Final Project Report

Segmentation of Tumour and Extracting Useful Features From Mammography Images

Group

MA_21

Project Supervisor

Dr Golnoush

Team Members

Daniel Kee

Jaclyn Neoh

Tan Sook Mun

Table Of Contents

1.0 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.). As a result, physicians routinely advise women above the age of 40 to perform an annual breast examination. It is a vital habit that women have to adopt in order to ensure early detection and mitigation is done to prevent patients regressing into late-stage cancer. Breast cancer screening is done by a radiologist manually scrutinizing the mammography images for any signs of tumor or abnormalities. As a result, human error can occur during this process. A wrong diagnosis can have disastrous consequences. The radiologist may develop fatigue or lack attention to detail after many repetitions.For this reason, physicians are exploring different methods to help improve the overall process of breast screening. Following this, the main approach this paper is exploring is creating a new method for diagnosis by adopting the approach of Computer-Aided Diagnosis (CAD).

In this project, there are 3 main stages to this new method of diagnosing breast cancer. The first stage is to remove the pectoral muscle of the breast mammography image. This stage essentially is to ready the image for the next step. Stage two is the segmentation and feature extraction of the tumor in the mammography image. Finally, the last stage is to apply classification and predict what is the diagnosis of breast cancer. This phase will determine the type the tumour is which is benign or malignant.In our project, we focus on the second stage. This is where segmentation and feature extraction method.

The team comprises three members. The members are Daniel Kee, Jaclyn Neoh and Tan Sook Mun. The role of each member is technical lead, quality assurance and project manager. The project supervisor for our team is Dr Golnoush. Our team has created a breast cancer diagnosing system. Basically, the stage two that was mentioned above is encapsulated into a single system that is accessible by an easy to use UI. This system allows users to apply different segmentation methods and also choose breast cancer images that the user wants to predict. The entire document is dedicated in to going in depth to describe the process and end product of this project

2.0 Literature Review

The literature review below is taken from the project proposal we had written in Semester 1 2021. We had added more insights and discussion that we had during the implementation phase.

2.1 Introduction

Tumors in mammography images produced from scans alone might be able to be seen with the naked eye, neither can we determine whether it is cancerous or not. Therefore, image processing is needed for processing the image to a point where the tumor is spotlighted and statistics and features can be extracted from it. The general process of image processing is as follows: pre-processing, segmentation and feature extraction. Below are the details of what was implemented.

2.2 Pre-processing

Pre-processing is a process of preparing an image to go through the segmentation and feature extraction process. A morphological operation, the "Opening And Closing" algorithm, was used to pre-process the images. As the name implies, opening was first applied where "opening basically opens up strips; it generally smoothen the contour of an image by breaking necks and eliminating thin protrusions" and closing was applied to resulting image where "Closing basically closes gaps; it tends to smooth contour sections, but as opposed to opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes, and fill gaps in the contour". (FIT3081 Week 5 Lecture 1).

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. The segmentation methods that we used in our project are Region Growing and Hammouche algorithms.

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 was hard-coded to be 180 as it produces the best results out of the other values we tested.

Furthermore, the other implemented segmentation method is known as Hammouche's Algorithm. Wavelet theory and Genetic Algorithm is used when implementing this algorithm. To give some background, a wavelet is an oscillation, and its amplitude starts and ends at 0, ascending then descending. Images are not the only artifact that it can be used on to extract information out of. Next, 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). Generally, the algorithm simulates the process of genetics, and it uses the principle of survival of the fittest in certain factors in the segmentation process, and produces the most optimized segmented mammography image.

2.4 Feature Extraction

The feature extraction process will be implemented using two separate methods, both unrelated to each other. The first method implemented is the LESH algorithm.

The LESH algorithm works by converting an image into a combination of local energies along different orientations. As stated in the name, it will be shown as a histogram, and each bin is a local energy level. The values will be used for training in the SVM classification.

Other than that, GLCM is implemented as our second feature extraction algorithm. The characteristics that we extracted out of the method are Contrast, Correlation, Energy and Homogeneity. GLCM is available as a built-in function in MATLAB.

2.5 Related Work

There are many other methods of image segmentation and feature extraction of mammography images. Here, we discuss the different possibilities of algorithms that we can use to segment and feature extract our mammography images had we not chosen our current algorithms.

Pereira et al. (2014) used Top-hat morphological operation & Otsu's thresholding for their preprocessing on mammography images. This preprocessing method is mainly used to remove unwanted objects that are not part of the mammography image. For example, some mammography images might have a tag that includes information about the individual that mammography image came from. Before any image processing, this has to be removed, and therefore, this method is useful for removing those tags.

Other pre-processing methods in global gray-level thresholding, as used by Mousa et al. (2005), as well as a method called histogram equalization. Other than that, Gaussian filtering could be one of the options for pre-processing too.

In segmentation, Domingue et al. (2009) used dynamic programming (D2PBT) and constrained region growing as the implementation of image segmentation. Convoluted neural networks were used in Tsochatzidis et al. (2021) image segmentation method. They call it the U-net architecture, and it is renowned for its use in medical image segmentation.

Furthermore, a CNN is also used to extract features from segmented images for which objects could be shown at different orientations and scale, based on the paper from SIngh et al. (2020).

In the list of methods used by Wajid et al. (2015), they use a Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve the image quality after segmentation.

Other segmentation methods discussed are as follows. The list includes Deformable models, Multiresolution method, Coupled Surface and Geodesic Minimal Path. Newer 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. More recent methods include Active Shape Model, Segmentation using Artificial Intelligence and Clustering. This list of methods are provided by Chowdhary et al. (2020).

Other feature extraction methods of note by Chowdhary et al. (2020) are the Gray level run length method and Harlick features. This is also known as gray level based feature extraction, according to the author. The author also lists out filter-based feature extraction methods, which include Gabor texture features, Learning vector quantization and Symbolic dynamic filtering. Furthermore, the author also lists out component analysis based feature extraction methods, which include Principal component analysis and Independent component analysis.

