FIT3161 Project Proposal

Developing a New Method for Diagnosis the Breast Cancer in Mammography Images

Semester 1, May 2021

Segmentation of Tumour and Extracting Useful Features From Mammography Images

Group Name : MA_21

Project Supervisor : Dr Golnoush

Team Members :Daniel Kee

Jaclyn Neoh

Tan Sook Mun

TABLE OF CONTENTS

1. Introduction

According to the national cancer institute, breast cancer is classified as one of the most common cancers for women. (National Cancer Institute,n.d.). Doctors usually recommend women from the age of 40 to have regular annual screening with mammograms for breast cancer. This practice is vital because early detection and mitigation can help avoid the patient regressing into late-stage cancer. One of the prevalent practices in the medical industry to detect breast cancer is by a radiologist manually checking the mammography for breast cancer. However, this method leaves room for human error. The radiologist may develop fatigue or lack attention to detail after many repetitions. An error in diagnosis is extremely dangerous and may be fatal for patients. Considering the problem statement above, we hope to address this issue by developing a new method for diagnosis by applying the approach of Computer-Aided Diagnosis (CAD). With this new approach we hope to reduce human error and improve the accuracy of tumour detection in mammography images.

The main objective of the project is to develop a new method for the diagnosis of breast cancer in mammography images. In this project, there are a few distinctive phases in this project. Firstly, the pectoral muscle has to be removed from the mammography. This is a crucial step because essentially it helps prepare the image for the next phase.
Next, the breast tumour will be segmented out from normal tissue and the features of the tumor will be extracted. In the last phase, there will be a classification process by using machine learning. This phase will decide whether the tumour is benign or malignant.

The phase our project is working on will be the second phase. We will have to propose an accurate segmentation algorithm to segment out the abnormal tumour from normal breast tissue in the mammography image and extract any useful features of the tumour. The outcome of this project would be a diagnosing system that takes in a dataset of mammography images with the pectoral muscle removed. Next, segmentation and feature extraction methods will be applied to the images. Lastly, there will be an output of useful features of the tumour and a confusion matrix to show the accuracy of the applied segmentation and feature extraction algorithms.

2. Literature Review

2.1 Introduction

The image processing of mammography images to find tumors have been studied and researched over time, and there are many methods created and implemented. Image processing generally consists of a few parts. These include preprocessing of the image, segmentation of the image, extracting features from the image, in other feature extraction, and postprocessing of the image. Currently, our main focus of our project is to implement an algorithm for segmentation and feature extraction, preprocessing our image along the way if needed. Our objective is to take note of all the research papers, methods and implementations found that expounds on our main focus, analyze its advantages and disadvantages, benefits and limitations, then propose a modified and optimized algorithm based on the studies we found.

2.2 Pre-processing

Pre-processing is a process of preparing an image to go through the segmentation and feature extraction process. Morphological operations combined with the resulting mammography image from running the original image with Otsu's thresholding will be used to remove the artefact from the original mammography image.

According to Chudasama et al. (2015), "The base of the morphological operation is dilation, erosion, opening, closing expressed in logical AND, OR notation and described by set analysis". Its advantages are it can enlarge pixels of interest in an image and remove 'noise' in an image. However, in some cases, it may distort the original information illustrated in the original image.

For the image that resulted from Otsu's thresholding of the mammography image, it will be part of a process which includes subtracting this resulting image from the original image.

2.3 Segmentation

Segmentation involves bringing out the region of interest; in our case, bringing out from the mammography image any abnormal shape which could be a tumor. One of the methods of segmentation is Optimized Region Growing, with a dynamic selection of threshold using the Dragonfly Algorithm.

Region growing is a process of examining neighbouring pixels of initial seed points and to determine whether these neighbours will be considered as a single region. The process will be a repeated process until a certain condition is met, then it comes to a halt.

The initial seed point will be determined using the Dragonfly Algorithm. The Dragonfly Algorithm is "based on the two different types of swarming behaviors of Dragonflies in nature", and "balances the phases of exploration and exploitation by imitating the natural

swarm interaction of dragonflies for navigating, food search and enemy avoidance" (Díaz-Cortés et al., 2018). The advantage of using the Dragonfly Algorithm for the Optimized Region Growing is that the threshold selection, which is the seed point, will be dynamic. In other words, the seed point will not be a constant value or hard-coded for every image.

Furthermore, the other proposed segmentation method is known as Hammouche's Algorithm. It involves wavelet theory and Genetic Algorithm.

A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. It can be used to extract information from many different kinds of data, not just images. One of the examples are audio signals. In one of the proposed segmentation methods, Hammouche's Algorithm, the corresponding gray level histogram will be obtained from a mammography image and will be decomposed using a wavelet to extract useful information. The wavelet used is 'db2'. Wavelet theory is very well known and is referenced in research papers often. Gao et al. (2020) used a wavelet transform-based image segmentation method, using the "Haar wavelet" to choose a threshold, and then use that threshold to mark the region of image segmentation. Also, Gavlasov´a et al. (2006) uses the Haar wavelet as well for image decomposition.

Genetic Algorithm consists of "randomized search and optimization techniques guided by the principles of evolution and natural genetics, having a large amount of implicit parallelism", according to Maulik et al. (2000). This involves having a search space, with its parameters encoded in the form of strings called chromosomes (Maulik et al., 2000). A group of them are called populations, and based on the principle of survival of the fittest, a few of the strings are selected and each is assigned a number of copies that go into the mating pool. This is repeated for generations until a termination condition is satisfied. This whole process imitates the whole biological idea of genetics. In this proposed segmentation method, a chromosome will be selected in a similar way. The chromosome is represented as a binary string, and the threshold values needed as part of the image segmentation process is found at the index positions where the bits are 0.

2.4 Feature Extraction

The feature extraction process will be implemented using two separate methods, both unrelated to each other. The first method implemented uses wavelet theory, similar to the wavelet theory method in Hammouche's algorithm above; however, the wavelet used here is "db4". The grounds for using wavelet theory is similar to that of Hammouche's algorithm.

Feature extraction for this method of implementation consists of obtaining the wavelet decomposition structure of the segmented mammography image, extracting the coefficients from it and performing calculations on it like normalization and energy computation. The number of features extracted will then depend on the number of energy values per feature.

Furthermore, the other feature extraction implementation involves the GLCM. According to Gadkari (2004), GLCM "contain(s) information about image texture characteristics such as

homogeneity, gray-tone linear dependencies, contrast, number and nature of boundaries present and the complexity of the image". The advantage of GLCM is that many statistical information can be obtained from just a single function that gives a GLCM of a certain segmented mammography image.

2.5 Related Work

In this section, we will talk about the findings and analysis of the different research papers we found.

Pereira et al. (2014) used Top-hat morphological operation & Otsu's thresholding for their preprocessing on mammography images. By applying this operation on a mammography image only the artifact of the original image has been shown. Taking the original image and subtracting from it the artifact-only image gives a mammography image without the artifact. They used Hammouche's computational algorithm (wavelet theory and genetic algorithm (GA) for the segmentation. They used a semi automatic approach and it had 2 stages where the detected region shape was compared with the original image and the area of region was verified, taking into account that the information extracted from both mammographic views. The second step was done manually.

Domingue et al. (2009) used dynamic programming (D2PBT) and constrained region growing as the implementation of image segmentation. The resulting image in then run on the feature extraction algorithm which is not specified, giving 6 features: contrast between foreground region and background region, coefficient of variation of edge strength, two measures of the fuzziness of mass margins, a measure of spiculation based on relative gradient orientation, and a measure of spiculation based on edge-signature information

Tsochatzidis et al. (2021) implemented a segmentation algorithm using convoluted neural networks (CNN), and it is know as the U-net architecture, well known for medical image segmentation.

SIngh et al. (2020)'s implementation of image segmentation of mammography images includes two stages. The first stage is to segment the breast tumor ROI using a conditional Generative Adversarial Network, then classify its binary mask using a CNN based shape descriptor. After that, a CNN is used to learn features from the segmented image to represent objects at different scales and orientations.

