INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

BEng/BEng.(Hons) in Software Engineering

Final year Project 2014/2015

Interim Report

For

Project Title: **Identify Inherited Diseases based on DNA (IIDDNA)**

By

Iddamalgodage Don Lahiru Manohara - 2010070

Supervised By: Mr. Achala Chathuranga Aponso

………………………..   …..……………..

Signature of Supervisor   Signature of Student

Table of Contents

1. Abstract

In complexity of the genetic inherited diseases, doctors are difficult to pinpoint the correct diseases and diseases related information. Genes are playing major role of the diseases pathogenesis. There are many patterns exists on the genes related on one disease. So it is difficult to classify weather set of genes and their patterns are actually related to the particular disease. To overcome this problem, proposed solution is machine learning based prediction method to identify inherited diseases correctly. IIDDNA is supporting system for doctors and medical parsons to gain knowledge about disease and diseases related information to provide better service to the patients.

2. Aims and objectives of the projects

In TOR (Term of Reference) document mentioned aims are research, design, development and evaluate. Achievements are explains in following table.

Objective/ Aim	Achieved yes/no	Comments
Research	Yes	Research was done by submitting different documents in time to time. Term of reference, literature review.
Design	Yes	Design was done by using different modeling methods and designing classification model.
Development	Yes	Development was done by using different code snippet ruining and creating prediction model.
Evaluate	No	Evaluation and testing not completed. But some evaluation was done by during the implementation stage.
Term of Reference Document	Yes	Term of reference completed and submitted.
Literature Review Document	Yes	Literature review document completed and submitted.
Software Requirement Specification document	Yes	Software requirement document completed and submitted including interview and online survey.
Design Specification	No	The document is not completed but some model diagrams are completed.
Application to demonstrate the solution	Yes	Almost completed but need to add more features on the application.
Prototype evaluation	No	Not completed. But did some evaluation during the development.

3. Main problems were faced during the project

3.1 Finding out correct data set

There are many data source available on the internet but it is difficult to find out correct data set for IIDDNA project. Also it is difficult to find out correct data attributes for particular disease.

Steps of the solution

1. Gather data source related to inherited diseases from the internet.
2. Filtering out data which is most appropriated for the project.
3. Identify the relation between attributes and diseases.

3.2 Difficulty of understanding problem domain

The domain area of the IIDDNA project is difficult to understand. Some concepts of the biology are needed to understand for better outcome of the project.

Steps of the solution

1. Explore knowledge sources are available in the internet.
2. Asking question on the domain experts to solved out the problem.

3.3 Difficulty of creating classification model

There are many classification models available on the machine learning area. Select correct classifier and best fit classifier for given data set are difficult. Some classification models are displaying good accuracy but some are not classify correct disease accurately. Also theories behind the classification model are difficult to understand.

Steps of the solution

1. For checking the accuracy, trial and error methodology used from the data set.
2. To understand classification model, discuss with expert who has a good knowledge in that area.
3. Understating the theories exploring some resources are available on the internet.

3.4 Trying out different implementation approaches

There are many technical ways exists for solve particular problem such as machine learning techniques, expert systems, genetic algorithms, mathematical models. But the challenge is most appropriate techniques select all among the existing technologies. It is very difficult because without testing these methods cannot become a decision.

Steps of the solution

1. Testing different techniques most researchers are suggested.
2. Trying out sample codes from different technologies.

4. Deviation of the project activity schedule

Activity	Start date	End date	Reason	Percentage %
Term Of Reference	08/08/2014	10/06/2014	Submitted on time	100
Literature Review	08/08/2014	12/31/2014	Submitted on time	100
Requirement Analysis	10/15/2014	11/22/2014	Submitted on time	90
Design the prototype	10/28/2014	02/03/2015	Some components are not design yet.	50
Implementing the prototype	11/25/2014	02/24/2015	Some components are not implemented yet. Prediction model is progressing on.	60
Testing	12/15/2014	01/15/2015	Testing was done by during the implementation but did not follow any testing mechanism.	10
Evaluation	01/15/2015	02/21/2015	Some evaluation was done by during implementation. Such as getting feedback from the experts.	10
Final Report	08/08/2014	04/27/2015	Not start yet. But some part of the report is covering from the previously written documents.	10