2.6 Conclusion

A user interface will be the final outcome of our project. A user will input a mammography image of their choice, and the system will run image segmentation and feature extraction on it, and then predict the condition of the tumor, whether it is cancerous or non-cancerous. If the tumour is cancerous, the system can also output the type of cancer for instance benign or malignant. There will also be an admin user interface which will be solely used by the software developers. It will run a SVM classification on a dataset of mammography images and then outputs the confusion matrix and accuracy to help us make sure our algorithms implemented are truly working reliably and correctly. We also allowed the developer to choose the percentage of dataset they want to change. During the implementation process, we want to allow the developer to have as much control as possible.

3.0 Outcomes

3.1 What has been implemented

We implemented the "Opening and Closing algorithm" which is a morphological operation as our pre-processing algorithm. This algorithm helps remove structures yet fill holes without affecting the remaining parts of our input images.

We also implemented 2 segmentation algorithms, which are the Region Growing algorithm and the Hammouche algorithm. The segmentation algorithms are used to segment the tumour in the pre-processed dataset. As stated in the literature review, region growing is a segmentation algorithm which examines the neighbouring pixels of initial seed points and determines 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, whereas, Hammouche algorithm combines both the wavelet theory and the genetic algorithm.

Moreover, we also implemented 2 feature extraction algorithms which are GLCM, a feature extraction algorithm that "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" (Gadkari (2004)), and LESH algorithm which is an algorithm that works by converting an image into a combination of local energies along different orientations. After feeding the segmented images into the feature extraction algorithm, the feature extracted in the form of a table will be passed to the SVM classification model to predict if the input mammography images is cancerous or not, and if so, what kind of tumour it contains.

Furthermore, we also implemented the UI for the system. We used a built- in tool in MATLAB to design the UI. In the system UI there are 3 interactive pages. The first page is the main page. In this main page, it welcomes the user and allows the user to choose which interface they want to go next. Users are able to choose to go to the Admin page or the User page. Each page has their own set of functionality and purpose. In the admin page, users are allowed to apply segmentation and feature extraction methods on their chosen dataset. Additionally, this interface also allows users to choose the percentage of the dataset they want to train. After inserting the appropriate data into the system, the user is able to run the SVM. Here the combination of segmentation, feature extraction methods and lastly SVM is run into the dataset. The SVM part of the system will separate the dataset into training and testing based on the specified percentage. The system will print the output into a text box. The output consists of the number of classes, the accuracy of the combination of methods applied and the confusion matrix. The User interface allows the user to choose an image and predict what is the condition of the breast mammography. The condition could be cancerous or non cancerous. If the breast is cancerous, the system may output if the cancer is malignant or benign depending on the number of classes input into the system.

After the combination of the preprocessing algorithm, segmentation algorithms, feature extraction algorithms, SVM classification model, together with the UI, we got ourselves a working diagnosis system for the prediction.

3.2 Product Delivered

The final product of this project is a functional system that encapsulates all functionality proposed into a UI system. This system allows users to choose and apply different segmentation and feature extraction algorithms. The system is also able to take in new images and apply the best combination of segmentation and feature extraction method onto the image. Then using the resulting SVM model, it classifies the type of tumour. More detailed information of the system UI is mentioned in the methodology section. We also wrote an extensive user guide on our system that is submitted separately from this assessment. Before deploying our system, we also wrote some test cases to ensure the system is working as expected. These are all documented in the software testing assessment.

3.3 How are requirements met

In this section, we will look at what was written in the project proposal and compare it to what requirements have been achieved. Based on the project proposal in semester 1, our team had managed to successfully implement most of what was proposed.

Below is a table of Functional and non-functional requirements taken from the project proposal.

Table 1- Functional and non-functional requirements

Item	Category	Requirement
1	Functional	The segmentation method is used to segment out the tumour from normal breast tissue
2	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
3	Functional	The feature extraction method is used to detect the features of the tumour
4	Functional	The output of confusion matrix and features extracted to see the accuracy and data of tumour
5	Non- Functional	The segmentation method is accurately segmenting out the tumour from breast tissue
6	Non- Functional	The feature extraction method accurately detect the segmented tumour features
7	Non-Functional	Create an intuitive UI for the diagnosing system, so a user manual will be unnecessary.
8	Non-Functional	There is the availability of multiple segmentation and featuring extraction methods.

The breast cancer diagnosing system we had implemented does satisfy item 1. Our system is able to apply Hammoche Algorithm and Region Growing algorithm on the segmentation of tumours. For item 2, our system does apply to preprocessing on images before applying segmentation methods. However, our system does not apply post-processing. The system does satisfy requirement item 3. The system applies feature extraction using the LESH or GLCM method. The LESH extracted 512 features while the GLCM extracted 4 features. As for item 4 in the table, the system does output the confusion matrix of the resulting combination of algorithms used. IN the admin interface, after the system runs the different algorithms, it will output the different classes detected, the accuracy of the applied methods and also the confusion matrix from the SVM. For item 5 & item 6, the overall accuracy of the system differs on the combination of algorithms the user had applied. Admittedly, the accuracy of the algorithm we had applied is not as high as we would have liked. As for item 7 in the table, the UI we had implemented is simple and easy to use. We had tested the usability with a few volunteers and the given feedback was overall not bad. We also implemented input validation in the UI. Lastly, for item number 8, our system does have multiple choices of segmentation and featuring extraction methods. For segmentation methods, users are able to choose between the choice of Region Growing and Hammoche Algorithm. Likewise, for the feature extraction method users are able to choose between LESH and GLCM.

