HIGH LEVEL DOCUMENTATION (HLD)

THYROID DISEASE DETECTION

Revision Number:

Last date of revision:

Document Version Control

Date Issued	Version	Description	Author
1-11-23	1	Initial HLD-V1.0	Komal Madiwal

Contents

Abstract

The thyroid gland is a vital organ in the human body, secreting two hormones—total serum thyroxin (T4) and total serum triiodothyronine (T3)—that regulate metabolism. Common thyroid disorders include Hyperthyroidism, associated with excessive thyroid hormone circulation, and Hypothyroidism, often due to a lack of thyroid hormone generation and poor alternate therapy.

Machine learning, a powerful tool in disease prediction, is used to analyse and classify thyroid diseases based on a dataset from the UCI machine learning repository. Various algorithms such as SVM (support vector machine), decision tree, logistic regression, KNN (K-nearest neighbours), decision tree, random forest, and xgboost are employed to predict the risk of thyroid disease in patients.

Concurrently, a user-friendly web application has been developed using Flask, a web framework. This application enables individuals to easily input their data, making it convenient to receive predictions about their risk of thyroid disease

1.Introduction

1.1Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary details to the current project description to represent a suitable model for coding. This document isalso intended to help detect contradictions prior to coding, and can be used as referencemanual for how the modules interact at a high level.

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 feature and the architecture of the project
- List and describe the non-functional attribute like:

    1. Security
    2. Reliability
    3. Maintainability
    4. Portability
    5. Reusability
    6. Application compatibility
    7. Resource utilization
    8. Serviceability

1.2 Scope

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

1.3 Definitions

- TDD – Thyroid Disease Detection

2 General Description

2.1 Product Perspective

The Thyroid Disease Detection solution system is a data science-based machine learning model that helps us detect whether the individual is suffering from the disease or not.

2.2 Problem Statement

To create an AI solution for detecting thyroid disease and to implement the following usecases.

- To detect thyroid disease and its type in healthy person.
- To detect thyroid disease and its type in unhealthy person.

Here unhealthy person means person already affected by thyroid disease.

2.3 Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. In first use case, we will take input from a healthy person who is not suffering from thyroid disease and see whether proposed solution is going to detect it or not. And in second use case, we will take inputfrom an unhealthy person, already suffering from thyroid disease and check our solutionwhether it is performing or not in right way.

2.4 Further Improvements

The Thyroid disease detection solution can be added with more use cases in health care domain. TDD solution can also be synchronized with other health care domainsolution to give one step extra confirmation of health to those people who has littlesymptoms of thyroid disease also.

2.5 Data Requirements

Data requirement completely depend on our problem statement.

We need data of people who have already gone with thyroid blood test to know whether they are suffering from thyroid disease or not. If yes then what kind of thyroid disease they are suffering from. We will be required thesemany attributes, in which some will be personal details and some will be attributes from blood test.

- Age: age of the patient
- Sex: sex patient identifies(male or female)
- on thyroxine: whether patient is on thyroxine
- on antithyroid meds: whether the patient is on antithyroid meds
- sick: whether patient is sick
- pregnant-whether patient is pregnant (bool)
- thyroid surgery- whether patient has undergone thyroid surgery
- I131_treatment- whether patient is undergoing I131 treatment
- Lithium test: Lithium is concentrated by the thyroid and inhibits thyroidal iodine uptake
- Goitre test: A goitre can sometimes occur when your thyroid gland produces too much thyroid hormone(hyperthyroidism).
- Tumour test: Thyroid cancer occurs when cells in your thyroid undergo genetic changes (mutations). The mutationsallow the cells to grow and multiply rapidly. The cells also lose the ability to die, as normal cells would. The accumulating abnormal thyroid cells form a tumour.
- TSH level measure: It supervise thyroid gland, TSH released by pituitary gland. Normal TSH range for an adult:0.40 - 4.50 mIU/mL (milli-international units per litre of blood).
- T3 level measure: Hormone released by thyroid, should be in normal range.
- T4 level measure: Low T4 is seen with hypothyroidism, whereas high T4 levels may indicate hyperthyroidism.Normal T4 range for an adult: 5.0 – 11.0 ug/dL (micrograms per decilitre of blood).
- FTI (Free T4 or Free Thyroxine: The free T4 index (FTI) is a blood test used to diagnose thyroid disorders. The FTI isobtained by multiplying the (Total T4) times (T3 Uptake) to obtain an index. Normal FT3 range: 2.3 - 4.1 pg/mL (picograms per millilitre of blood).

- Thyroxine-binding globulin (TBG): The TBG blood test measures the levelof a protein that moves thyroid hormone throughout your body.