Berner et al. (2013) has 3 proposals for image segmentation, the watershed segmentation, level-set segmentation and the breast mass contour segmentation, the preferred segmentation algorithm.

Wajid et al. (2015) has a preprocessing method that i includes breast image segmentation and filtering, which is followed by image normalization to improve image quality and reduce noise, then apply the Contrast Limited Adaptive Histogram Equalization (CLAHE). Furthermore, a feature extraction algorithm is implemented using LESH. LESH works by calculating a histogram of the local energy pattern within the image.

Mousa et al. (2005) implements preprocessing in a few stages. The first stage is image pruning, which is just cropping out the artefacts from the mammography image. Then, global gray-level thresholding is implemented, with the upper threshold as 240 and the lower threshold as 120. After that, histogram equalization is performed. After the preprocessing step, feature extraction is performed using wavelet decomposition. The wavelet 'db4' is used. The whole process consists of extracting coefficients from the wavelet decomposition structure, then performing normalization, then the energy computation. The number of features will depend on the number of energy values per feature.

Chowdhary et al. (2020) lists out segmentation methods in categories, namely obsolete methods, ancient methods and recent methods. Obsolete methods include Deformable models, Multiresolution method, Coupled Surface and Geodesic Minimal Path. Ancient methods include Thresholding, Markov Random Field Approach, Graph Cut, Appearance Model, Automatic and Semi-automatic Segmentation, Class based Segmentation, Target Tracking and Atlas based Segmentation. Recent methods include Active Shape Model, Segmentation using Artificial Intelligence and Clustering. Furthermore, this research paper also lists out feature extraction methods in categories, namely gray-level based, filter-based and component analysis based. Gray-level based includes Local binary pattern, Gray level co-occurrence matrix, Gray level run length method and Harlick features. Filter-based includes Gabor texture features, Learning vector quantization and Symbolic dynamic filtering. Component analysis based includes Principal component analysis and Independent component analysis.

Punitha et al. (2018) implements preprocessing using Gaussian filtering. Then, the segmentation method used is optimized region growing with initial seed point chosen using the dragonfly algorithm. For feature extraction, features are extracted using GLCM.

2.6 Conclusion

The final product of our project will be a user interface where users can input the mammography image they want to analyze, then choose a segmentation algorithm out of 2 algorithms and output the final result, and then if desired, choose a feature extraction algorithm out of 2 algorithms and output the final result. Overall, we will definitely preprocess the image first, then run the desired segmentation algorithm, then run the desired feature extraction algorithm.

3 Project Management Plan

Below are the details of our project management plan.

3.1 Project Overview

The objective of this project is to propose a new segmentation technique to segment out the tumour in a mammography image, as well as displaying the features that are extracted based on the segmented image.

This project consists of major work activities including segmenting the tumour out from the mammography image as well as extracting the features of the tumour so that the tumour in the mammography image can be classified as benign or malignant. A user interface is also required as the final product of this project will be an executable file that can read the mammography images, segment the tumour in the mammography image and extract the feature of the mammography image.

3.1.1 Milestones

The major milestones of our project consist of:

1. Pre-processing the data set
2. Implementation of the segmentation algorithm
3. Implementation of the feature extraction algorithm
4. Documentation of the finding and results of the output
5. Creation of evaluation metric
6. Testing and getting the results of the segmentation and feature extraction algorithm
7. Designing and creating the user interface
8. Making the prototype of the system
9. Final testing of the system.
10. Documentation and pitching of the project

3.2 Project Scope

Below is the breakdown of the project scope. It is broken down into the deliverables, product characteristic and product user acceptance criteria

3.2.1 Deliverables

In this section contains 2 tables which are project deliverables and product deliverables

Table 1 - Project Deliverables Table

Project Deliverables
Team Contract
Work Breakdown Structure
Meeting Minutes
Requirements Traceability Matrix
Risk Register
Scope statement
Project Timeline
Project Design
Budget Plan
Project Proposal

Table 2 - Product Deliverables Table

Product Deliverables
Collection of research papers for project reference
Summary/analysis of different segmentation and feature extraction algorithm
Collection of mammography images with the pectoral muscle segmented out
Blueprint of algorithm
Algorithm code
Test cases for algorithm
Output and confusion matrix of successful segmented of mammography image
System Diagnosis with interactive UI

3.2.2 Product Requirements

This section contains the table for Functional and Non-functional Requirements. The requirements traceability matrix of the project is appended in the appendix under Appendix A of the report.

3.2.2.1. Functional and Non-functional requirements

Table 3 - Functional and Nonfunctional Requirements Table

Category	Requirement
Functional	The segmentation method is used to segment out the tumor from normal breast tissue
Functional	The pre-processing and post-processing of the images need to be applied to significantly reduce noise and prep the image to apply feature extraction
Functional	The feature extraction method is used detect the features of the tumour
Functional	Output of confusion matrix and features extracted to see the accuracy and data of tumor
Non- Functional	The segmentation method is being accurately segment out the tumor from breast tissue
Non- Functional	The feature extraction method accurately detect the segmented tumor features
Non-Functional	Create an intuitive UI for the diagnosing system, so a user manual will be unnecessary.
Non-Functional	There is the availability of multiple segmentation and featuring extraction methods.

3.2.2.2. Requirements Traceability Matrix

The requirements traceability matrix of the project is appended into the Appendix A of the report.

3.2.3 Product Characteristics

The final product will be a breast cancer diagnosing system. It will have a UI that allows users to select the different segmentation methods and another option to select a feature extraction method. The user will have to select or insert the dataset that they want. Next, the user will select the segmentation and feature extraction method. Finally there will be an output of segmented mammography images and an evaluation matrix that shows the accuracy of the method. There will also be a table of features extracted from the images.

3.2.4 Product user acceptance criteria

The table below is the product user acceptance criteria.

Table 4 - Product User Acceptance Criteria Table

Requirement	User Story	Acceptance Criteria
The segmentation need to accurately segment out the tumor from normal breast tissue	As a user I want the tumor to be accurately segmented out	Before performing segmentation process, the images will be pre-process so the images will be clearer and easier to segment out the tumor
The feature extraction process need to provide sufficient details of the tumor	As a user, i want to be able to know accurately the features of the tumor	Before performing feature extraction, the images will be post-process to reduce noise and increase accuracy
The output of the system need to contain a confusion matrix to show the accuracy rate	As the developer I want to be able to know the accuracy rate of the algorithm applied	The system need to output a confusion matrix to show the accuracy of the algorithm applied
The UI of the system must be easy to use and without bugs	As a user, I need the UI to be intuitive and easy to use	The system has an easy and intuitive UI
The diagnosing system needs to be able to be given an input	As a user, I want to be able to choose or insert the dataset I want	The system as a input bar which enables user to input or select from existing dataset

3.3 Project Organisation

This section contains the breakdown of the project organisation. It consists of the process model of the project and the responsibilities of each member.

3.1 Process Model

Our group decided collectively to use a predictive approach for this project. One of the reasons is because the project requires a lot of research and analysis on segmentation and feature extraction methods before starting to design and propose our algorithm. It is vital that each stage of the project has to be completed before moving on to the next stage. Furthermore, our group had clearly identified the objective and consulted with our project supervisor and concluded that the requirements are unlikely to change during the project lifecycle. This also gives the reason why the agile methodology is not required for our project. Agile methodology is based on the project requirements and is constantly changing. Below shows the detailed breakdown of the project framework.