3.4 Justification of decisions made

On the user side of the interface, the combination of segmentation and feature extraction we applied is Hammouche and GLCM. The reason why we went with this combination is because it produces the best results with an overall accuracy of 63.85%. Moreover, we decided not to have a post-processing algorithm as the post-processing algorithm is dependent on the segmentation algorithm we used and the segmentation algorithm we went with did not require any post-processing. Furthermore, in comparison with the algorithm that we proposed, the reason why we chose to implement a different algorithm is due to the fact that we didn't not manage to acquire the code of the algorithm we proposed.

3.5 Discussion of all results

The final product of our project will be displayed with user interfaces.

The admin interface, as we call it, is meant for us, the developers of this project. Its main purpose is to run SVM on a dataset of mammography images that has been segmented and feature extracted, and then use the results to calculate how accurate was the prediction of the condition of the tumor using our algorithms. In other words, is our algorithm working properly?

The user interface is meant for the people we developed this project for. They will input a single mammography image, and the algorithm will segment and extract the feature of the image, and then give the result to the user, whether it is cancerous or non-cancerous.

In terms of the running SVM in the admin side of the project, we ran SVM in terms of classifying it into 2 classes and also for 3 classes. The 2 classes are categorized as cancerous and non-cancerous, whereas the 3 classes are categorized as non-cancerous, benign and malignant.

Below displays the combination of results that we have tested out over the past few weeks.

3.5.1 Region Growing + GLCM

Table 2- Results with Region Growing with GLCM

No.	Confusion Matrix	Class Order	Accuracy
1	45 18 29 5	2	0.5511
2	43 19 26 9		0.5422
3	56 14 19 8		0.4889
4	47 0 15 14 0 4 13 0 4	3	0.5022
5	5 17 0 15 42 0 4 14 0		0.5200
6	45 18 0 14 3 0 14 3 0		0.5156

3.5.2 Hammouche + GLCM

Table 3- Results with Hammouche with GLCM

No.	Confusion Matrix	Class Order	Accuracy
1	64 0 33 0	2	0.6356
2	63 0 34 0		0.6400
3	63 0 34 0		0.6400
4	64 0 0 13 0 0 20 0 0	3	0.6356
5	63 0 0 17 0 0 17 0 0		0.6400
6	63 0 0 22 0 0 12 0 0		0.6400

3.5.3 Region Growing + LESH

Table 4- Results with Region Growing with LESH

No.	Confusion Matrix	Class Order	Accuracy
1	41 21 23 12	2	0.5422
2	5 35 9 48		0.6178
3	48 15 18 16		0.5289
4	47 5 10 11 3 4 11 2 4	3	0.4933
5	3 19 0 3 51 3 5 13 0		0.5867
6	54 6 3 13 2 2 13 2 2		0.5111

3.5.4 Hammouche + LESH

Table 5- Results with Hammouche with LESH

No.	Confusion Matrix	Class Order	Accuracy
1	40 24 21 12	2	0.5822
2	40 23 24 10		0.5733
3	39 24 16 18		0.4889
4	38 13 13 8 2 3 10 3 7	3	0.4756
5	48 8 7 11 2 4 15 0 2		0.5511
6	45 9 9 18 0 4 4 4 4		0.5111

The results above included the dataset after preprocessing using the “Opening And Closing” algorithm and a training percentage of 70%, which has 322 images in the entire dataset. Based on the results, we can observe that Hammouche and GLCM not only gives the best accuracy out of all the other combinations, the results are consistent for both 2 class and 3 class datasets. Thus, the reason the user model is based on Hammouche and GLCM.

3.6 Limitations of project outcomes

There are a few limitations to our project. Firstly, it must be kept in mind that there are many algorithms in the world that have been implemented for image segmentation and feature extraction. As of now, our project has only implemented 2 algorithms each for segmentation and feature extraction each. This reduces the possibility of clearer segmentation images, better features extracted, and more accurate results.

Other than that, after classification of our segmented and feature extracted mammography images using SVM, the accuracy produced is relatively low. The accuracy fluctuates between 40 and 64 percent.

Moreover, another factor that could affect the performance of our model is the dataset we used. As our diagnosis system only works for images without pectoral muscle, we manually removed it ourselves and could have accidentally removed and/or added some noise into the dataset in the process. It is important to note that the user model is trained on the current dataset.

Furthermore, our project does not have an appropriate pre-processing method for our mammography images before we put them through segmentation and feature extraction. Pre-processing a mammography image filters unwanted parts of the image or and/or enhances the image to make it suitable for segmentation and then feature extraction. In our case, our mammography image cannot include the pectoral muscle. Instead of removing it using pre-processing methods, we manually remove every pectoral muscle part in every mammography image we have that has it.

3.7 Discussion of possible improvements and future works

For possible improvements, one way to improve the overall performance of the algorithm is to apply a different classification model, as the classification model is not in our project scope and was not the main focus as well as improving the pre-processing algorithm. Other than that, as an extension to the diagnosis system we created, other segmentation and feature extraction algorithms can be added to the diagnosis system.

3.8 Any other matter of relevance and interest

At the introduction section of the paper, it was briefly explained that our project handles the second phase of creating a new method of diagnosing breast cancer. To reiterate, the 3 stages of this topic is removing the pectoral muscle from breast mammography, applying segmentation and feature extraction and lastly classification and identification of the tumour. There are 3 separate teams implementing each stage respectively. Our team would be interested to incorporate all 3 team work into a single cohesive system. It would be interesting to build a single system that allows users to apply all 3 functionality that was mentioned previously. Subsequently, it will also be interesting to know how other teams implement their project. Perhaps we could learn a thing or two from other teams' implementation.

4.0 Methodology

4.1 Algorithms

For segmentation, we have implemented the Region Growing algorithm and Hammouche's algorithm. Hammouche's algorithm produced a better segmented mammography image than the region growing method. For region growing, we did not implement an optimized version of region growing as proposed before. This is because of the unavailability of the author of the research paper that we obtained the understanding of the optimized version of region growing. Therefore, there is no chance of clarifying or debugging the attempted code we wrote.

