

# High Level Design (HLD)

## Thyroid Disease Prediction Model

Revision No. 1.0  
Last Revision Date: 3/12/2023

## Document Version Control

Date Issued	Version	Description	Author
3/12/2023	1	Initial HLD – V1.0	Ravi Yadav

## Contents

Document Version Control.....	2
Abstract.....	4
1 Introduction .....	5
1.1 Why this High-Level Design Document?.....	5
1.2 Scope.....	5
1.3 Definitions .....	5
2 General Description.....	6
2.1 Product Perspective .....	6
2.2 Problem statement .....	6
2.3 PROPOSED SOLUTION.....	6
2.4 Technical Requirements.....	6
2.5 Data Requirements.....	6
2.6 Tools used .....	7
2.7 Constraints.....	8
2.8 Assumptions.....	8
3 Design Details.....	9
3.1 Process Flow.....	9
3.1.1 Model Training and Evaluation .....	10
3.2 Event log .....	11
3.3 Error Handling .....	11
4 Performance.....	12
4.1 Reusability.....	12
4.2 Application Compatibility .....	12
4.3 Resource Utilization .....	12
5 Conclusion.....	13

## Abstract

Disease Prediction using Machine Learning is the system that is used to predict the diseases from the symptoms which are given by the patients or any user. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease. Gradient Boosting classifier is used in the prediction of the disease which is a supervised machine learning algorithm. The probability of the disease is calculated by the Gradient Boosting algorithm. With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care

# 1 Introduction

## 1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail 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 a 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 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. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

## 1.3 Definitions

<i>Term</i>	<i>Description</i>
FTI	Free thyroxine index, specifies the free T4 serum
TSH	Thyroid stimulating hormone, a test used to measure the thyroid hormones level
T3	Triiodothyronine hormone, controls the metabolism activity
T4	Thyroxine hormone, a thyroid hormone

## 2 General Description

### 2.1 Product Perspective

Thyroid disease prediction model is supervised machine learning model which is classifies either a person suffering from thyroidal illness or not.

### 2.2 Problem Statement

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 main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.

### 2.3 PROPOSED SOLUTION

Thyroidal illness causes due to uncontrolled thyroid hormone release from thyroid gland which affects the various level of various hormones in human body like TSH (Thyroid Stimulating Hormone), T4 (Thyroxine Hormone), T3 (Triiodothyronine) etc. Thyroid disease prediction model will predict the result either positive or negative based on these hormone level values.

### 2.4 Technical Requirements

This document addresses the requirements for detecting the thyroid disease of a person. The model requires the TSH, T4, T3, FTI and age value.

- TSH value can be obtained through the blood test of a person. It is an initial test to test the thyroid function.
- T3 value, it is also obtained by the blood test. This hormone controls the metabolism of a body.
- FTI value, it used to measure the serum T4 in the body. It is an optional test.
- T4 value, it is used to measure both free and bound thyroxine hormone.

### 2.5 Data Requirements

Data requirement completely depend on our problem statement.

- We need at least 4000 to 5000 rows in dataset and mandatory columns like FTI, TSH, T3, T4 having numeric values and also a label defining the class of person either negative, hyperthyroid, hypothyroid (i.e. Positive class) etc.
- Each label should be in equal proportion to the other class labels.
- All data could be in string datatype and in comma separated format.

- In features where values are numeric should not contain any alphabetic or special character.
- Class feature having disease status i.e. either negative or positive (hyperthyroid or hypothyroid) should not have any special character.

## 2.6 Tools used

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



- VS Code is an integrated development environment provide support to many programming languages like C, C++, Java, Python, Javascript and many others.
- Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive.
- Scikit-learn is a free, open-source machine learning library for Python. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means, etc.
- NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.
- Seaborn is a Python data visualization library based on matplotlib.
- GitHub is a code hosting platform for version control and collaboration. It lets you and others work together on projects from anywhere.
- Rendêr is a unified cloud to build and run all your apps and websites with free TLS certificates, global CDN, private networks and auto deploys from Git.
- Docker is a software platform that allows you to build, test, and deploy applications quickly. Docker packages software into standardized units called containers that have everything the software needs to run including libraries, system tools, code, and runtime.
- Flask is a Python web application framework that allows developers to build lightweight web applications.
- Apache Cassandra is a NoSQL, open-source, distributed database. It's designed to store data for applications that require fast read and write performance.

## 2.7 Constraints

Thyroid disease prediction system must be user friendly, as automated as possible and users should not be required to know any of the internal working.