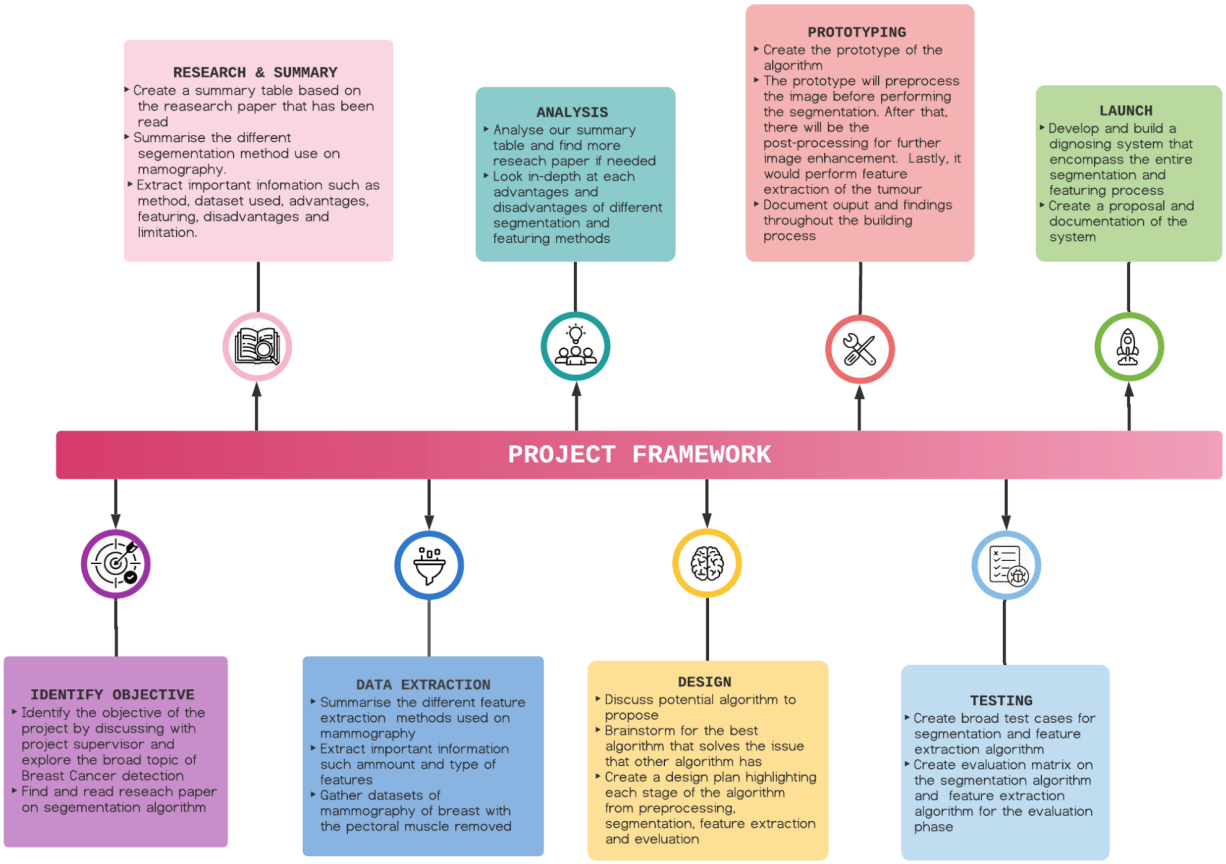


Figure 4 - Project Framework

The diagram above shows the overall framework of the project. The entire project can be summarised into eight different stages. The early stages of the project is to identify the objective of the project. After consulting with the project supervisor, it is clear that the main objective of this project is to propose an accurate segmentation algorithm to segment out the tumour and a feature extraction algorithm to extract useful features of the tumour. We also have to build a diagnosing system to run our algorithms. The bulk of the project relies heavily on research and analysis. We will have to do an in-depth research on different segmentation and feature extraction algorithms that have been proposed in the industry. We will then extract any useful information from the research papers we have read. Consequently, we will have to find datasets of mammography with the pectoral muscle removed. With all the data we have gathered, we will need to analyze all the information. After these phases are done, only then we will be able to design our algorithm. During the prototyping stage is where we will be building the algorithm and diagnosing system. Before launching the project, we will have to apply extensive tests onto our diagnosing system to ensure it works well. Lastly, will be our launching of the product. In this stage we will have to complete the documentation and the pitching of the project.

3.2 Project Responsibilities

Below has 3 table detailing the team members and each of their responsibilities

Table 5 - The Team Table

The Team	
ROLES	NAME
Project Manager	Tan Sook Mun
Quality Assurance	Jaclyn Neoh
Technical Lead	Daniel Kee

Table 6 - Project Responsibilities Table

Project Responsibilities	
Functions	Member Accountable
Compiling Dataset	Daniel Jaclyn Sook Mun
Applying Pre-processing	Daniel
Proposing Segmentation Algorithm	Sook Mun
Applying Post-processing	Jaclyn
Proposing Feature Extraction Method	Sook Mun
Testing and evaluating algorithm	Jaclyn
Designing UI of Diagnosing System	Daniel
Preparing Project Documentation	Daniel Jaclyn Sook Mun

Table 7 - Team member Contribution for Project Proposal Table

Team Member's Contribution for Project Proposal	
Introduction	Sook Mun
Literature Review	Daniel
Project Overview	Jaclyn
Project Scope	Sook Mun
Project Organisation	Sook Mun
Project Management Process	Jaclyn
Schedule and Resource Requirement	Sook Mun
External Design	Daniel
Proposed Methodology	Sook Mun
Test Planning	Jaclyn
Conclusion	Daniel
Style and Presentation	Daniel Jaclyn Sook Mun

3.4 Project Management Process

This section is dedicated to managing the processes in a project. We will look into Risk Management as well as into monitoring and controlling mechanisms which contain Stakeholder and Communication plan and Review and audit mechanisms.

3.4.1 Risk Management

Project risk management is a very important step in the project management process. It can help save your project when the risk turns into a problem as you can save a lot of time planning how to combat it especially during the development stage.

As this project is a rather new experience for the 3 of us, it is quite likely that the number of risks that can occur during the implementation phase is rather high. Therefore, we took the chance to identify, analyse and manage the risk factors by brainstorming on possible risks between ourselves, as well as together with our supervisor.

The risk register of our project based on all the risks that we identified can be found in the appendix B.

3.4.2 Monitoring and Controlling Mechanisms

It is very important to monitor the progress of the project throughout the duration of the project to ensure that this project will be a success. We will start the monitoring and controlling process by creating a Stakeholder Management Strategy shown below to monitor all of the stakeholders as the plan provides guidelines on how to engage each stakeholder involved in this project.

Table 8 - Stakeholder Management Strategy

Stakeholder/Role	Level of Interest	Impact	Potential Management Strategy
Supervisor	High	High	Dr Golnoush is very responsive to the emails that the team has sent to clarify some doubts and to seek help. She has met the team throughout the planning stage and guided the team to understanding the project in greater detail. Have a lot of meetings with her to keep her updated on our progress as well as sending her our progress through email as well.
Project Manager	High	High	Sook Mun is very committed to the team and has shown a lot of interest in this topic. Have the team meetings to keep the team updated on her progress as well as to give her the platform to address any issues she faces. Other than through email, this can also be done on Whatsapp. The team's kanban board and the Gantt Chart will also be useful to give her a clear vision of the task.
Quality Assurance	High	High	Jaclyn is very devoted to the team. Have the team meetings to keep the team updated on her progress as well as to give her the platform to address any issues she faces. Other than through email, this can also be done on Whatsapp. The team's kanban board and the Gantt Chart will also be useful to give her a clear vision of the task.
Tech Lead	High	High	Daniel is very committed to the team as shown time and time again. Have the team meetings to keep the team updated on his progress as well as to give him the platform to address any issues she faces. Other than through email, this can also be done on Whatsapp. The team's kanban board and the Gantt Chart will also be useful to give him a clear vision of the task.