For feature extraction, we implemented it using GLCM and LESH. GLCM stands for grey level co-occurrence matrix. We can obtain features and statistics from the GLCM like contrast, correlation, energy and homogeneity. As for LESH, it stands for local energy-based shaped histogram. Each bin in the histogram are the energy levels of the image that it is run on. Energy is a form of measurement in feature perception. LESH is a replacement of the proposed feature extraction method using wavelet theory. In the same predicament, the author was unavailable for the clarification of the code for the feature extraction method using wavelet theory.

We have also implemented SVM for classification of our mammography images. This takes a dataset of mammography images and classifies them, then produces a confusion matrix and the accuracy of the classification. This was an extra feature that we implemented that was not proposed beforehand. We decided that we want to implement SVM because we think it would enhance our overall project greatly. SVM will have the role of ensuring that our algorithms for segmentation and feature extraction actually work. This will be reflected in the level of accuracy produced from classification from the SVM.

4.2 User Interface

The final design of the user interface we had implemented consisted of 3 main components. The first component is the main page where the user is allowed to select which interface they want to go next. Compared to the design proposed in semester 1, we had added this extra main page during the implementation phase. The second interface is the Admin interface where users are allowed to choose a dataset they want to input and apply segmentation and feature extraction methods of their choice. There is also extra functionality added in this interface such as a slider to choose the percentage of users who want to train their dataset, the input of CSV file of actual breast conditions for training and a lamp to indicate when the program is not running, running and completed. This extra functionality added was not in the originally proposed idea. Compared to what was proposed in the project proposal, the current implementation of UI is much more improved. In the project proposal design, the user has to click run twice. The first run is for the segmentation method and the second is the feature extraction method. We improved this design by having a single "run SVM" button that runs both segmentation and feature extraction. As for the User interface, the user is able to select an image they want to apply segmentation and feature extraction and get the results of the diagnosis. The output also contains information on what methods have been applied to the image. Compared to the initial proposal, we had improved the UI by only having to click a single button to run both the segmentation and feature extraction. The overall layout of the design is also more simple and cleaner. However, the main fundamental change that was implemented in the User interface was that users are not allowed to choose the segmentation and feature extraction that they want. The reason for this is because in the admin page we had run all possible combinations of the segmentation and feature extraction method and selected the one with the best accuracy to be used in the User interface. In essence, the admin interface is to test out different combinations of methods while the user page is to apply the best combination. Another essential point, the current implementation also incorporates input validation in the UI. This was not considered in the initial proposal.

4.3 Comparison of Initial and Implemented UI

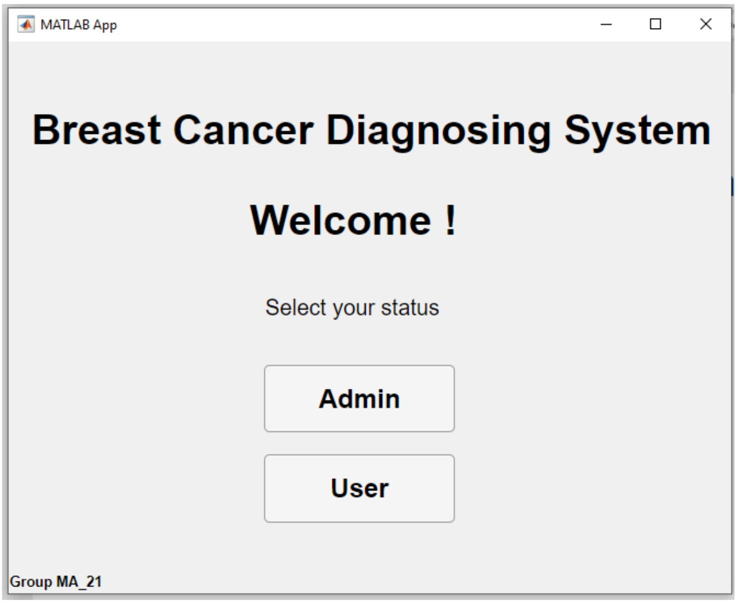
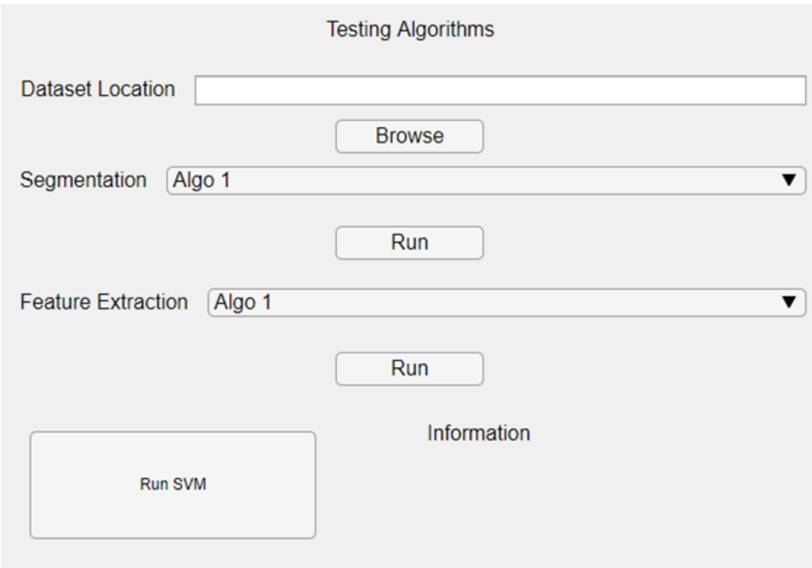