## 2.8 Assumptions

The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset that comes through user. Supervised machine learning based thyroid detection model is used for detecting the above-mentioned use cases based on the input data. It is also assumed for detecting the above-mentioned use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

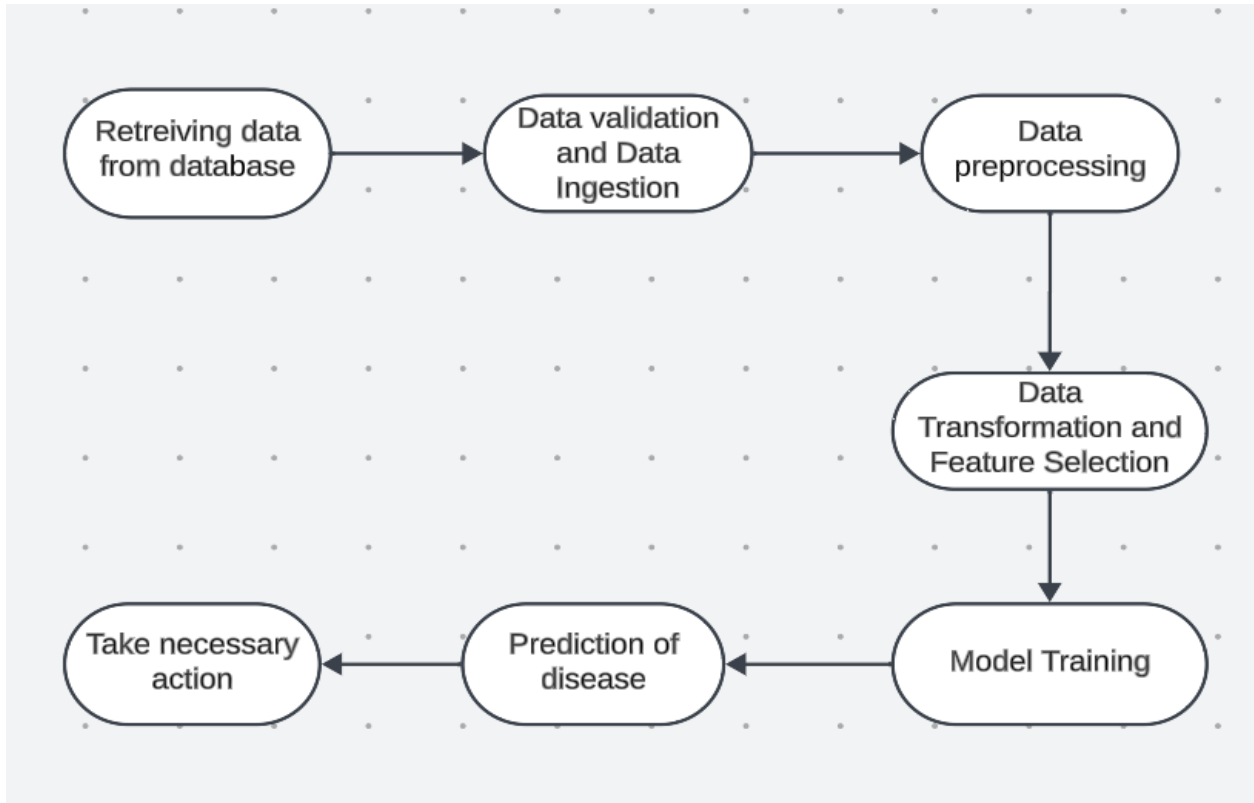


### 3 Design Details

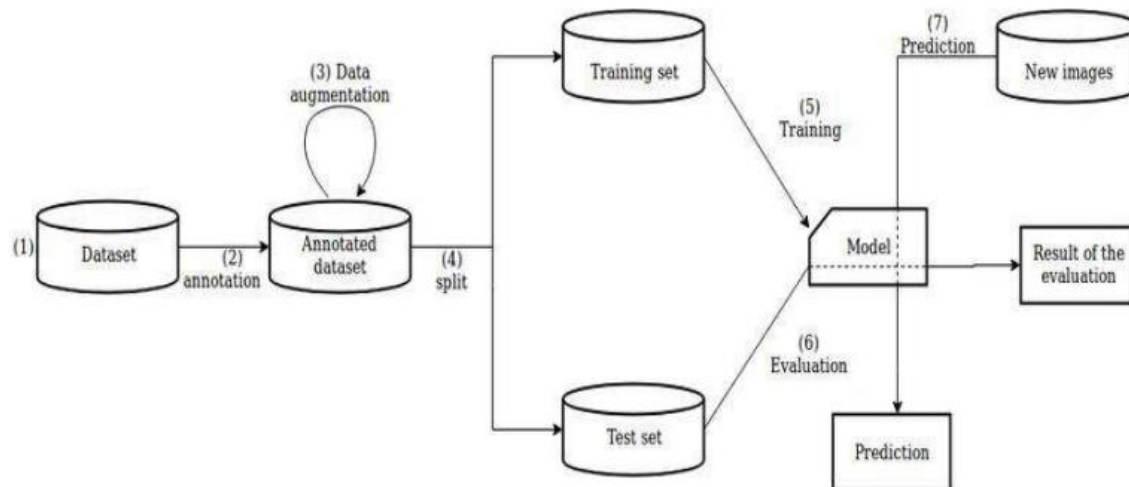
#### 3.1 Process Flow

For identifying the person condition, we will use a machine learning base model. Below is the process flow diagram as shown below.

##### Proposed methodology



### 3.1.1 Model Training and Evaluation



## 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 as 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

Model retraining is very important step to improve the performance based on additional updated dataset.

### 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

# render

## 5 Conclusion

Thyroid disease prediction model will detect either a person has thyroid disease or not based on various hormone features. So, a person can take necessary action based on the condition

## 6 References

1. Google.com for images of various frameworks tools
2. <https://www.lucidchart.com/>