High Level Design (HLD)

Thyroid Disease Detection

Revision Number:
Last date of revision:

Document Version Control

Date Issued	Version	Description	Author
11-10-2021	1	Initial HLD-V1.0	Umasangavi G

Contents

Abstract

At least a person out of ten is suffered from thyroid disease in India. The disorder of thyroid disease primarily happens in the women having the age of 17–54. The extreme stage of thyroid results in cardiovascular complications, increase in blood pressure, maximizes the cholesterol level, depression and decreased fertility. The hormones, **total serum thyroxin (T4)** and **total serum triiodothyronine (T3)** are the two active thyroid hormones produced by the thyroid gland to control the metabolism of body. For the functioning of each cell and each tissue and organ in a right way, in overall energy yield and regulation and to generate proteins in the ordnance of body temperature, these hormones are necessary .

The basis of classification of thyroid disease is **euthyroidism**, **hyperthyroidism** and **hypothyroidism** which are denoting normal, excessive or defective levels of thyroid hormones. The state euthyroidism depicts the normal production of thyroid hormones and normal levels at the cellular level by the thyroid gland. The state hyperthyroidism is clinical symptom due to excessive circulation and intracellular thyroid hormones. The state hypothyroidism is most of due to the lack of thyroid hormone generation and poor alternate therapy.

Cure of disease is a regular concern for the health care practitioners, and the errorless diagnostic at the right time for a patient is very important. Recently, by some advanced diagnosis methods, the common medical report can be generated with an additional report based on symptoms. The different questions like ''what are the causes for affecting the thyroid?'', ''Which age group of people are affected due to thyroid?'', ''what is the relevant treatment for a disease?''. **All these answers we can find on implementing machine learning methods on Health care data.** Health care data can be processed and after implementing with certain methodologies; it can provide information that can be used in diagnosis and treatment of diseases more efficiently and accurately with better decision making and minimizing the death risk.

The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not. The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Try out different machine learning algorithm that's best fit for the above case

1   Introduction

1.1   Why 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 is also intended to help detect contradictions prior to coding, and can be used as reference manual 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:

    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology 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 which help us to detect the thyroid disease in people and take necessary action.

2.2 Problem Statement

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

- 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 input from an unhealthy person, already suffering from thyroid disease and check our solution whether 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 domain solution to give one step extra confirmation of health to those people who has little symptoms of thyroid disease also.

2.5 Data Requirements

age:	continuous,?.
sex:	M,F,?.
on_thyroxine:	f,t.
query_on_thyroxine:	f,t.
on_antithyroid_medication:	f,t.
thyroid_surgery:	f,t.
query_hypothyroid:	f,t.
query_hyperthyroid:	f,t.
pregnant:	f,t.
sick:	f,t.
tumor:	f,t.
lithium:	f,t.
goitre:	f,t.
TSH_measured:	f,t.
TSH:	continuous,?.
T3_measured:	f,t.
T3:	continuous,?.
TT4_measured:	f,t.
TT4:	continuous,?.
T4U_measured:	f,t.
T4U:	continuous,?.
FTI_measured:	f,t.
FTI:	continuous,?.
TBG_measured:	f,t.
TBG:	continuous,?.

2.6 Tools used

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

- Virtual Studio Code is also used as IDE
- Heroku is also used for deployment of the model.
- Python, Flask is used for backend development
- Github is used as Version Contol System.