Figure 1: The main page



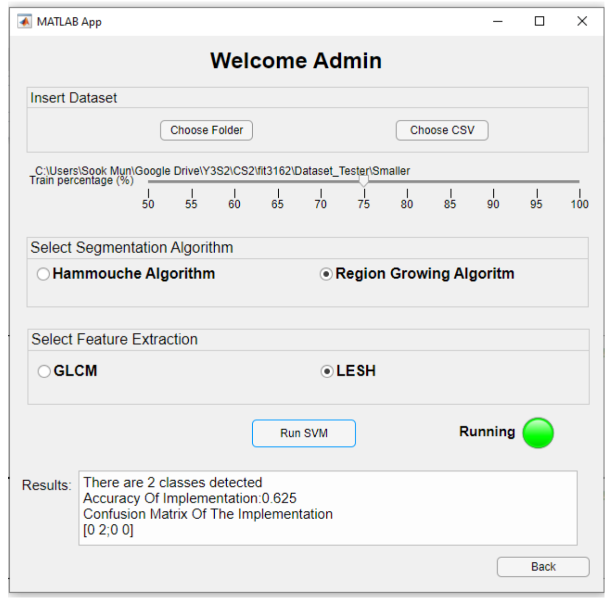
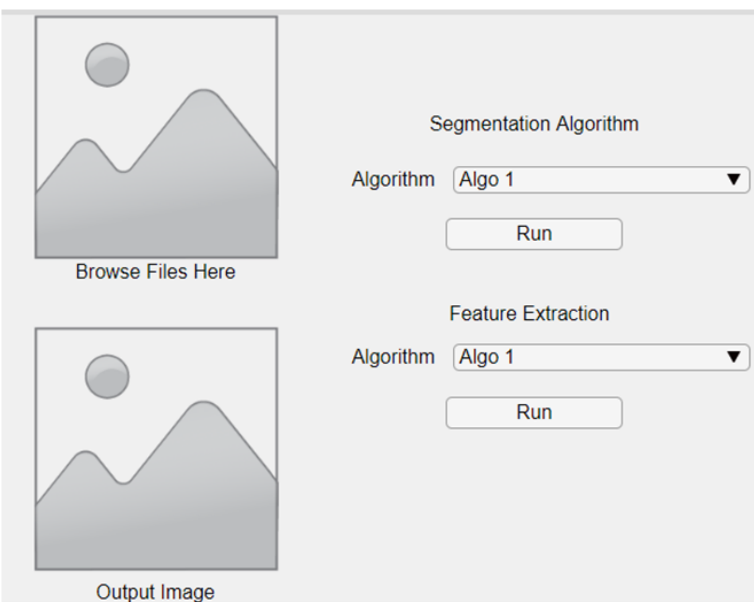


Figure 2: Initial proposed Admin Interface

Figure 3: Implemented Admin Interface



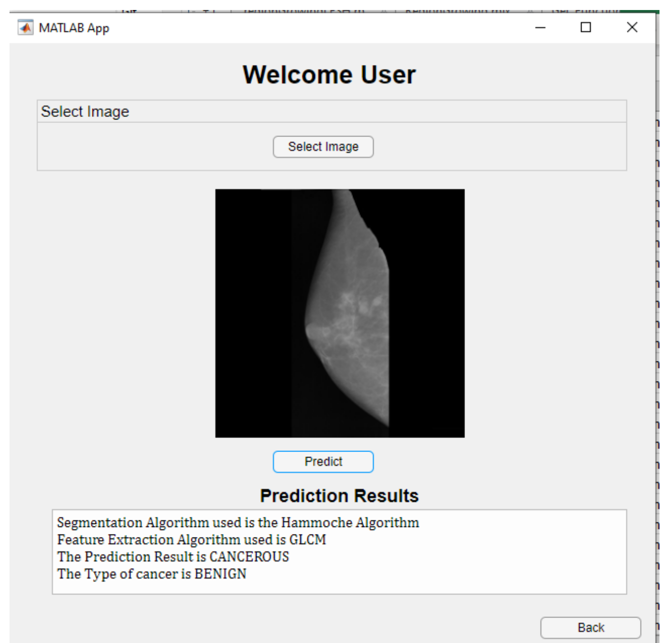


Figure 4: Initial proposed User Interface

Figure 5: Implemented User Interface

4.4 Implementation

The tools used to implement the projects are mainly MATLAB. MATLAB was used to implement and code all segmentation and feature extraction algorithms using the Image Processing Toolbox. As for the UI, we used the MATLAB build-in tool “app design”. To obtain the code base for the algorithms we have to contact multiple researchers that are authors of the research paper we had read. Through this contact, we politely request their codebase. Once we received the code base, we tried running it. More often than not, we had to figure out how to run the codebase. When we are having difficulty, we ask for help from the researchers or from our project supervisor. There was an instance where the codebase that was given to us was in a different language. We had to translate by using google translate to understand the comments and documentation of the codebase. We also implemented the SVM functionality so we are able to predict the resulting segmentation and feature extraction value. Initially, the result we were receiving from the SVM classifier was bad. We tested our SVM code with another dataset which is the Iris dataset that was built in MATLAB. After some trial and error, we managed to get the SVM working. Lastly, we implement the UI of the system. We modularise our functions into different mat files. This way, the code structure will be cleaner and easier to debug. We also made all functions return values instead of writing the output into a CSV file. This is to reduce the file dependency of the system. We also implement input validation in the UI and the functions of our system during the software testing phase. For the UI a warning will pop up when the user enters the wrong data while for functions, an assert error will be raised when the wrong input is entered.

5.0 Software Deliverables

5.1 Summary of software deliverable

5.1.1 Deliverables

The deliverables for our projects is a breast cancer diagnosing system. The source code is uploaded into gitLab. The system is able to take in images and apply segmentation and feature extraction onto it. Then the system outputs the confusion matrix and accuracy of the selected combination of method.The system is also able to take in an image and predict the type of cancer or type of tumour.