Furthermore, in order to monitor the progress made by each team member, each team member is required to fill up a Weekly Progress Report by the end of each week. This is to assist in ensuring that sufficient progress is made to ensure that the project will be completed on time. The Weekly Progress Report will be reviewed by the Project Manager each week after each team member provides a brief presentation of their Weekly Progress Report during the second meeting of the week. The Project Manager will then take appropriate measures to keep the team in check if needed. Examples of measures that can be taken consist of having an emergency meeting to discuss the root of the delay, giving the team a prep talk and reassigning the tasks. Another way to track the team's progress is through the team's kanban board as well as the Gantt Chart. A template of the Weekly Progress Report can be found in the appendix as Appendix C as well as the Gantt Chart as Appendix D and a screenshot of the kanban board the team is using as Appendix E. A template of our Meeting Minutes can also be found in Appendix F.

In addition, communication between the team as well as with the supervisor is not only very important in monitoring the progress made, but also in ensuring the success of the project. We will also need to review and audit throughout the project to improve the quality of our project.

3.4.2.1 Stakeholder and Communication plan

Communication is key to every project's success. To ensure that the team stays on track, each team member is required to communicate with the rest of them when they have any doubts, face any problems or require any assistance. This can be brought up through a quick communication medium like Whatsapp or during the team meetings. The team will also be meeting up with the team's supervisor, Dr Golnoush, to give a constant update on progress of the project. The meeting with her also creates the opportunity for clearing any doubts, seeking her guidance as well as receiving her feedback. Another tool that can be used to communicate with her is through email. The communication metric is as follows:

Table 9 - Communication Matrix

Communication Types	Objective of Communication	Medium	Frequency	Audience	Owner	Deliverables	Format
Kick-Off/ Project Briefing	Provide a brief introduction to the project topic	Zoom	Once	Project Team and Project Supervisor	Project Manager	-Agenda -Meeting Minutes	Soft copy saved on Google Drive
Weekly status report with supervisor	Provide update to the Project Supervisor on the progress of the project	Zoom	Once a week	Project Team and Project Supervisor	Project Manager	-Agenda -Meeting minutes	Soft copy saved on Google Drive
Team Meetings	Provide an update on the progress made. During the second half of the second meeting of the week, there will be a quick recap session of the week	Zoom	2 times a week	Project Team and Project Supervisor	Project Manager	-Agenda -Meeting Minutes -Notion	Soft copy saved on Google Drive
Test Plan Meeting	Review Test method and results	Zoom	When needed	Project Team	Quality Assurance	Code Report	Report saved in Google Drive
Code Review	Review the efficiency and structure of the code	Zoom	When needed	Project Team	Tech Lead	-Code Base -Code Report	Code pushed into Git Report saved in Google Drive
Final meeting	To close the project	Zoom	One	Project Supervisor and project team	Project Manager	All document	All document of the project in the Google Drive

3.4.2.2 Review and audit mechanisms

It is really important to review and audit throughout the project in order to ensure the quality of the software deliverables. In this section, we will talk about quality assurance, versioning control and documentation.

In terms of quality assurance, the Quality Assurance is assigned to carry out code reviews whenever it deemed necessary. She is also responsible for preparing test cases to test the algorithm. This will help us in the long run as it will help ensure that the accuracy of the final algorithm is high, as it helps us spot certain cases that can cause our algorithm to fail. At the same time, the evaluation metric will be created to test the accuracy of the algorithms.

On the other hand, in terms of versioning control, we will be using GitLab as a version control tool for our code, as we can easily look at the previous version of the code as well as who made the changes. Gitlab also allows us to collaborate with each other when writing the code. In terms of documentation, all documentation will be stored in Google Drive. Each document needs to have their version number in their title as well as the type of document they are.

Last, in terms of documentation, the code itself will have comments so that the person who looks at our code can understand what the algorithm is doing. We will also prepare a code report which provides further information about the diagnosis system. We believe that is sufficient documentation needed for our software. We do not foresee anyone needing any training if they were to take over our project as it is rather straightforward unless they do not have any prior experience with Matlab.

3.5 Schedule and Resource Requirements

In this section shows the detailed breakdown of the schedule and resource requirements

3.5.1 Schedule

Deadline of the project: October 2021

Expected duration of the project: 3 months

The required tasks that must be completed in order to satisfy the project agreement is as below :

1. Compile datasets of mammography images with the pectoral muscle segmented out
2. Propose segmentation and feature extraction method
3. Algorithm Testing
4. Creating an evaluation matrix and format of output

5. Developing UI

7. Documentation

The broken down process and activities of the project is depicted in the form of a Work Breakdown Structure (WBS) as Appendix G and a Gantt chart as Appendix D which are appended in the appendix of the report.

3.5.2 Resource Requirements

Below detailed the resource requirements of the project. The skills set requirements are the list of skills required for the project to be completed. The Hardware specification is the hardware we need to implement the project. The software specification tables list the software needed to complete the project.

Table 10 - Skill Set Requirements

Skill Set Requirements	
Programming knowledge in Matlab	The code for the entire segmentation and featuring process is coded in Matlab
Knowledge of Image Processing	Every members need to have in-depth knowledge of image processing because it is the main framework of the project

Hardware Requirement

Table 11 - Hardware Specification

Hardware Specification	
Operating System	Windows 10 Home
CPU	Intel i5
GPU	GTX 1080
RAM	8GB
External Hardware	Hard Disk 1Tb

Software Requirement

Table 12 - Software Specification

Software Specification	
MATLAB 2020B	This software is use for running, creating and testing our segmentation and featuring algorithm
Image Processing Toolbox	Image Processing Toolbox provides built-in functions for image processing,algorithm development, visualization and more. This toolbox is required to build our segmentation and feature extraction algorithm.
Matlab Compiler	Matlab Compiler is an add-on app that enables you to package and deploy your Matlab programs into executable files, web applications and more. We will need this to deploy our diagnosis system at the end of the project.
Statistic and Machine Learning Toolbox	This toolbox provides functions that describe, analyse and model data. This toolbox will be used to code out our classification model to evaluate our segmentation and feature extraction algorithm for testing.
GitLab	This software is used for version control of our codes. Each member can contribute to writing the code by pushing and pulling the project code base. With this tool we can see each change and version of our codes.
Notion	This software is used for project management.Is has a collaborative to-do list, kanban board and a roadmap which features the timeline of the project
Whatsapp	This is used for day to day informal communication
Google Drive	This is used for storing all our project files (assignments,documents,summary and etc)

4. External Design

The following is what our final product user interface would look like and what users who use our application would see:

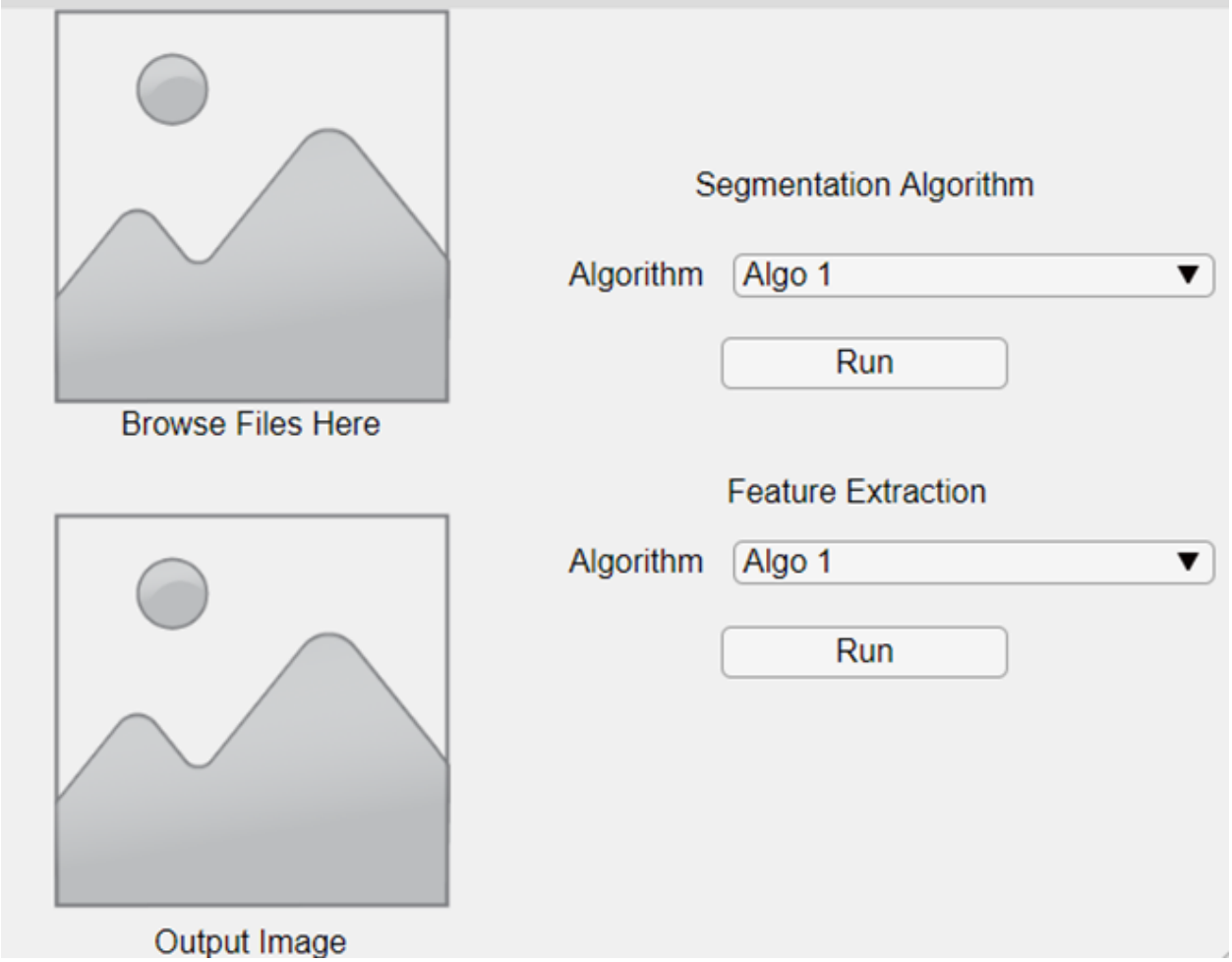


Figure 5 -User Interface

The top image is the place where users enter their mammography image, and they can choose what segmentation algorithm they want and click run, and the output image will appear at the image position at the bottom left. If desired, they can also choose a feature extraction method and click run after running segmentation, then the output image will again appear at the bottom left part. The assumption of this user interface is that users are assumed to run the segmentation algorithm first before running the feature extraction algorithm.

Furthermore, we also have a separate user interface solely for us developers to test whether our segmentation and feature extraction algorithms are working well or not. The user interface is as follows:

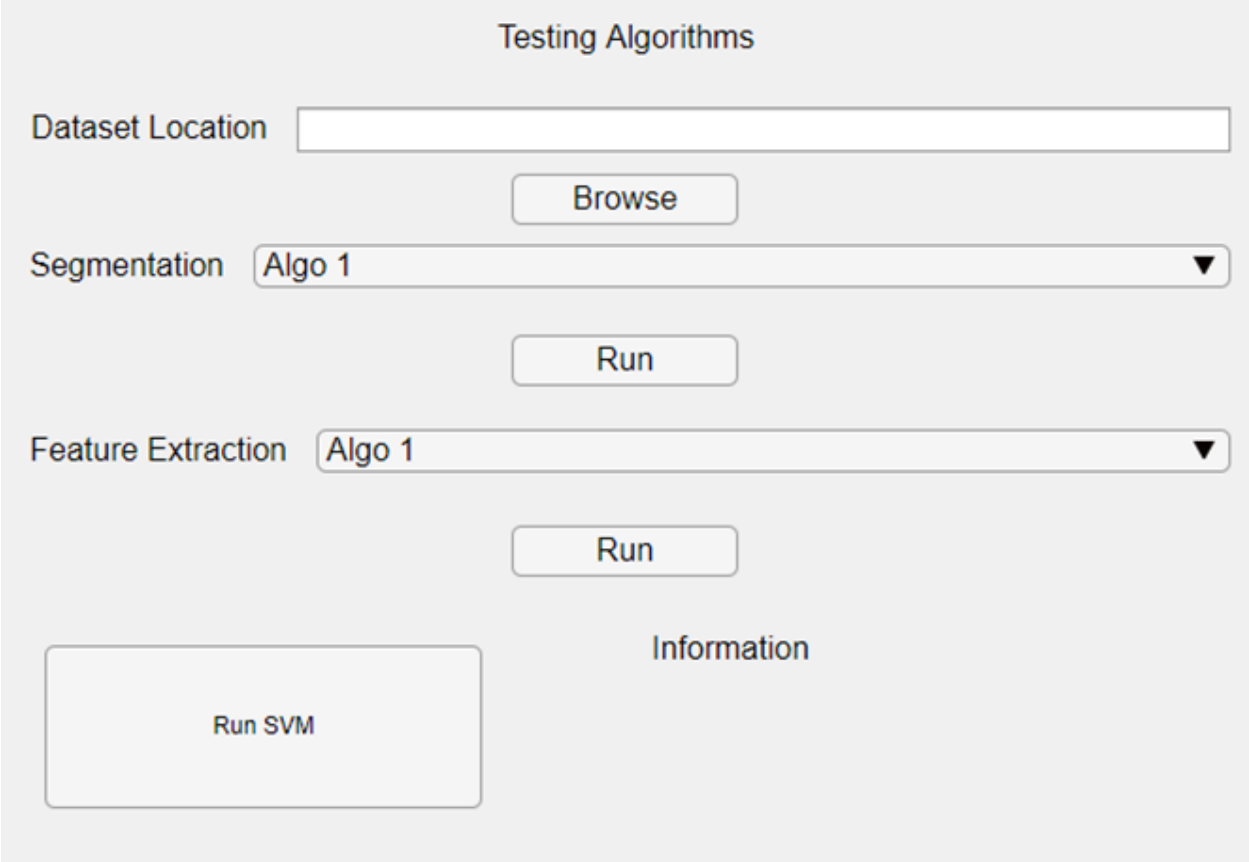


Figure 6 -Developer Interface

The Dataset Location input is for inputting a dataset of mammography images. We can then run for each image the segmentation algorithm and feature extraction. For a training set, each image will have labels to inform whether the image has a tumor or not, and whether it is malignant or benign. Running SVM on the training set will give information in the information section regarding accuracy, sensitivity, specificality etc., relative to the labels on each training set image.

Once the training set images, after running segmentation and feature extraction algorithms on it and running an SVM on it to say that its accuracy is good enough, we build a classifier out of it. We then use this classifier to classify our test set (which has already gone through the segmentation and feature extraction algorithms), then output the information to inform us whether our segmentation and feature extraction algorithm works or not.

