Low Level Design (LLD)

Mice Protein Expression

Contents

**Abstract 4**

1 Introduction 4

1.1 Why this Low-Level Design Document? 4

1.2 Scope 4

1.3 Constraints 5

1.4 Risks 5

1.5 Out of Scope 5

2 Technical specifications 5

2.1 Dataset 5

2.1.1 Dataset Overview 5

2.1.2 Input Schema 5

2.2 Predicting Disease 6

2.3 Database 6

2.4 Deployment 6

3 Proposed Solution 6

4 Model training / validation workflow 7

5 Test cases 9

6 Key performance indicators (KPI) 10

Abstract

Down syndrome (DS) is a chromosomal abnormality (trisomy of human chromosome 21) associated with intellectual disability and affecting approximately one in hundred live births worldwide. The over expression of genes encoded by the extra copy of a chromosome in DS is believed to be sufficient to perturb normal pathways and normal response to stimulation causing learning and memory deficits.

1. Introduction

1.1 Why this Low-Level Design Document ?

The purpose of this document is to present a detailed description of Deep EHR System. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stake holders and the developers of the system and will be proposed to the higher management for its approval.

The aim is to identify subsets of protein that are discriminant between the classes.

1.2 Scope

This software system will be a Web application. This system will be designed to detect the diseases at earliest for better disease management. Early detection of any preventable disease is important for better disease management.

1.3 Constraints

The dataset contains only one type of disease i.e Down Syndrome (trisomic sample/mice).

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Undefined protein sample is the out of scope for the project.

2. Technical specifications

2.1 Dataset

|  |  |  |
| --- | --- | --- |
| **Disease** | **Finalized** | **Source** |
| Down syndrome | Yes |  |

2.1.1 Dataset Overview

The data set includes the expression levels of 77 proteins/protein changes that generated measurable signals in the cortex's nuclear fraction. There are 72 mice in all, with 38 control mice and 34 trisomic mice (Down syndrome). In the experiments, 15 measurements of each protein per sample/mouse were recorded. As a result, there are 38x15, or 570 measurements for control mice and 34x15, or 510 measurements for trisomic mice. There are 1080 measurements per protein in the dataset. Each measurement can be thought of as a separate sample/mouse.

2.1.2 Input schema

|  |  |  |
| --- | --- | --- |
| **Feature name** | **Datatype** | **Size** |
| Name of protein | int | - |

2.2 Predicting Disease

* When the input is given the system classifies accordingly.
* After the classification the sample will be recognized as control or trisomic.

2.3 Database

System needs to store request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The system should be user friendly.
2. The system stores each and every data given in input.

2.4 Deployment

1. AWS



3. Proposed Solution

Here we are using the protein sample of proteins implanted in a mice to identify the trisomic mice/sample. Some of the Machine Learning algorithms are used to train the model using python programming language.

4. Model Training / Validation workflow

Start

Data Preprocessing

Data Description

End

Output

Data Classification

Classifies Accordingly

Model Training

Data Analysis

Feature Selection

Data Input (sample)

Model Start

Data Visualization

5. Test Cases

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Case** | **Steps to perform test case** | **Module** | **Pass / Fail** |
|  |  |  |  |

6. Key performance indicators (KPI)

* Comparison of accuracy of model prediction and doctor’s prediction.
* Number of protein samples.
* Types of data input.