Within this system, there are multiple mat files that contain the functionality of the system. The table below detailed each mat file and its corresponding function.

Table 6- Details of functions

File Name	Description
app1	The UI of the Main page
app2	The UI of the Admin page
app3	The UI of the User page
fun_custo	Helper function of Hammouche
hammocheGLCM	Implementing the combination of hammoche and GLCM
hammocheLESH	Implementing the combination of hammoche and LESH
lowpassfilter	Helper function for LESH
phasecong3	Helper function for LESH
pre_process	Function to pre -process the image
regionGrowingFun	Implementation of Region Growing
regionGrowingGLCM	Implementing the combination of Region Growing and GLCM
regionGrowingLESH	Implementing the combination of Region Growing and LESH
segmenta	Implementation of Hammouche algorithm
SVM_GLCM	Implementing SVM when user selects GLCM in the combination
SVM_LESH	Implementing SVM when user selects LESH in the combination

We also created test suites to ensure the correctness of the system. The table below detailed each test suit and its functionality.

Table 7- Details of test files

File Name	Description
test_featex_lesh	This is the test file to test functionality of feature extraction LESH
test_module	This is the test file for testing functionality of combinations of our algorithms
test_segmentation_ham	This is the test file to test functionality of segmentation Hammouche’s algorithm
test_segmentation_rg	This is the test file to test functionality of segmentation region growing
test_svm	This is the test file to test functionality of SVM classification
test_algorithm_perf	This is the test file to test the accuracy of the different combinations of segmentation and feature extraction algorithms.

5.1.2 Proof of Deliverables

Figure 6 : Main Page

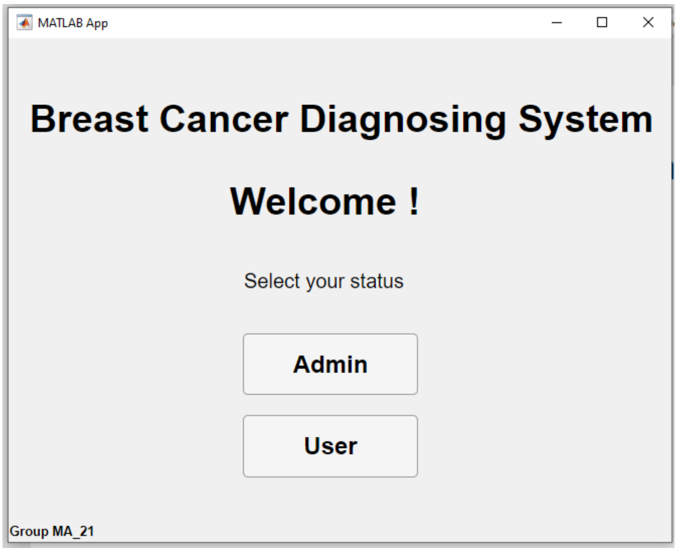
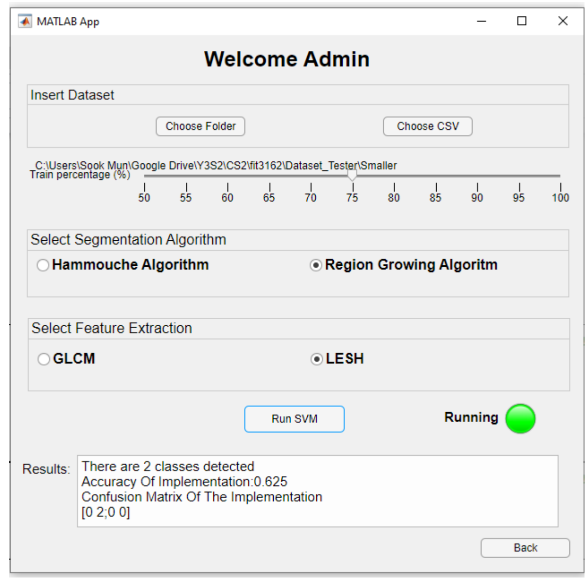


Figure 7: Admin Interface



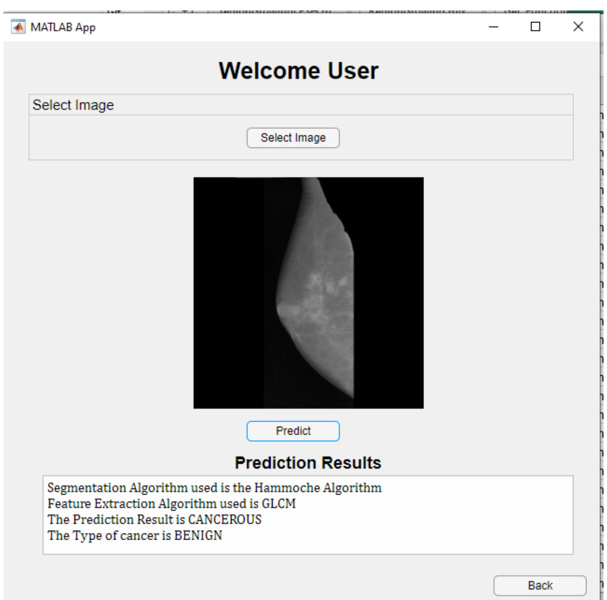


Figure 8 : User Interface

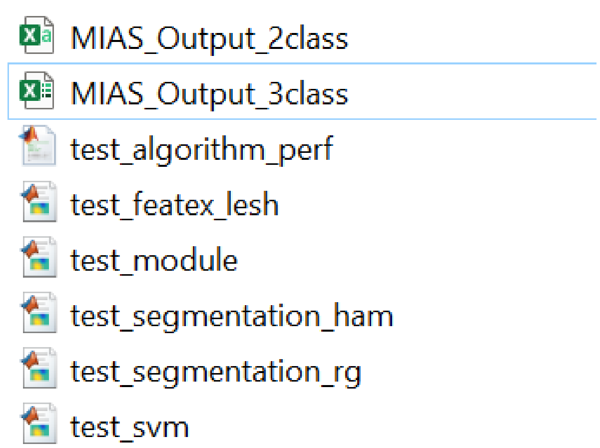


Figure 9 : Unit Testing Files

5.2 Summary and discussion

5.2.1 Robustness

We determine robustness based on the accuracies taken from the repeated running of the SVM classification. The accuracy values for each time the SVM runs the value fluctuates. This shows that our project is not very robust.