5. Methodology

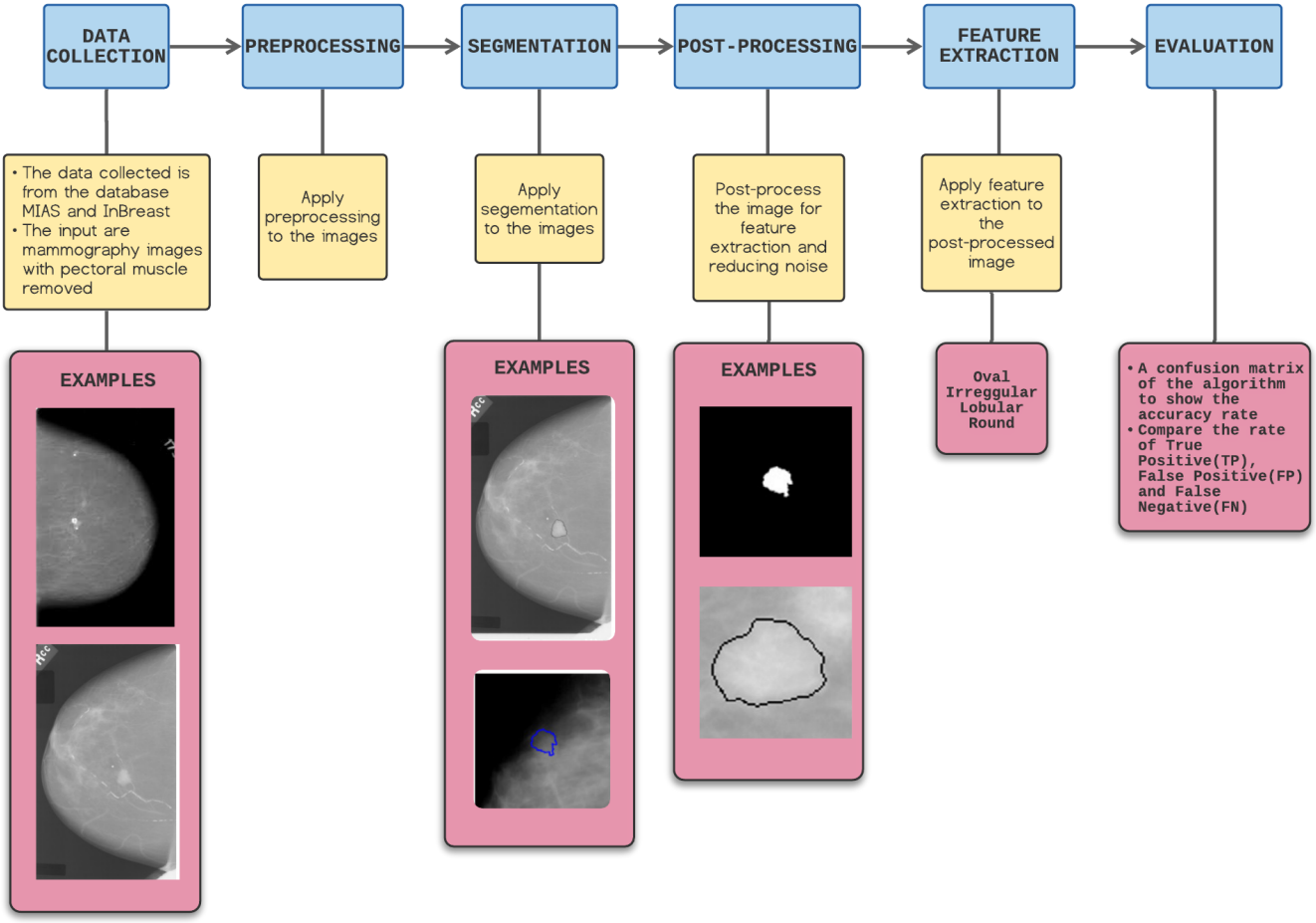


Figure 7 - Algorithm Framework

The diagram above represents the algorithm framework of our proposed methodology. After comprehensive research and analysis, our group proposes there should be 6 stages in the algorithm framework. The initial stage would be data collection. The algorithm will only accept input that consists of mammography of the breast that has the pectoral muscle removed. Next would be the pre-processing phase of the mammography images. This has to be done in-order to increase the accuracy of the segmentation algorithm that will be applied afterwards. The segmentation phase will segment out the tumor from the breast tissue. Succeeding this stage would be the post-processing stage. This stage is crucial because it eases the feature extracting process and decreases noise in the image. Then, we will perform feature extraction onto the post-processed image. This stage will extract any useful features of the tumor. Lastly, we will evaluate our algorithm by

producing a confusion matrix. This is so we are able to identify the accuracy rate of the algorithm applied.

This project also consists of building a diagnosing system that consists of multiple segmentation and feature extraction algorithms. After exhaustive research to find the best and most comprehensive methods we decided on 2 segmentation and 2 feature extraction methods.

For segmentation methods, we have chosen to implement optimized region growing, selecting threshold using the dragonfly algorithm, and Hammouche's algorithm. Region growing is like a clustering algorithm, where neighbouring pixels from an initial seed point are examined and evaluated on whether they should be added into the region. As for Hammouche's algorithm, it uses wavelet theory to break down the gray level histogram of the mammography image, then find the threshold levels using a chromosome encoded as binary.

For feature extraction we had chosen Gray-Level Co-Occurrence Matrix (GLCM) and wavelet theory. The approach of the GLCM method is to outline the texture of the image by calculating the pairs of pixels with specific values and spatial relationships in the image. The rationality of this decision is because for GLCM, it is useful to have data on contrast, correlation, energy and more in order to inspect the mammography image. Consequently, the reasoning behind choosing wavelet methods is because we can analyze the image without any loss of the original data. In our case, we would want to retrieve what our mammography image is trying to show, which is the shape of the tumor.

6. Test Planning

This section is divided into 3 sections which are Test Coverage, Test Method and Test Cases.

6.1 Test Coverage

Quality of dataset

We need to feed some dataset into our algorithm to test the quality of our segmentation and feature extraction algorithm. Therefore, we need to ensure that our dataset has good quality to reduce the percentage of false negatives(detected correctly but labeled false). We also need to ensure that the images in our dataset have no pectoral muscle.

Implementation of the segmentation and feature extraction algorithm

The main feature of our diagnostic system is to segment out the tumour as well as extracting its feature. As it is extremely crucial that the result produced by our

segmentation algorithm as well as the feature extraction algorithm be accurate, we need to ensure that there is sufficient testing.

6.2 Test Methods

Quality of dataset

We will be using mammography images that do not contain pectoral muscle as it is not in our scope to remove the pectoral muscles out.

Implementation of the segmentation and feature extraction algorithm

To measure the performance of the segmentation algorithm and feature extraction algorithm, the following steps will be taken:

1. Split the dataset into 2 parts, one for training the model and one for testing the model.
2. Use a built-in machine learning algorithm to classify the tumour as malign or benign.
    a. As classification of the tumour is not in our scope, we will be picking one classification method to classify the tumour for the testing.
    b. SVM will be used for the classification.
3. Create a confusion matrix to compute the accuracy, precision and recall.
4. Plot a ROC curve to compute the benefit of using the model over not using it.
5. Calculate the Area Under the Curve(AUC) to check if there is any discrimination.

6.3 Test Cases

We can verify that our algorithms for segmentation and feature extraction works through the use of SVM, where we will run a training set through our segmentation and feature extraction algorithm, then use SVM to acquire information like accuracy, sensitivity, specificality, etc., which is then compared to the information given in the labels of each image of the training set. We can then use SVM on the training set to acquire the same information, and compare it to the information of the training set to see whether our segmentation and feature extraction algorithm works.

7. Conclusion

In conclusion, our project to build a diagnosing system is feasible. The diagnosing system will allow users to choose segmentation and feature extraction methods they want to apply to their dataset of mammography images with the pectoral muscle removed. The users also have at least 2 feature extraction and 2 segmentation methods to choose on. We also plan extensive testing of our algorithms. We have a developer UI that is meant to train and test the algorithm accuracy. In short, this is only the tip of the iceberg on the subject of image processing of mammography images. There are many more algorithms that are more well known and there are still some newfound algorithms being researched today. This implementation of ours for segmentation and feature extraction of mammography images to detect tumors in breasts are based on these researches and studies, and we hope that this whole project will bring some benefit to us and to the community.

