# AUTOMATIC DETECTION OF CALCIFICATIONS IN BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING CLASSIFIERS

**FATINA HAM BINTI YAHYA HAM**

**SCHOOL OF HEALTH SCIENCES**

**UNIVERSITI SAINS MALAYSIA**

**2024**

# AUTOMATIC DETECTION OF CALCIFICATIONS IN BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING CLASSIFIERS

by

**FATINA HAM BINTI YAHYA HAM**

Dissertation submitted in partial fulfilment of the requirement for the degree of
Bachelor of Health Sciences (Honours) (Medical Radiation)

JUNE 2024

# CERTIFICATE

This is to certify that the dissertation entitled "**AUTOMATIC DETECTION OF CALCIFICATIONS IN BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING CLASSIFIERS**" is the bona fide record of research work done by **FATINA HAM BINTI YAHYA HAM** during the period from October 2023 to June 2024 under my supervision. I have read this dissertation and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation to be submitted in partial fulfilment for the degree of Bachelor of Health Science (Honours) (Medical Radiation).

| Main Supervisor: | Co-Supervisor: | Co-Supervisor: |
|---|---|---|
| ……………………………… | ……………………………… | ……………………………… |
| Madam Siti Aishah Abd Aziz | Dr Rosniwati Ghafar | Dr Lau Chiew Chea |
| Senior Lecturer | Lecturer | Lecturer |
| School of Health Sciences | School of Health Sciences | School of Medical Sciences |
| Universiti Sains Malaysia | Universiti Sains Malaysia | Universiti Sains Malaysia |
| Health Campus | Health Campus | Health Campus |
| 16150 Kubang Kerian | 16150 Kubang Kerian | 16150 Kubang Kerian |
| Kelantan, Malaysia | Kelantan, Malaysia | Kelantan, Malaysia |

Date: June 2024

# DECLARARTION

I, Fatina Ham Binti Yahya Ham hereby declare that the dissertation entitled "**AUTOMATIC DETECTION OF CALCIFICATIONS IN BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING CLASSIFIERS**" is the result of my own investigations, except where otherwise stated and duly acknowledged. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Universiti Sains Malaysia or other institutions. I grant Universiti Sains Malaysia the right to use the dissertation for teaching, research and promotional purposes.

………………………………

FATINA HAM BINTI YAHYA HAM

Date: June 2024

# ACKNOWLEDGEMENT

In the name of Allah S.W.T., the Most Gracious and the Most Merciful. All praise to Him for His blessings and for giving me the opportunity to complete the research project for my bachelor's degree. This journey has been long and challenging, but it is an enjoyable phase of life that I will always remember. This report would not have been possible without the guidance and help of several individuals who contributed in various ways to the preparation and completion of my final year project. Special thanks to my parents, Yahya Ham Abdullah and Roslina Ismail, for their endless support in morals, beliefs, financial aid, and necessities needed for this thesis. I also extend my heartfelt thanks to my younger brothers, Fairil Ham Yahya Ham and Faikhil Ham Yahya Ham, for their significant help and support in the research process. My deepest gratitude to my knowledgeable supervisor, Madam. Siti Aishah Abd Aziz, for her supervision over the two semesters of my study. Her constructive suggestions, advice, passion, and sincerity in helping her students excel are incomparable. I also greatly appreciate my co-supervisors, Dr. Rosniwati Ghafar and Dr. Lau Chiew Chea, for their invaluable help, especially during data collection, which would have been impossible without their support. I am deeply thankful to my fiancée, Muhammad Haiqal Ramli and close friends, Farahamizah Othman and Nurul Shazliana Mohamad Daud, for their unwavering love, emotional, financial, and mental support throughout this journey. Their assistance was invaluable, especially during stressful times and cherished moments. I could not have completed this study without them. Lastly, I cannot forget the contributions of everyone mentioned here. I am forever grateful for their presence and support throughout these years. I wish them all happiness and success in their future endeavors.

