Capstone Completion Plan

# Milestone 1: Project Proposal and Requirements Analysis

The first milestone requires revision in terms of its focus from general cancer care to breast cancer. Besides this, several updates to the success measures, solution, project cost and schedule are also required. This is detailed below:

## Project Overview and Project objectives

There is no need to update the problem and background as breast cancer is a subcomponent of the general problem that is cancer. In terms of the project objectives, it terms must be updated as the data collected are not mere images, but rather DICOM files that contain both images and patient metadata (data such as the patient’s sex, age, etc.).

## Project Scope

The scope will be set back to fulfill the guidelines. Although the desired result would be to classify the images solely on the image data, there are other factors that influence the outcome of a diagnosis and can assist the model to make better and more accurate predictions. The stakeholders table shall also be expanded to include the radiologist in charge of examining the images and passing the images through the application for model predictions. The Work Breakdown Structure also requires update to the status column, the actual completion column, and the resource column.

## Project Success Measures

For this part of the milestone, only one more assumption will be included. This assumption pertains to the data gathered within the header files of the DICOM files.

## Project High-Level Solution

No changes are to be made to this section.

## Project Controls

The risk management table requires the inclusion of overfitting from model parameters as well as the model memorizing the data due to the lack of samples. The roles and responsibilities will also be expanded to include the radiologist, the data analyst, and the data engineer which may oversee furthering the data pipeline which in turn will distribute the data for the data analyst to train the model and update the application.

## Project Cost and Schedule

The project schedule hours will have to be reworked together with the resource column. Other aspects of the project, such as the page design will have to be included within the Work Breakdown Structure table

## Timeline for Completion

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task | Start Date | End Date | Days | Time (hours) | Done? |
| Project Overview and Project Objectives | | | | |  |
| Update objectives to represent DICOM files | 11/3/2022 | 11/4/2022 | 1 | 1 | Yes |
| Project Scope | | | | |  |
| Change scope to use image data AND header file data. | 11/3/2022 | 11/4/2022 | 1 | 1 | Yes |
| Add more stakeholders into table | 11/3/2022 | 11/4/2022 | 1 | 1 | Yes |
| Update the status of the work breakdown table | 11/3/2022 | 11/4/2022 | 1 | 1 |  |
| Project Success Measures 1 Hour | | | | |  |
| Include assumption about header files | 11/3/2022 | 11/4/2022 | 1 | 1 | Yes |
| Project Controls 4 Hours | | | | |  |
| Include overfitting to risk table | 11/4/2022 | 11/5/2022 | 1 | 1 | Yes |
| Include low sample size to risk table | 11/4/2022 | 11/5/2022 | 1 | 1 | Yes |
| Expand roles and responsibilities table | 11/4/2022 | 11/5/2022 | 1 | 2 | Yes |
| Project Cost and Schedule 17 Hours | | | | |  |
| Rework the schedule hours | 11/5/2022 | 11/6/2022 | 1 | 1 | Yes |
| Include the page design in work breakdown table | 11/5/2022 | 11/6/2022 | 1 | 3 | Yes |
| Total | 11/3/2022 | 1/6/2022 | 3 | 13 | Yes |

# Milestone 2: Data Pipeline

The data pipeline design summary together with the data pipeline architecture will require further redesign to better fit the new updates done to the application as the scope has been shrunk from highlighting the tumor, identifying the stage, and determining whether the tumor is benign or malignant to simply determining whether the tumor is benign or malignant based on the image and metadata provided.

## Design Planning Summary

The project deliverables of this section will need to be changed to better reflect the current scope of the project (which is to simply diagnose whether a tumor is benign or malignant).

## Overview of Model Pipeline Design

This section requires a single update on the encoding component of the algorithm. Rather than converting the categorical data for training into numbers, an encoder from TensorFlow’s platform was used to encode the parameters.

## Detailed Model Pipeline Design

The Detailed overview section will require further updates as the scope of the project has been changed to create a model that only diagnoses a tumor as either benign or malignant. Besides this, the hardware technology used must be updated to fit the latest update to the RTX 3090TI GPU.