5.2.2 Security

Security of implemented code is the endeavor of coding in a way that ensures elimination of vulnerabilities that could be exploited by external malicious individuals or code. In our project, security is not relevant as there is no need for confidential communication between software and users.

5.2.3 Usability

Usability is how easy users are able to use our final product of our project. Our project has good usability. It makes use of input validation for inputs the user gives. The user guide given to users is easy to read and understand. Any errors users make to any inputs to the software will produce a meaningful message to users instead of the error logs dumped into the output. Furthermore, not only does it cater to users, it also caters to the admin, which is us. It makes any errors easier to understand for us to debug.

5.2.4 Scalability

Scalability means the ability of code to adapt to a growing dataset. We have first tested our software using a small dataset of mammography images for our SVM classification. We then gradually grew that dataset and tested our software again. In all cases, the software runs successfully. We can then conclude that our software will be reliable with larger datasets.

5.2.5 Documentation and Maintainability

Documentation for this project is done in several aspects. First of all, a user guide is written stating the instructions and guidelines on how to use our software. Other documentations are on the admin side (which is us) which include meeting minutes, weekly progress reports, project and team management reports and so on.

In terms of maintainability, as this project has a very clear purpose of doing image processing on mammography images, it is not relevant as our objectives are absolute and it has been met.

6.0 Critical Discussion

Considering the outcome of the entire project, it can be concluded that the project was a success. The reason for this statement is because what we had implemented does satisfy most if not all of the requirements that were proposed in the project proposal. The team had started implementing the project before the semester began to help give a bit of a head start. Before semester 2 began, the dataset had already been prep and the git repository created. The team also contacted the author of the research paper that proposed the dragonfly algorithm. Throughout the semester we keep trying to contact the authors of the research paper to ask for more clarification and the source code for the algorithm. This was a difficult task because we did not expect the time it takes for the authors to reach back to us. This may have caused the team to be a bit behind schedule. In some cases when we do receive the source code, we find that we are unable to run the code. For instance, we proposed the Dragonfly Algorithm in the project proposal. However, when receiving the source code for the Dragonfly algorithm, we had trouble running the code. We tried reaching out to the author but still were not able to solve the issue. In the end we had to pivot to a different algorithm. This was one of the issues that caused a setback in our progress. During the planning of the project proposal, we did not expect that contacting the authors of the research paper would take so long.

Subsequently, the other bump in the road we encountered is our dataset. For the entire project, we worked on the MIAS dataset. We needed the dataset of images and also the list of actual conditions of the breast for the SVM functionality of the system. Unfortunately toward the end of the implementation, we notice there was some duplicated data in the list of actual conditions that was given when downloading the dataset.For instance in the given list there was duplicated data for image 5. The duplicated data caused the wrong reading of the actual conditions of any subsequent image. When inspecting the list, it was found that multiple images had duplicates. We had to manually check for duplicates and run our algorithm again. This was very time consuming and we had to redo our analysis on applying different combinations of algorithms. With the new analysis we found that the combination we assume to be the best was not. This caused us to make last minute changes to our code and documentation. In hindsight we should have checked the dataset given and not put blind trust in what we were given.

To summarize, we did not deviate much from the initial project proposal. This contributed to the fact we had put a lot of effort into planning the project proposal. We also consulted a lot with our project supervisor and she had given tons of insight on what to expect during implementation. She also had guided the team a lot in the entire project from giving advice and helping to give more insight into what we are supposed to implement. Any time when we are having an issue she will assist us as much as she can. As mentioned before, the major deviation from the project proposal is the Dragonfly Algorithm. We proposed to implement the optimised region growing which is the Dragonfly algorithm. However, we were not able to do so thus pivoting to implementing the normal region growing. We realise that planning the project is different from Implementing. The deduction from this is from the many roadblocks we had encountered. No matter how much you plan or mitigate certain tasks, during implementation there will be unexpected circumstances that pops up. The only thing we as a team do is to pull full effort on solving issues whenever it arises.

7.0 Conclusion

In conclusion, our project is to build a diagnosing system which supports 2 types of end users, the admin and normal users. At the user side, we allow the user to input a singular image, which predicts the input based on the best combination of segmentation and feature extraction algorithm which happens to be Region Growing and GLCM, to be cancerous or not, and if so, the type of tumour it contains(malignant/benign). In the admin side of the diagnosis system, the diagnosis system will run on an input of a folder of images as well as a csv file containing the name of the images and the breast condition. The segmentation algorithm and feature extraction algorithm of the user's choosing will then be applied to the dataset and a table of the features extracted will be fed on the SVM model alongside the percentage of training dataset implied by the user. The accuracy will then be displayed on the user interface.

As mentioned above, one of the biggest limitations of our project is due to the pre-processing algorithm, dataset used and limited amount of segmentation and feature extraction algorithm used. With more time, resources and research on the following limitations, I believe that the performance of our diagnosis system will improve.

