



University of Reading  
Department of Computer Science

An AI-assisted decision-making system for thyroid nodule  
classification  
Project Logbook

Stefanos Stefanou

*Supervisor:* Huizhi Liang

A report submitted in partial fulfilment of the requirements of  
the University of Reading for the degree of  
Bachelor of Science in *Computer Science*

April, 2021

## 0.1 Introduction

The development of this project was inspired by the Agile development practices, and was split in periods of approximate 1.5 week ( 11 days) each (this is an approximation as this was dynamic, and varied depending on the university workload). In the end of each period or 'spring' a meeting was held with the projects supervisor for an update on the project, as well as for guidance and support. Below there is two sections. One section is a descriptive or brief section that outlines the progress on a particular period, and the second part, is a descriptive part that has details on the development process and the problems encountered.

## 0.2 Brief Table

29/09/2020	Initial comments on PID
6/10/2020	Final Comments on PID (title)
13/10/2020	Literature report and setting goals
20/10/2020	Comments on the literature report/is DDIT a suitable dataset?
27/10/2020	DDIT problems and solutions, next target: TDID and formal classification process
03/11/2020	Week 6, No meeting was held
10/11/2020	Promising results, objective towards service/UI perspective set
17/11/2020	Not Performed Due to personal reasons
24/11/2020	Promising results, Work continuous on schedule
1/12/2020	Not Performed/No updates, Development of the platform continues on schedule
1/12/2020	System development continues on plan with a demo scheduled for the next meeting
8/12/2020	Minor delays in delivery of Frontend, Top priority the presentation and the Frontend
11/01/2021	Frontend is ready, and the majority of the study for start developing backend is done
25/01/2021	Part of the backend is ready, development continues on plan, demo of the completed features
01/02/2021	The information backend is ready. Profile/Notifications/Login/Support NINO
08/02/2021	Migration FE-BE ready, meeting with researcher held, SVC model ready
16/02/2021	Full FE cycle + use cases ready on schedule. Task-Backend(Predictor) starts on schedule.
23/02/2021	Meeting with doctors from RBH. Much appreciation and potential. Great ideas for future work
29/03/2021	Task-Backend(Predictor) ready on schedule. SVC Works, issue with ResNet migration.
12/04/2021	ResNet migration completed and software ready. Extensive testing and presentation left
19/04/2021	Testing and video presentation completed successfully. Final progress meeting for TMC Project.

## 0.3 Analytic Report

### 0.3.1 29/09/2020

Some advice on good report writing is given by the supervisor, Mrs. Liang. Additionally, some typos regarding the PID draft and some clarification on the various subsections of the PID Template are also given. Some questions on context and expectations are also answered.

### 0.3.2 6/10/2020

Some final comments on PID regarding the title and the type of the system (web or standalone) are given. The PID is now at the 'Ready' state and gets approval from Mrs. Liang.

### 0.3.3 13/10/2020

Some information regarding the literature report is discussed, and short-term goals are set. The first goal is to perform some minor transformations into the DDIT dataset and building the first classifier in python and various technologies. The study-first-write-later approach to this project is agreed upon. Under this approach, a period of non-rapid progress will exist. The approach, as mentioned earlier, will allow the student to properly study the underlying subject and its mathematical formalism before engaging with the topic.

### 0.3.4 20/10/2020

Some guidance regarding the literature report and how it needs to be more 'research-based' was given. Additionally, we discussed the importance of proving that DDIT is a suitable database for our needs. Plan on how we will prove it based on a simple python-classifier. This goal is to be met by our next meeting.

### 0.3.5 27/10/2020

The first results regarding the viability of the DDIT dataset are discussed. The integration library is tested, and the first classifier using scikit-learn is implemented. 97/98 of samples can be successfully predicted using this approach. Some significant problems are also pointed out, as the inconsistency of the dataset in certain areas and the very few valid points that can be used. The fact that maybe more datasets are needed is pointed out. The next target is defined. Next week, the TDID dataset needs to be tested, and the formal classification process needs to be followed, splitting our data points in an 80-20% setup.