  In all the above-mentioned attributes if attribute is having binary answer, thenwe need it in Boolean and for measures we need them in float values.

2.6 Tools used

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

- Virtual Studio Code is also used as IDE.
- Matplotlib, Seaborn and Plotly are used for visualization of the plot.
- AWS is used for deployment of the model.
- Python, Flask is used for backend development
- Github is use for version control system

2.7 Constraints

The Thyroid Disease Detection solution system must be correct enough that it not misleads any report and as automated as possible and users should not be required to know any of the workings.

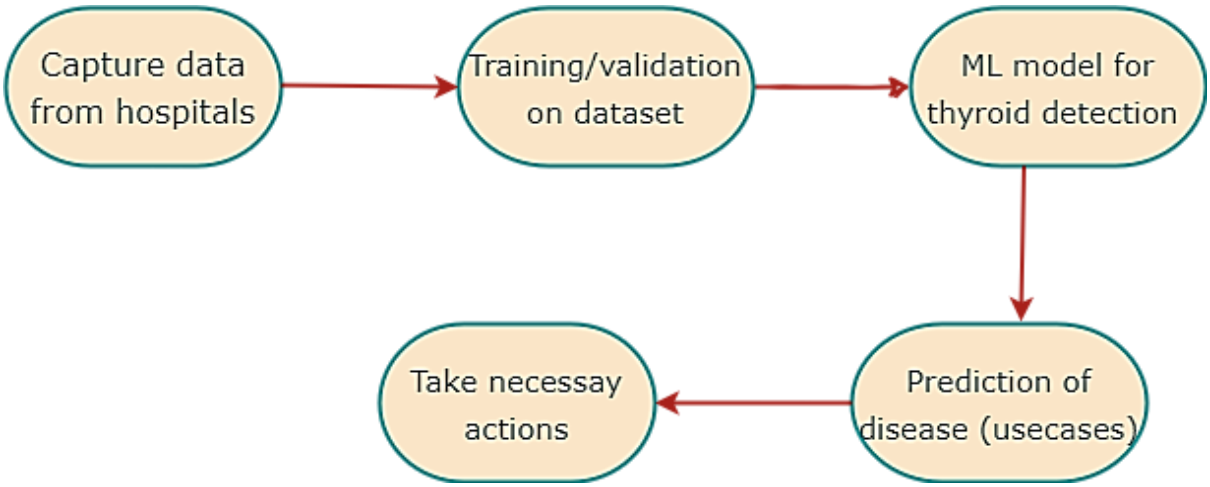
2.8 Assumptions

The primary goal of the project is to execute the predefined use cases on new datasets originating from hospitals where the Thyroid Disease Detection solution is installed, capturing reports of individuals within their campus.