2.7 Constraints

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

2.8 Assumptions

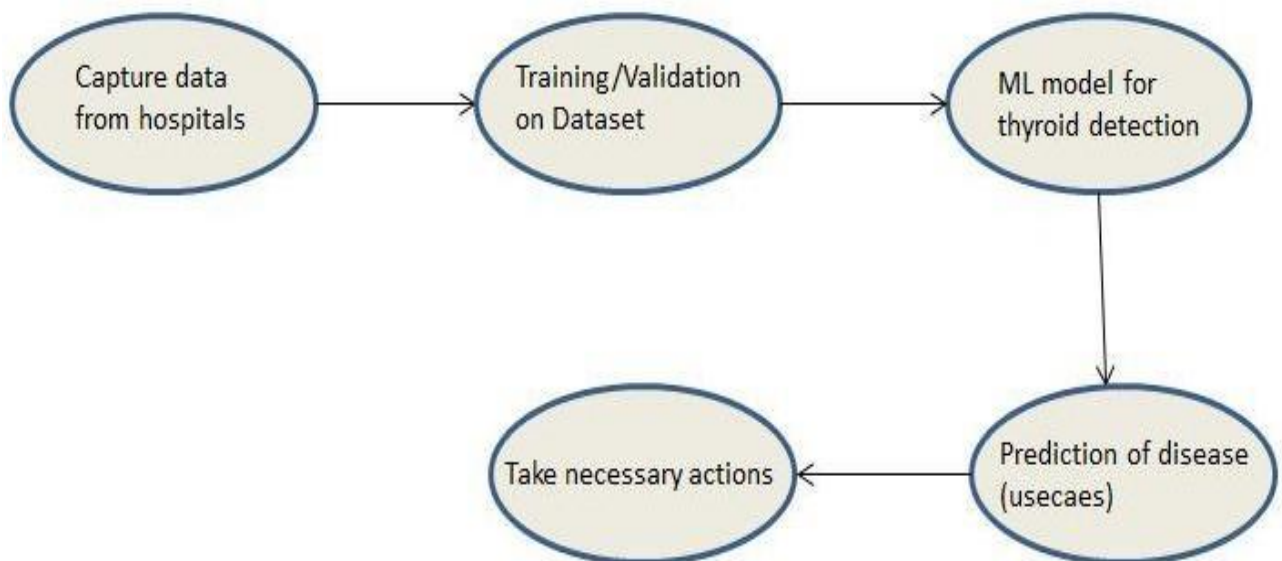The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through Hospitals which has this solution install in their campus to capture people reports.
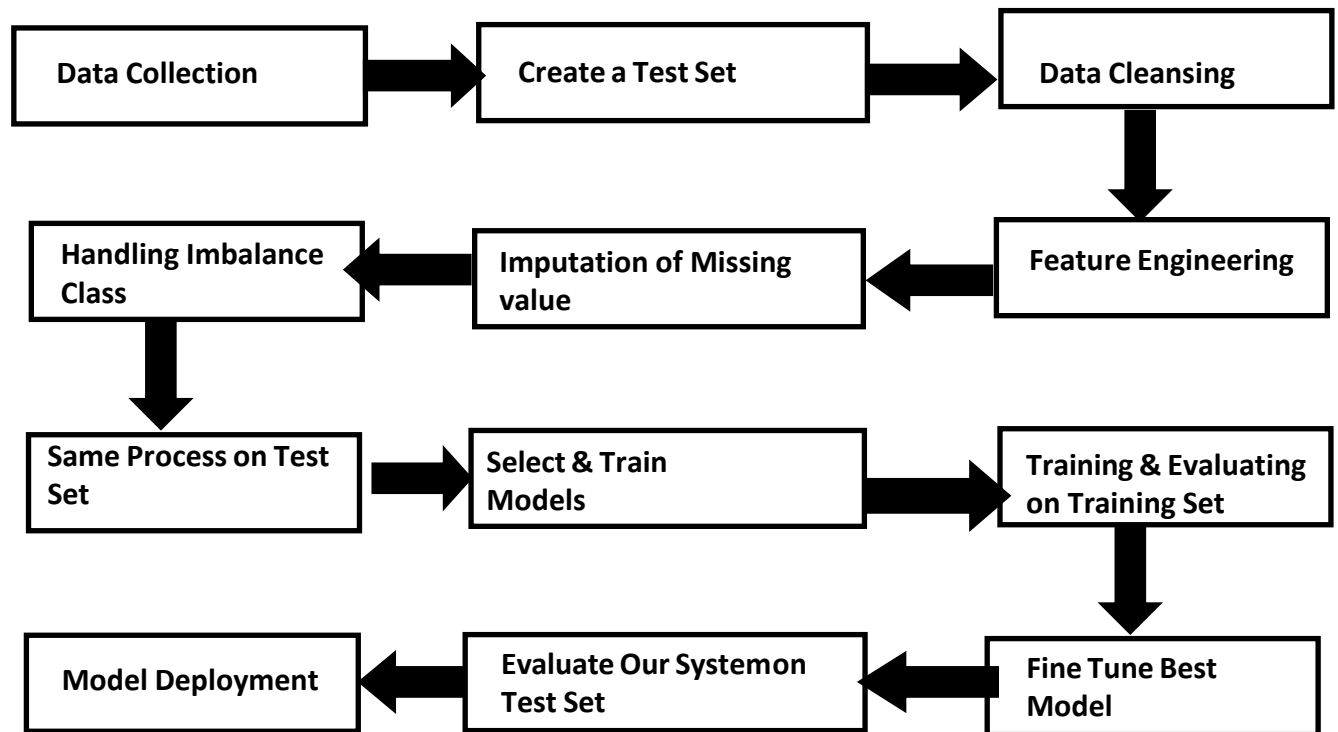
3 Design Details