### 0.3.6 03/11/2020

Week 6, No meeting was held.

### 0.3.7 10/11/2020

Opening, the results of the two previous weeks are being presented.

- Integrate TDID into the codebase.
- 80-20 setup : 75-80 %accuracy.
- Study: Complete Revision of Probability and Statistics.
- Progress on the study of the topic.
- Five papers were found, studied, and added to the literature report.
- TIRADS idea, multiple labels classification. Is it viable?

After the initial presentation, some feedback in the process is given, and the next objectives are set, namely...

- Try different classifiers, with different hyperparameters, compare results.
- Multiple label classification. Compare results.
- Introduction of complex image filters. Compare results.
- Paper prototype, what are the features that we want to support?
- Define a draft of Use cases, Interactions.
- Service perspective of the project, maybe a mock of UI.
- What are the end-users?

### 0.3.8 17/11/2020

No meeting was held due to personal reasons.

### 0.3.9 24/11/2020

Completed objectives are stated in the following list

- Try different classifiers with different hyperparameters, compare results.
- Paper prototype, what are the features that we want to support?.
- Define a draft of Use cases, Interactions.
- Service perspective of the project, maybe a mock of UI.
- What are the end-users?.

Not completed objectives are stated below.

- Multiple label classification. Compare results.
- Introduction of complex image filters. Compare results.

Some notes from the supervisor.

- Paper prototype: align with the universitys report template
- Define a draft of Use cases, Interactions : Write for those in the report
- Service perspective of the project: Architecture okay, create a mini presentation for the doctors, feedback is needed
- What are the end users: only hospitals? needs to be investigated.

Next time objectives

- Align with the universitys report template.
- Mini presentation.
- Write about the user interactions.
- Write about the softwares architecture in the report.
- Minimal functional version of the software.
- Multiple label classification. Compare results.
- Introduction of complex image filters. Compare results.

### 0.3.10 1/12/2020

No Meeting was held, because of no further updates. System develepoment continues on plan with a demo scheduled for the next meeting

### 0.3.11 8/12/2020

Objectives

- (1)Align with the universitys report template (DONE)
- (2)Mini presentation (To be completed)
- (3)Write about the user interactions (DONE)
- (4)Write about the softwares architecture in the report (DONE)

- (5) Minimal functional version of the software (In progress)
- (6)(carryover) Try different classifiers ,variance of with parameters, compare results (Postponed)
- (7)(carryover) Multiple label classification. Compare results (Postponed)
- (8)(carryover) Introduction of complex image filters. Compare results (Postponed)

Because of university deadlines, the development of a minimally functional version of the software is postponed until the 22nd of December. Presentation to be completed by the next meeting, Tasks (1),(3),(4) are completed, and tasks (6),(7),(8) are postponed until next year. Top priority, the delivery of the frontend for the next few weeks.

### 0.3.12 11/01/2021

Things Achieved on Christmas

- Minimal functional version of frontend

The front end is ready, and most of the study to start developing the backend is done. For the next few weeks, the first microservice(backend-1) will be developed. with the new scheduling, tasks(6),(7), and (8) will be postponed until after the end of the development of backend-2.

### 0.3.13 25/01/2021

Part of the backend is ready, develepoment continues on plan, 20 Tasks are complete, demo of the following

- Login System
- Signup System
- Profile Endpoint
- Django endpoint
- Heroku deployment
- Split into multiple projects for ease of use and proffesionalism

The following plan was agreed

- 20-30 Jan: Backend Microservice 1 (Information) + Migration FE/BE
- 1-10 Feb: Backend Microservice 2 (Task)+Migration
- 10 Feb-15 March: Study of machine learning patterns + University Deadlines
- 15 March - 15 April: Complex models,image filters+prediction focused period
- 15 April-20April: Presentation work and safety padding

### 0.3.14 01-02

- All the required functionality of Information Backend is ready(demonstration performed),
- Part of FE-BE Migration is ready. Develepoment continues on plan.
- Profile Section Implemented.
- Notifications System Implemented.
- Login System Implemented.