## Timeline for Completion

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task | Start Date | End Date | Days | Time (hours) | Done? |
| Design Planning Summary | | | | |  |
| Set the scope to only diagnose tumors based on whether tumor is malignant or benign | 11/6/2022 | 11/7/2022 | 1 | 1 | Yes |
| Overview of Model Pipeline Design | | | | |  |
| Update description of encoding | 11/6/2022 | 11/7/2022 | 1 | 1 | Yes |
| Detailed Model Pipeline Design | | | | |  |
| Change Hardware GPU to latest. | 11/6/2022 | 11/7/2022 | 1 | 1 | yes |
| Total | 11/6/2022 | 11/7/2022 | 1 | 3 |  |

# Milestone 3: Implementation

The main issue now with the Implementation section is to join it together into a singular document. Despite this, there are some modifications to be made in regard to the implementation plan, Mapping of Functional Requirements, and the Source Code Listing

## Mapping of Functional Requirements

The functional requirements only need the addition of the online aspect of the application. There are certain functions which provide the background strength to the application and allows data to be parsed and displayed in an orderly manner.

## Source Code Listing

Currently the documentation for the source code listing has been automated using sphinx. The only step left is to reprint the new source code listing and add it to the Milestone document.

## Implementation Plan

This will require the largest change as the Tkinter library will no longer be in used in exchange for the Dash library to create a web app where the user will upload the DICOM files for making predictions.

## Timeline for Completion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Start Date | End Date | Days | Time (hours) |
| Mapping of Functional Requirements | | | | |
| Add web-app implementation to the functional requirements | 11/7/2022 | 11/8/2022 | 1 | 3 |
| Source Code Listing | | | | |
| Update documentation to include web design. | 11/8/2022 | 11/9/2022 | 1 | 1 |
| Implementation Plan | | | | |
| Remove Tkinter from the plan | 11/8/2022 | 11/9/2022 | 1 | 1 |
| Implement the use of Dash in plan | 11/8/2022 | 11/9/2022 | 1 | 2 |
| Total | 11/7/2022 | 11/9/2022 | 2 | 7 |

# Milestone 4: Performance Analysis and Presentation

Although some concepts in testing, such as the confusion matrix and making predictions on a sample that the model has never seen before is part of the testing process, no automation has been developed for it yet. Therefore, most of the 4th Milestone is still required to finish the Capstone project.

## Testing

For this section, the documentation for the testing will have to be developed. Although some testing has been planned, it has not been fully conceptualized and implemented yet. Documentation for the testing also requires development

## Module Test Cases

The test cases will be automated using python’s unittest or doctest libraries. As an alternative to the built-in python libraries, pytest can also be utilized to test the application. The main purpose of these tests is to observe whether the independent functions are functioning as intended. To add on to the modular test format, the components together towards building a small model and finally the test the model’s ability to make predictions. The main purpose of the last test it confirms whether the procedures are repeatable using the same algorithms.

## Project Completion Phase

The presentation is yet to be created. The presentation will be composed of the initial presentations created for the project proposal with the addition of the benefits that the solution may provide.

## Project Presentation Phase

The project presentation will be completed using screencast software for approximately 7 to 10 minutes. The presentation is almost ready, the only part of the presentation that is not ready yet is the solution component which requires explanation and re-scoping to better fit the current project.

## Final Project Submission

An online version of the project is already available at the following site:

[wpeguero/capstone: Capstone Project information. This contains the algorithm used for the Master's in Data Science Thesis. (github.com)](https://github.com/wpeguero/capstone)

The only requirement on this side is to update it by following git protocols.

## Timeline for Completion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Start Date | End Date | Days | Time (hours) |
| Testing | | | | |
| Create Separate Test document | 11/8/2022 | 11/9/2022 | 1 | 2 |
| Create plan for testing | 11/8/2022 | 11/9/2022 | 1 | 2 |
| Module Test Cases | | | | |
| Create test case for model prediction | 11/9/2022 | 11/10/2022 | 1 | 2 |
| Create test case for the ETL pipeline | 11/9/2022 | 11/10/2022 | 1 | 2 |
| Create test case for the model training | 11/10/2022 | 11/11/2022 | 1 | 3 |
| Project Completion Phase | | | | |
| Develop presentation of solution | 11/10/2022 | 11/11/2022 | 1 | 1 |
| Project Presentation Phase | | | | |
| Download and install screencast software | 11/11/2022 | 11/12/2022 | 1 | 1 |
| Create script for presentation | 11/11/2022 | 11/12/2022 | 1 | 2 |
| Screencast and begin presentation | 11/11/2022 | 11/12/2022 | 1 | 1 |
| Final Project Submission | | | | |
| Update the Github site with upgrades | TBD | TBD | TBD | TBD |
| Total | 11/8/2022 | 11/12/2022 | 4 | 16 |