8. Appendix

This link contains the high definition images and documents of the appendix https://drive.google.com/drive/folders/1ov65etB5QP1bLrQ4t0_JzZcRaXpuAh6h?usp=sharing

Appendix A- Requirements Traceability Matrix

REQUIREMENTS TRACEABILITY MATRIX					
ID	*Requirements (Functional or Non-Functional)*	*Assumption(s) and/or Customer Need(s)*	*Category*	*Source*	Status
1	Compilation of mammogram images with pectoral muscle segmented out	Need images to run and test the the algorithm on. The image could also be used to help us analyse existing algorithm and gain more insight on them	Functional	Dr Golnoush	Completed
2	Blueprint for building propose algorithm	A rough design plan of our algorithm. This is to show the logic and thought put into our propose segmentation	Functional	Dr Golnoush	Completed
3	Build a diagnosing system that incorporates segmentation and feature extraction algorithms	The diagnosing system need to allow user to apply segmentation and feature extraction onto their chosen dataset	Functional	Dr Golnoush	In Progress
4	The diagnosing system has a few options of different segmentation techniques	The user is able to choose different segmentation methods to suit their needs	Non-Functional	Dr Golnoush	In Progress
5	The diagnosing system has a few options of different feature extraction techniques	The user is able to choose different feature extraction method to suit their needs	Non-Functional	Dr Golnoush	In Progress
6	Test cases for the algorithm	There need to be test cases for the propose algorithm to ensure it is accurate and working properly	Functional	Dr Golnoush	In Progress
7	Build a evaluation metric to evaluate the algorithm	By building an evaluation metric, we can test our algorithm to see it's performance and think of ways to imporve the accuracy	Functional	Dr Golnoush	In Progress
8	Create an intuitive UI for the diagnosing system, so a user manual will be unnecessary.	Customer need to be able to easily use the diagnosing system	Non-Functional	Dr Golnoush	In Progress
9	Write a documentation detaling our project	Our assumption is that this happens after the algorithm is evaluated. This report will be read by other researches or doctors that want to implement or gain insights on our proposed algorithm.	Functional	Dr Golnoush	In Progress

Appendix B- Risk Register

No.	Rank	Risk	Description	Category	Triggers	Root Cause	Potential Responses	Risk Owner	Probability	Impact	Status	Score
	1	similar open source code found	useful reference of the algorithm can be found on the internet	positive	doing research for own project and stumble across it	Algorithm is famous and commonly used	Use them as reference to see what we can improve upon, citing them in our work too	Technical Lead	8	8	Open	64
	2	Dataset assessability	We may not be able to access the dataset we planned on using if the source's decided not to provide free access to it anymore	Data science	No access to the dataset	Dataset is no longer public	Find another dataset	Quality Assurance	3	10	Open	30
	3	Supervisor may be busy to meet	Our supervivor may not be able to meet with the team when needed	Schedule	Our supervisor writing us an email that suggest that she couldn't make it to the meetings	Due to personal reasons or due to their workload	Reschedule the meeting if the supervisor could not attend the meeting. Another alternative is to communicate to email	Project Manager and Project Supervisor	2	9	Open	18
	4	Team members may fall sick or received side effects from taking the Covid vaccination	If one of the member is sick, that person has to take time-off to recover. Durring this period, Malaysia is badly affected by Covid-19 and not only that, we have to take into account of that our members might receive side effects from the Covid-19 vaccination	Schedule	The members informing the rest that they are ill	Illness which includes getting Covid-19, side effects from the Covid-19 vaccination and other illnesses	Assign the task to another teammate	Project Team	3	5	Open	15
	5	Project is delayed due to the change of requirements	Unplanned new work/requirement that must be accommodated/change of direction	Schedule	Our supervisor informed the team that there is an update on the scope	Specifications not specific enough in the first, resulting in when it is being implemented many problems come out	Rescheduling of task.	Project team and project supervisor	2	7	Open	14
	6	Financial attraction of project to investor	Because this project could helped save lifes, we might attract outside investors	Positive	Potential user/investor looking for algorithms on segmentation and feature extraction of mammography images	The success of the project	Go into discussion with potential users/investors, further improve and research on how to improve our application and algorithms	Project Team	1	8	Open	8

Legend	Numerical
high	7
medium	5
low	3
Very low	1

Appendix C - Weekly Progress Report

WEEKLY STATUS REPORT

Name: Week/Date:

What have you completed?

Are there any obstacles hindering your progress?

Deadlines

☐ On Track ☐ At Risk ☐ Off Track

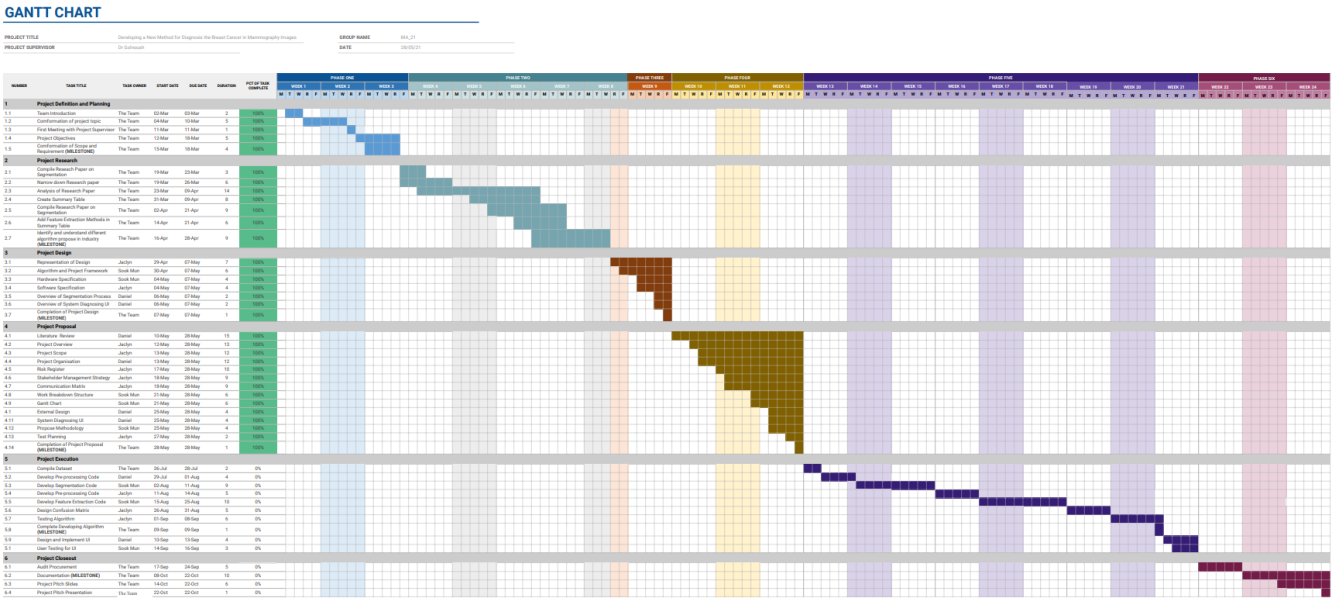
Do you need any clarification on your task?

☐ yes ☐ no

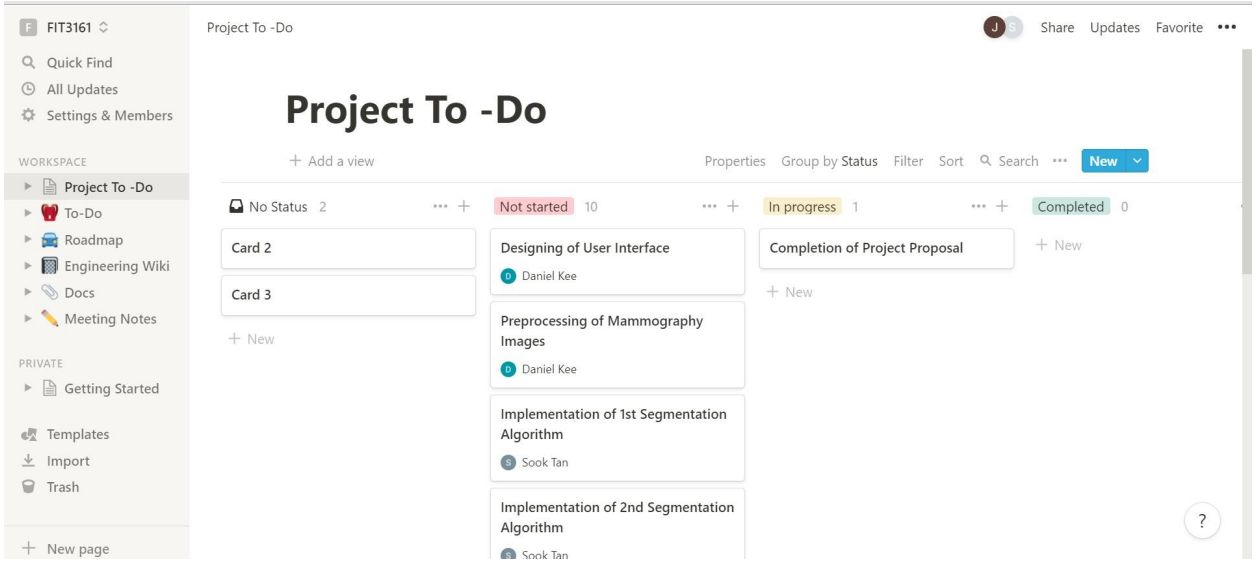
If yes, what would you like to clarify?