- Necessary study materials completed.
- Support NINO on patients.

Targets for next week

- Information Backend Ready
- Migration Complete
- The basic model to be runnable, and shuffling mechanism to be implemented and tested.
- PostDoc meeting to be held next week with advices on improving the model
- Doctors meeting to be scheduled

### 0.3.15 08-02

- Information Backend is ready on schedule
- Migration FE-BE is ready on schedule(Demonstration performed)
- The basic model to be runnable, and shuffling mechanism implemented
- PostDoc meeting held. Introduction to the system, we discussed various approaches on enhancing the system
- Doctors meeting for feedback to be done in 2 weeks time

Some notable tasks following

- Comments on patient (comments on scan will follow later)
- Search by nino
- About page
- Title+logo
- Error handling
- Refactoring
- Patient add
- Patient list
- Notifications pill
- Profile username upper left
- Contract testing
- Shuffling mechanism on the basic model
- + 2 bugs fixed

For Next week

- Task-Backend to be implemented
- Task-Backend - FE Migration to be performed
- Finalization of user interactions
- Start working on the Doctors presentation
- Start working on the Project poster presentation for February 26th

### 0.3.16 16-02

#### 25 Completed Tasks

- Full Cycle is implemented
- All interactions are implemented
- Comments on Patient and on Scan
- Patient View and Scan view
- View All scans for a given patient
- Form alignment and Table alignment, all is uniform
- Notification system completed
- Change ST marys hospityal to RBH

#### Whats Left

- Image upload and storage on TE (mocked)
- Classification Handling on TE (will use pykka actors for those)
- Putting an actual model
- Studing my last 2 books
- Complex models migration
- Update the final report

How? from now untill 29th march, will go to low-productivity mode 1) Will finish poster video 2) Will finish presentation 3) Will Implement 1,2,3,6,4

Those to be expected by 29th march; by implementing those, i will have three weeks for complex model and finish the project(manageable)

What low-productivity means? instead of  $5*7 = 35\text{hrs/week}$   $2*7 = 14\text{hrs/week}$

Some Suggestions have been given by Mrs Liang

- Implement Hint's for making the system user friendly
- Better Logo and Naming
- Agree/Disagree functionality

Those Have been agreed to have been implemented by the next week. Additionally, the next week, a doctors meeting will be held, and I will present my work for receiving useful feedback

### 0.3.17 23-02

The meeting with the doctors was an absolute success, where the potential of the project was outlined. Some useful feedback and relevant software was mentioned for me to be inspired and get ideas. Great for future development of the project.

### 0.3.18 29-03

52 Completed Tasks, most notable changes include...

- Image upload and storage on Task Backend (Base64 Encoding)
- Classification Handling on TE (will use pykka actors for those)
- Putting an actual model, the SVC Model works.
- The necessary books has been studied.
- Complex model migration still under construction
- Work on final report

Numerous issues were discussed, the most difficult one is the fact that the complex model, provided by the post-doc researcher uses the torch package. py-torch(the python version of torch) is a very heavy library in terms of storage. Heroku(our deployment service) has a maximum limit of the size per project(500mb). A workaround may be possible, if the more lightweight cpu-only version is used, i will have a look at the next week for that. Work on final report continuous on schedule.

### 0.3.19 12-04

36 Completed tasks. most notable changes include

- Complex model works according to plan.
- Work on final report continuous on schedule.

The problem with the maximum size per project was resolved by creating a different requirements.txt, one for my local machine(Mac OS) and one using a virtual machine for Linux x86-64 (using VirtualBox). The plan worked perfectly, and now it uses the -CPU packages instead of GPU-CUDA ones. this needs to be fixed if this project tends to go on production, probably we can deploy this service separately into a dedicated server with GPUS available for the models to utilize. For the final demonstration, the following things needs to be done.

- Final E2E testing.
- Extensive writing of unit tests.
- Video script and video editing of the final presentation.

### 0.3.20 19-04

The final meeting before the delivery of the video presentation. Some minor changes were proposed by Mrs. Liang. Implemented, and the video was submitted on the 19th.