8.0 Appendix

Source Code : https://gitlab.com/stan0111/fit3162/-/tree/main/Project

Below are snippets of our code that we had wrote

Main - Navigation through the applications

```
    methods (Access = private)

        % Button pushed function: AdminButton_2
        function AdminButton_2Pushed(app, event)
            app2
            app.UIFigure.Visible = 'off';
        end

        % Button pushed function: UserButton
        function UserButtonPushed(app, event)
            app3
            app.UIFigure.Visible = 'off';
        end
    end
```

User - When the predict button is clicked

```
function PredictButtonPushed(app, event)
    % segment it
    % extract feature extraction
    % feed into SVM
    if class(app.imageChosen) ~= "uint8"
        message = {'Image Not Selected!','Please select an image'};
        uialert(app.UIFigure,message,'Incorrect Input','Icon','error');
        return
    end
    I=pre_process(app.imageChosen); %pre-process the image
    %apply hammouche algorithm
    IM = segmenta(I);
    %applying GLCM
    [GLCM, ~] = graycomatrix(IM,'Offset',[1 1],'NumLevels',2,'GrayLimits',[],'Symmetric', true);
    stats = graycoprops(GLCM,'all');
    cell = struct2cell(stats);
    flipCell = cell.' ;
    Table = cell2table(flipCell);
    Table.Properties.VariableNames = {'Contrast' 'Correlation' 'Energy' 'Homogeneity'};
    %the creation of traning model
    %load the saved traning model
    load UserSVMModel GLCMModel
    %predict the image chosen by user
    TestOutputs = predict(GLCMModel,Table);
    result=string(TestOutputs);
```

```
    %ouput into the text Area
    app.ResultsTextArea.Visible = 'on';
    app.ResultsTextArea.Value{1}=char("Segmentation Algorithm used is the Hammoche Algorithm");
    app.ResultsTextArea.Value{2}=char("Feature Extraction Algorithm used is GLCM");
    if result=="N"
        app.ResultsTextArea.Value{3}=char(("The Prediction Result is NON-CANCEROUS"));
    elseif result =="B"
        app.ResultsTextArea.Value{3}=char(("The Prediction Result is CANCEROUS "));
        app.ResultsTextArea.Value{4}=char(("The Type of cancer is BENIGN "));
    elseif result == "M"
        app.ResultsTextArea.Value{3}=char(("The Prediction Result is CANCEROUS "));
        app.ResultsTextArea.Value{4}=char(("The Type of cancer is MALIGNANT"));
    elseif result == "C"
        app.ResultsTextArea.Value{3}=char(("The Prediction Result is CANCEROUS"));
    end
```

Admin - When user clicked on the SVM button

```
function RunSVMButtonPushed(app, event)
    %getting user input
    if ~isa(app.training_percentage,'double')
        message = {'Training Percentage Not Selected!','Please ensure you had input all appropriate data','Choose
        uialert(app.UIFigure,message,'Incorrect Input','Icon','error');
        return
    end
    if app.folder==""
        message = {'Dataset Folder Not Selected!','Please ensure you had input all appropriate data','Input a fol
        uialert(app.UIFigure,message,'Incorrect Input','Icon','error');
        return
    end
    if ~isequal(class(app.csvTable),"table")
        message = {'CSV File Not Selected!','Please ensure you had input all appropriate data','Input a csv file'
        uialert(app.UIFigure,message,'Incorrect Input','Icon','error');
        return
    end
    seg=app.SelectSegmentationAlgorithmButtonGroup.SelectedObject.Text;
    fea=app.SelectFeatureExtractionButtonGroup.SelectedObject.Text;
    app.RunningLamp.Color='yellow'; %changing the state of the lamp
    drawnow  %updating the UI
    % Hammouche
    if seg=="Hammouche Algorithm"
        if fea=="GLCM"
            %Hammouche Algorithm +GLCM
            outputTable = hammoucheGLCM(app.folder,app.csvTable);
            [con_matric,noClasses,accuracy]=SVM_GLCM(outputTable,app.training_percentage);
        else
            %Hammouche Algorithm +LESH
            outputTable=hammoucheLESH(app.folder,app.csvTable);
            [con_matric,noClasses,accuracy]=SVM_LESH(outputTable,app.training_percentage);
        end

    % Region Growing
    else
        if fea=="LESH"
            %Region Growing +LESH
            outputTable = regionGrowingLESH(app.folder,app.csvTable);
            [con_matric,noClasses,accuracy]=SVM_LESH(outputTable,app.training_percentage);

        else
            %Region Growing +GLCM
            outputTable= regionGrowingGLCM(app.folder,app.csvTable);
            [con_matric,noClasses,accuracy]=SVM_GLCM(outputTable,app.training_percentage);
        end
    end
    app.RunningLamp.Color='green'; % change lamp to green when segmentation, feature extraction and SVM has been
    drawnow %updating the UI
    %makign the results text Area Visible
    app.ResultsTextArea.Visible = 'on';
    app.ResultsTextAreaLabel.Visible= 'on';

    %displaying on the text area
    app.ResultsTextArea.Value{1}=['There are ' num2str(noClasses) ' classes detected'];
    app.ResultsTextArea.Value{2} =['Accuracy Of Implementation:' num2str(accuracy)];
    app.ResultsTextArea.Value{3} =('Confusion Matrix Of The Implementation');
    app.ResultsTextArea.Value{4} =mat2str(con_matric);
```

9.0 Annex

Section	Team Member
Introduction	Tan Sook Mun
Literature Review	Jaclyn Neoh Daniel Kee
Outcomes	
What has been implemented	Jaclyn Neoh
Product Delivered	Tan Sook Mun
How are requirements met	Tan Sook Mun
Justification of decisions made	Jaclyn Neoh
Discussion of all results	Daniel Kee
Limitations of project outcomes	Daniel Kee
Discussion of possible improvements and future works	Jaclyn Neoh
Any other matter of relevance and interest	Tan Sook Mun
Methodology	Tan Sook Mun Daniel Kee
Software Deliverables	
Summary of software deliverable	Tan Sook Mun
Summary and discussion	Daniel Kee
Critical Discussion	Tan Sook Mun
Conclusion	Jaclyn Neoh

10.0 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 October 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.