After your current task, what is the next task you will complete?

Appendix D - Gantt Chart



Appendix E - Kanban Board(Notion)



Appendix F - Meeting Minutes Template

Meeting Minutes

Meeting no:

Description: -

Date:

Time:

Location:

Attendees:

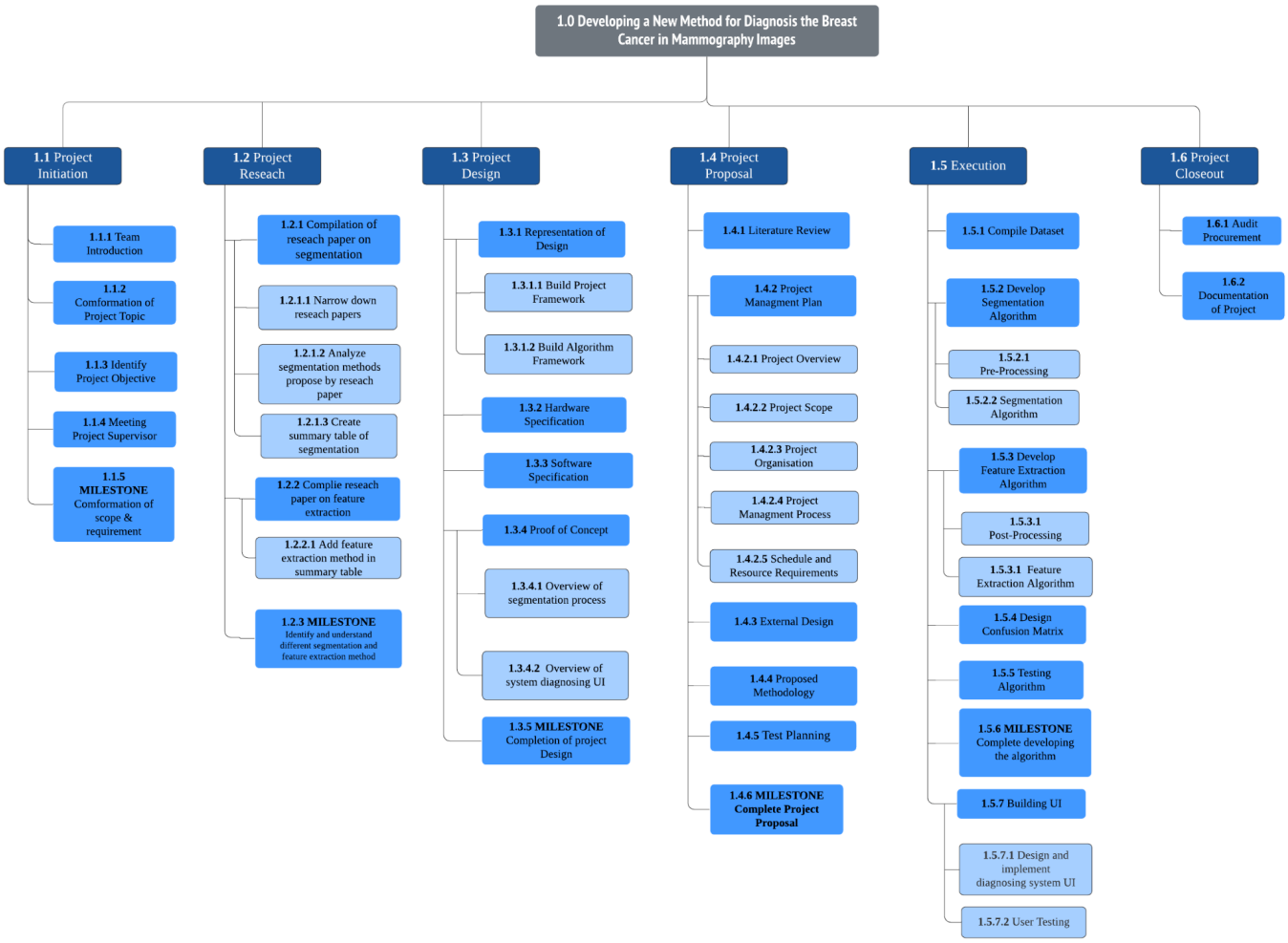
Absent: -

Chairperson:

Minutes taker:

Item No.	Item	Info (I) or Action Item (A)	Person in charge (PIC)	Due date	Comments

Appendix G - Work Breakdown Structure(WBS)



9. Reference

Berber, T., Alpkocak, A., Balci, P., & Dicle, O. (2013). Breast mass contour segmentation algorithm in digital mammograms. Computer methods and programs in biomedicine, 110(2), 150-159.

Chowdhary, C. L., & Acharjya, D. P. (2020). Segmentation and feature extraction in medical imaging: a systematic review. Procedia Computer Science, 167, 26-36.

Chudasama, D., Patel, T., Joshi, S., & Prajapati, G. I. (2015). Image segmentation using morphological operations. *International Journal of Computer Applications*, *117*(18).

Díaz-Cortés, M. A., Ortega-Sánchez, N., Hinojosa, S., Oliva, D., Cuevas, E., Rojas, R., & Demin, A. (2018). A multi-level thresholding method for breast thermograms analysis using Dragonfly algorithm. *Infrared Physics & Technology*, *93*, 346-361.

Domínguez, A. R., & Nandi, A. K. (2009). Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. Pattern Recognition, 42(6), 1138-1148

Gadkari, D. (2004). Image quality analysis using GLCM.

Gao, J., Wang, B., Wang, Z., Wang, Y., & Kong, F. (2020). A wavelet transform-based image segmentation method. *Optik*, *208*, 164123.

Gavlasová, A., Procházka, A., & Mudrová, M. (2006). Wavelet based image segmentation. In *Proc. of the 14th Annual Conference Technical Computing, Prague* (pp. 1-7).

https://www.cancer.gov/types/common-cancers(retrieved April 2021)

Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern recognition*, *33*(9), 1455-1465.

Mousa, R., Munib, Q., & Moussa, A. (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert systems with Applications, 28(4), 713-723.

Pereira, D. C., Ramos, R. P., & Do Nascimento, M. Z. (2014). Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm. Computer methods and programs in biomedicine, 114(1), 88-101.

Punitha, S., Amuthan, A., & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique. Future Computing and Informatics Journal, 3(2), 348-358.

Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., ... & Torrents-Barrena, J. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Expert Systems with Applications, 139, 112855.

Tsochatzidis, L., Koutla, P., Costaridou, L., & Pratikakis, I. (2021). Integrating segmentation information into CNN for breast cancer diagnosis of mammographic masses. Computer Methods and Programs in Biomedicine, 200, 105913.

Wajid, S. K., & Hussain, A. (2015). Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. Expert Systems with Applications, 42(20), 6990-6999.