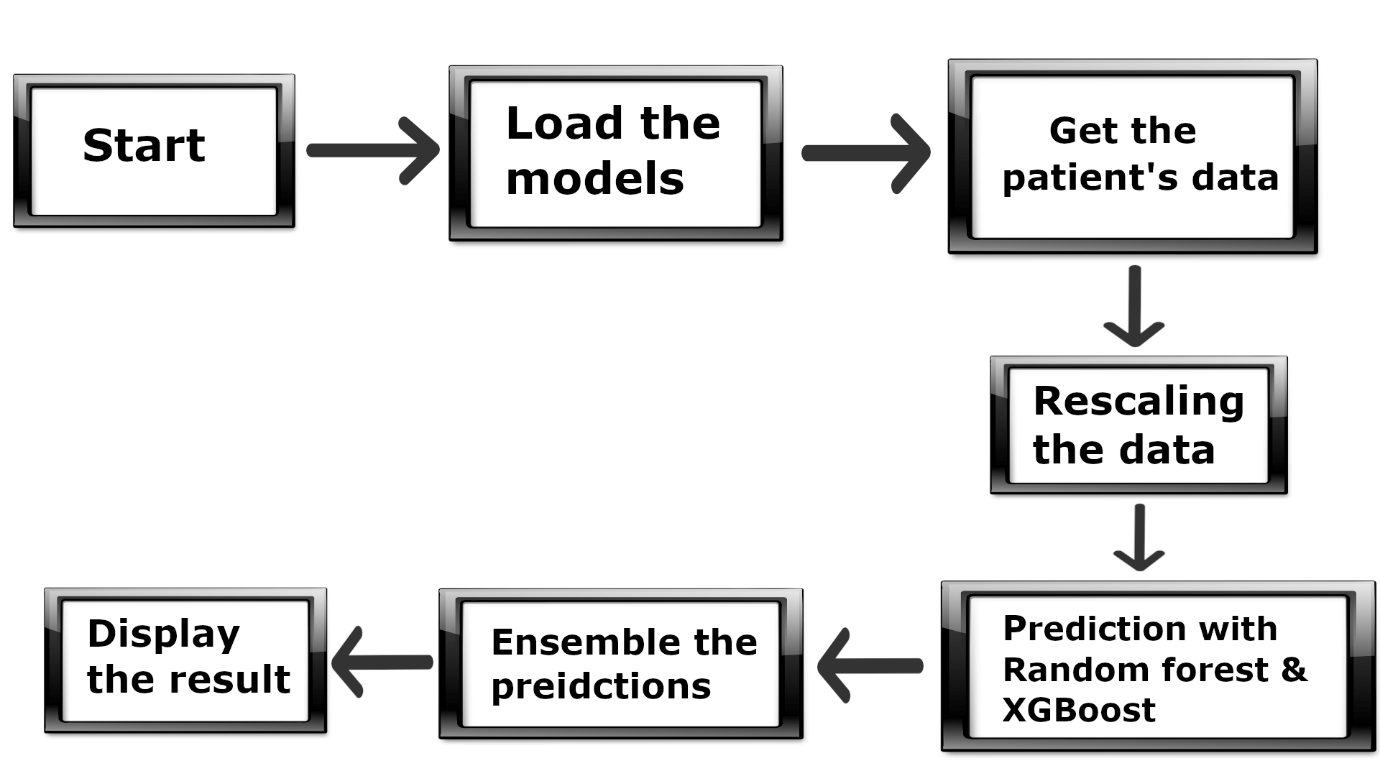
Proposed methodology



3.1.1 Model training and Evaluation



3.1.2 Deployment process



**3.2 Event Log**

The system should log every event so that the user will know what process is running internally.

Initial step by step description:

* The system identifies at what step logging is required
* The system should be able to log each and every system flow
* Developer can choose logging method. You can choose database logging or file logging.
* System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

**3.3 Performance**

Thyroid disease detection solution is used for detection of thyroid disease in patience based on the amount of hormones present in blood. This will help the medical team in diagnosing thyroidal disease in patience by minimising human error and increasing efficiency. Also, model retraining is very important to improve the performance.

**3.4 Reusability**

The code written and the components used should have the ability to be reused with no problems.

**3.5 Application Compatibility**

The different components for this project will be using Python as an interface between them. Each component will have its own risk task to perform, and it is the job of the Python to ensure proper transfer of information.

**3.6 Resource Utilization**

When any task is performed, it will likely use all the processing power available until that function is finished.

**3.7 Deployment**

****

**4 Conclusion**

The thyroid disease detection solution will help the medical teams to detect thyroidal disease in patience based on the quantity of 4 hormones. This solution eliminates human error and increases efficiency in diagnosing thyroid disease. After the prediction, necessary medical action will be taken by the concerned authorities.

**5 References**

1. <https://archive.ics.uci.edu/>
2. Google.com for images