3.1 Process Flow

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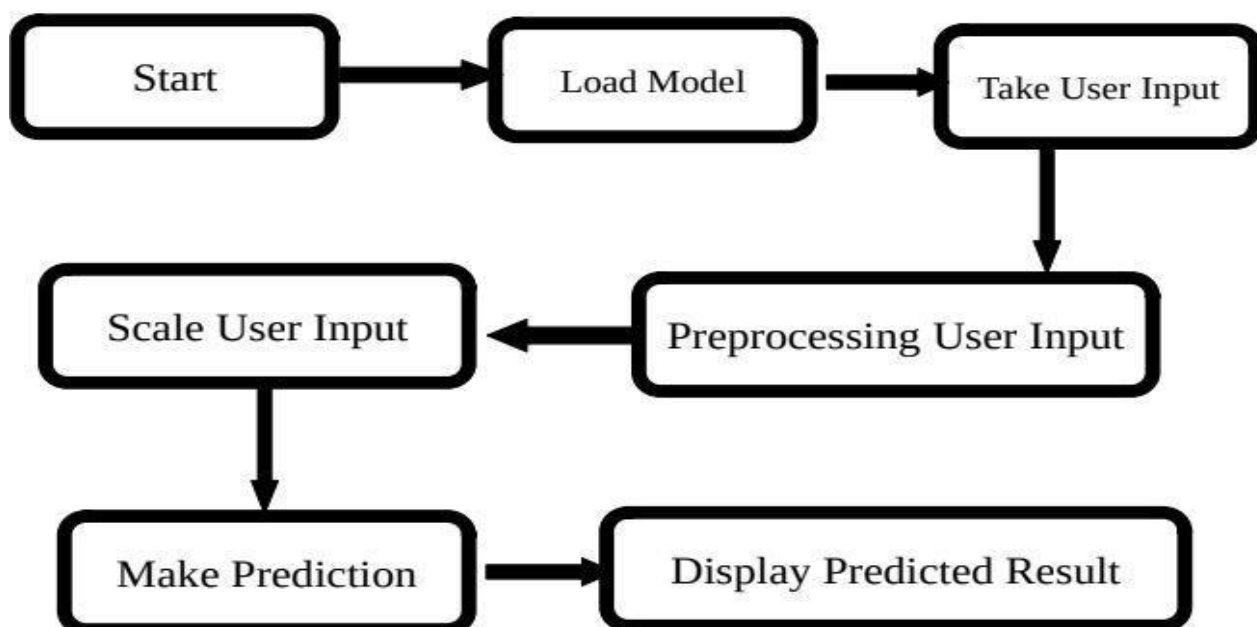
Proposed methodology

3.1.1 Model Training and Evaluation

```
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Data Collection  │ ───▶ │ Create a Test Set│ ───▶ │ Data Cleansing   │
└──────────────────┘      └──────────────────┘      └──────────────────┘
                                                              │
                                                              ▼
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Handling Imbalance│ ◀── │ Imputation of    │ ◀── │ Feature          │
│ Class            │      │ Missing value    │      │ Engineering      │
└──────────────────┘      └──────────────────┘      └──────────────────┘
         │
         ▼
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Same Process on  │ ───▶ │ Select & Train   │ ───▶ │ Training &       │
│ Test Set         │      │ Models           │      │ Evaluating on    │
│                  │      │                  │      │ Training Set     │
└──────────────────┘      └──────────────────┘      └──────────────────┘
                                                              │
                                                              ▼
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Model Deployment │ ◀── │ Evaluate Our     │ ◀── │ Fine Tune Best   │
│                  │      │ Systemon Test Set│      │ Model            │
└──────────────────┘      └──────────────────┘      └──────────────────┘
```

3.1.2 Deployment Process

```
┌──────────┐      ┌────────────┐      ┌──────────────┐
│  Start   │ ───▶ │ Load Model │ ───▶ │ Take User    │
│          │      │            │      │ Input        │
└──────────┘      └────────────┘      └──────────────┘
                                             │
                                             ▼
┌──────────────┐      ┌──────────────────────┐
│ Scale User   │ ◀── │ Preprocessing User    │
│ Input        │      │ Input                 │
└──────────────┘      └──────────────────────┘
       │
       ▼
┌──────────────┐      ┌──────────────────────┐
│ Make         │ ───▶ │ Display Predicted     │
│ Prediction   │      │ Result                │
└──────────────┘      └──────────────────────┘
```

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 easily debug 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? An error will be defined as anything that falls outside the normal and intended usage.

4 Performance

The machine learning based Thyroid Disease Detection solution will used for detection of thyroid disease in patients having symptoms of thyroid. So that necessary action will be taken ASP. Also model retraining is very important to improve performance.

4.1 Reusability

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

4.2 Application Compatibility

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

4.3 Resource utilization

When any task is performed, it will likely use all the processing power available until that 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 to alert 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

URL: https://archive.ics.uci.edu/ml/datasets/thyroid+disease