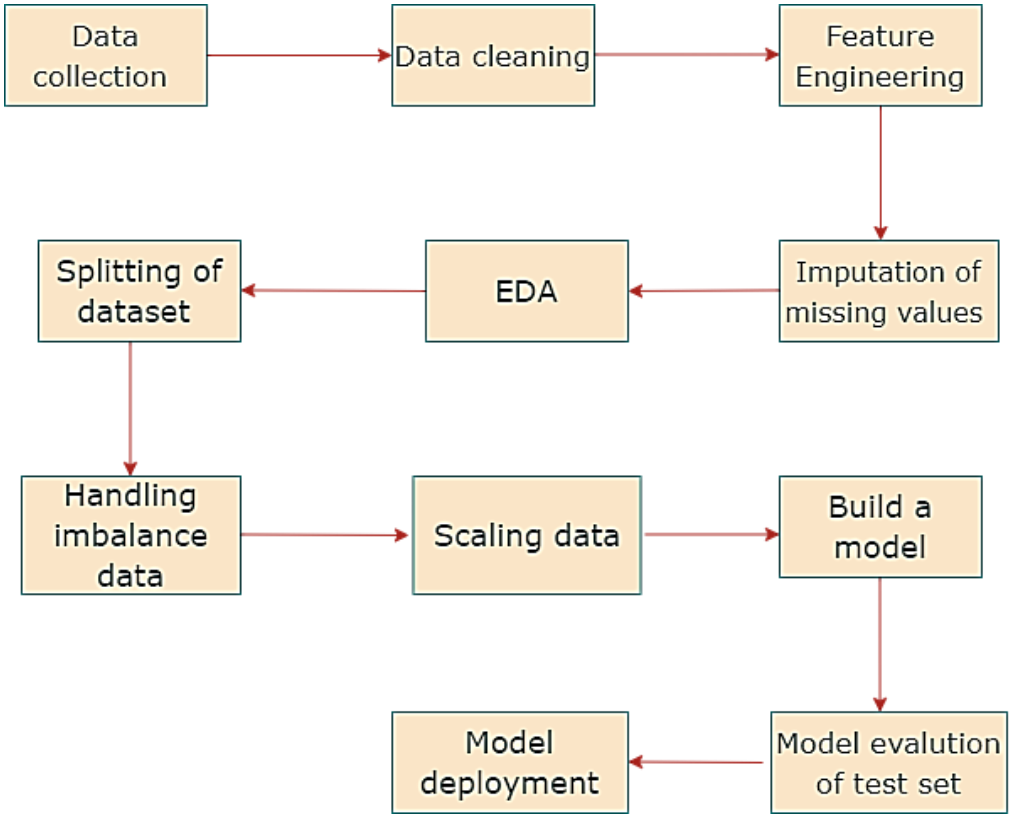
3.Design details

3.1 Process Flow

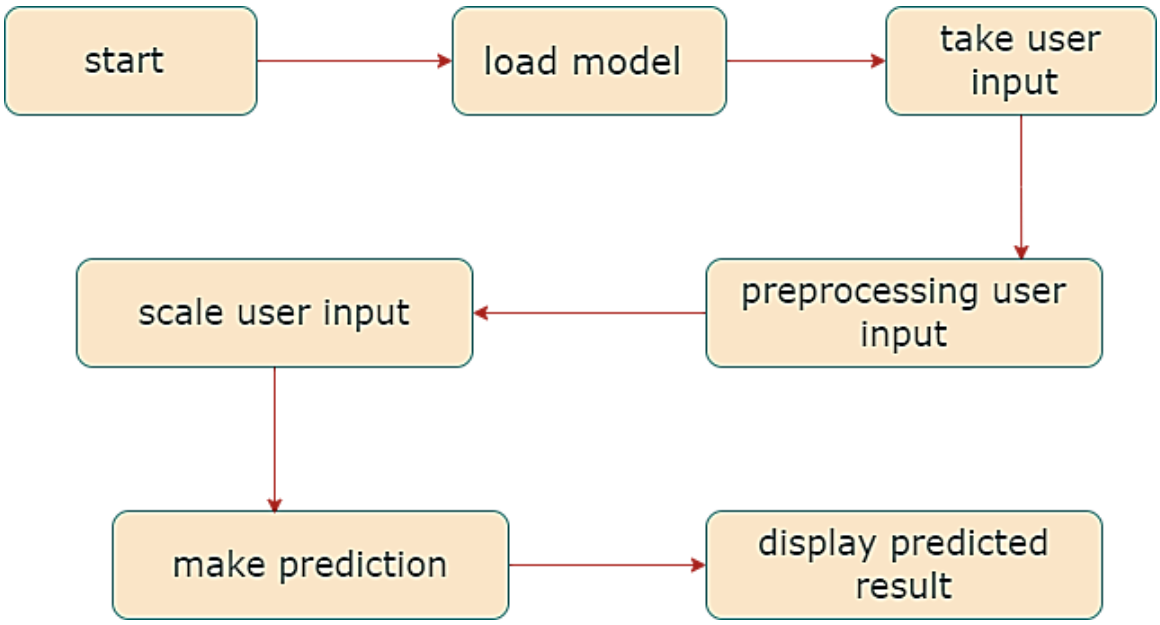
For detecting thyroid disease, we will use machine learning base model. Below is the processflow diagram is as shown below



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:

1. The System identifies at what step logging required.
2. The System should be able to log each and every system flow.
3. Developer can choose logging method. You can choose database logging/ File logging s well.
4. System should not hang even after using so many loggings. Logging just because we can easilydebug issues so logging is mandatory to do.

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? Anerror will be defined as anything that falls outside the normal and intended usage.

4 Performance

The machine learning-based Thyroid Disease Detection solution will be used to detect thyroid diseases in patients exhibiting symptoms, ensuring that necessary action is taken as soon as possible. Additionally, regularly updating (retraining) the model is crucial for continuous improvement in performance.

4.1 Reusability

The code written and the components used should have the ability to be reused withno problems.

4.2 Application Compatibility

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

4.3 Resource utilization

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

4.4 Deployment

5 Conclusion

Thyroid Disease Detection solution will take health-care domain data of those patients who have undergone diagnosis for thyroid to train our machine learning model and will evaluate its performance over use cases mentioned above. And then leverage its prediction to detect thyroid disease in people having symptoms of thyroid and able toalert people who is on positive side so that medical attention along with treatment will be given to that particular people as soon as possible. This solution should be as accurate as possible, so that chances of misleading reports will be taken good care of.

6 References

- UCI Machine Learning Repository For Data Set
- https://archive.ics.uci.edu/dataset/102/thyroid+disease