High Level Design (HLD)

Mice Protein Expression

Contents

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………………………………………………………………………………….5

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…………………………………………………………………………………………..7

2.8 Assumptions………………………………………………………………………………………..7

3 Design Details…………………………………………………………………………………………….8

3.1 Process Flow………………………………………………………………………………………..8

3.1.1 Model Training and Evaluation………………………………………………………8

3.1.2 Deployment Process………………………………………………………………………9

3.2 Event log………………………………………………………………………………………………9

3.3 Error Handling………………………………………………………………………………………9

4 Performance……………………………………………………………………………………………10

4.1 Reusability…………………………………………………………………………………………10

4.2 Application Compatibility…………………………………………………………………10

4.3 Resource Utilization…………………………………………………………………………10

4.4 Deployment………………………………………………………………………………………..10

5 Dashboards………………………………………………………………………………………………11

5.1 KPIs (Key Performance Indicators)……………………………………………………….11

6 Conclusion…………………………………………………………………………………………….....12

Abstract

Down Syndrome (DS) is a chromosomal abnormality (Trisomy of human chromosome 21) associated with intellectual disability and affecting approximately 1 in 100 life births world wide. The over expression of genes encoded by the extra copy of a normal chromosome in DS is believed to be sufficient to perturb normal pathaways and normal responses to stimulation causing learning and memory deficits. Although know pharmacatherapies for learning deficits in DS are available, because the incidence is high, their identifications.

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 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

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 Definations

|  |  |
| --- | --- |
| Term | Description |
| Database | Collection of all information monitored by this system |
| Down syndrome | A chromosomal abnormality associated with intellectual disability |
| Trisomy | A type of polysomy in which there are three instances of a particular chromosome, instead of the normal two |
| Genotype | Genetic makeup of an organism |
| Saline | A mixture of salt and water |
| Memantine | A medication used to slow the progression of moderate-to-sever Alzheimer’s disease |

2 General Description

2.1 Product Perspective

To analyze the protein expression dataset on protein influences that could have affected the recovering ability to learn among the trisomic mice.

2.2 Problem Statement

The extra copy of a normal chromosome in Down Syndrome alters the normal pathways and normal responses to stimulation causing learning and memory deficts.

2.3 PROPOSED SOLUTION

The solution proposed here is an mice protein expression model which can be implemented to perform above mentioned use cases. To classify normal and trisomic mice to treat trisomic mice as soon as possible.

2.4 Technical Requirements

This document addresses the requirements to classify normal and trisomic mice at early stage to avoid further suffering.

* The model system should be user friendly.
* The protein samples are given as input.

2.5 Data Requirements

Data requirements completely depend on our problem statement.

* The maximum types of data input which can be given are eight types.
* Which are :

1. c-CS-s: control mice, stimulated to learn, injected with saline

2. c-CS-m: control mice, stimulated to learn, injected with memantine   
3. c-SC-s: control mice, not stimulated to learn, injected with saline   
4. c-SC-m: control mice, not stimulated to learn, injected with memantine   
  
5. t-CS-s: trisomy mice, stimulated to learn, injected with saline   
6. t-CS-m: trisomy mice, stimulated to learn, injected with memantine   
7. t-SC-s: trisomy mice, not stimulated to learn, injected with saline   
8. t-SC-m: trisomy mice, not stimulated to learn, injected with memantine

2.6 Tools used

Python programming language and frameworks such as Numpy, Pandas, Jupyter Notebook, seaborn are used to build the whole model.

* Jupyter Notebook is used as IDE.
* For visualization of the plots, seaborn are used.
* GitHub is used as version control system.

2.7 Constraints

The built model (machine) system must be user friendly, as automated as possible and 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 (2.2 problem statement ) for new dataset that comes through model system which is trained to detect the disease. Machine learning based disease detection model is used 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 indentifying the trisomic mice, we will use a machine learning based model. Below is the process flow diagram is as shown below.

Proposed Methodology

Input (Protein sample)

Training/validat-ion on dataset

ML model for disease detection

Output (trisomic or control)

Prediction of trisomic mice/sample

3.1.1 Model Training and Evaluation

(7) prediction

New Protein samples

Training set

3) Data augmentation

(5) Training

Model

Annoted dataset

Dataset

(1)

Result of the evalution

(2) annotation

Test set

(6) Evaluation

Prediction

3.1.2 Deployment Process

Predicted Result

Make Prediction

Process Protein sample

Get Frames

Load Model

3.2 Event log

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

**Initial Step-By-Step Description :**

1. The system indentifies 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

When the protein sample is given to the model system it will categorize the amount of subset of proteins present in it and the noticed abnormal sample will be declared as trisomic by which the further suffering of the patient or individual can be decreased.

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. Dashboards

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the unveiled problems that if not addressed in time could cause catastrophes of unimaginable impact.

As and when, the system starts to capture the historical / periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors.

5.1 KPIs (Key Performance Indicators)

1. Key indicators displaying a summary of detecting the trisomic mice.
2. To differentiate the given protein sample.
3. Give the output in shorter time.

6 Conclusion

The designed model system will detect the trisomic mice based on various data used to train our algorithm, so we can identify the trisomic mice and treat the patient.

7. References

1. <https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression>
2. Google.com