## TABLE OF CONTENTS

CERTIFICATE ........ ii

DECLARARTION ........ iii

ACKNOWLEDGEMENT ........ iv

LIST OF FIGURES ........ ix

LIST OF TABLES ........ x

LIST OF EQUATION ........ xi

LIST OF SYMBOLS ........ xii

LIST OF ABBREVIATIONS ........ xiii

ABSTRAK ........ xvi

ABSTRACT ........ xvii

CHAPTER 1 ........ 1

1.1 Background of the Study ........ 1

1.2 Problem Statement ........ 3

1.3 Objective ........ 3

1.3.1 General Objective ........ 3

1.3.2 Specific Objectives ........ 3

1.4 Hypothesis ........ 4

1.4.1 Null Hypothesis ........ 4

1.4.2 Alternative Hypothesis ........ 4

1.5 Significant of Study ........ 4

1.6. Conceptual Framework ........ 5

CHAPTER 2 ........ 6

2.1 Mammography Imaging and Calcification Detection ........ 6

2.2 Preprocessing Techniques ........ 7

2.3 Feature Extraction Methods: Thresholding ........ 9

2.4 Machine Learning Algorithms ........ 12

2.5 ROC curve and AUC analysis ........ 15

2.6 Performance Evaluation Metrics ........ 17

2.7 Ensemble Learning ........ 20

CHAPTER 3 ........ 23

3.1 Study Design ........ 23

3.2 Study Location ........ 23

3.3 Selection Criteria ........ 23

3.3.1 Inclusion Criteria ........ 23

3.3.2 Exclusion Criteria ........ 23

3.4 Sample Size Estimation ........ 24

3.5 Data Collection ........ 24

3.6 Study Instruments ........ 25

3.6.1 Pictures Archiving and Communication System (PACS) (Version 6.0, GE Healthcare) .......... 25

3.6.2 Siemens Mammomat Revelation Mammography Machine .......... 26

3.6.3 MATLAB Software .......... 28

3.6.4 Laboratory Information System (LIS) .......... 29

3.7 Method .......... 31

3.8 Study Flowchart .......... 35

3.9 Image Analysis .......... 36

3.9.1 Image Pre-Processing .......... 36

3.9.2 Feature Extraction .......... 37

3.9.3 Classification .......... 39

3.9.4 Performance Evaluation .......... 41

CHAPTER 4 .......... 44

4.1 Results .......... 44

4.1.1 Mammogram Samples .......... 44

4.1.2 Output of Pre-processed Image .......... 52

4.1.3 Feature Extraction .......... 55

4.1.4 Data Splitting .......... 57

4.1.5 Classifications .......... 64

4.1.6 Data Training and Testing .......... 71

4.2 Discussions ........................................................................................................ 72

4.2.1 Mammogram Samples ............................................................................. 72

4.2.2 Image Pre-processing ............................................................................. 73

4.2.3 Data Testing ............................................................................................ 80

4.2.4 Performance Evaluation ......................................................................... 81

4.3 Limitation of Study ........................................................................................................ 84

4.4 Future Recommendations ........................................................................................... 85

CHAPTER 5 ......................................................................................................................... 86

CONCLUSION ...................................................................................................................... 86

REFERENCES ....................................................................................................................... 87

APPENDICES ........................................................................................................................ 96

APPENDIX A: MATLAB Code ........................................................................................... 96

# LIST OF FIGURES

Figure 1.1 Conceptual Framework ………………………………………………………………….. 5

Figure 2.1 Median Filter Calculation …………………………………………………………,, 9

Figure 2.2 ROC Curve (Nahm, 2021) ………………………………………………………... 17

Figure 3.1 Method ……….…………………………………………………………………... 34

Figure 3.2 Study Flowchart ………..……………………………...……………………….. 35

Figure 4.1 Original Data Training ………………………………………………………… 65

Figure 4.2 Hyperplane Training ……………………………………………………………... 66

Figure 4.3 Confusion matrix for training …………………………………………………. 66

Figure 4.4 AUC Training ……………………………………………………………………. 68

Figure 4.5 Confusion matrix testing ……………………………………………………… 69

Figure 4.6 AUC Testing ……………………………………………………………….…. 70

# LIST OF TABLES

Table 4.1 List of mammogram samples used in this study …………………………………………...44

Table 4.2 Sample of mammogram images before and after result ……………………………...52

Table 4.3 Comparison between original image and highlighted calcifications image ………… 57

Table 4.4 List of mammogram images used for training and validation set …………………….58

Table 4.5 Model accuracy for all SVM types ………………………………..……………… 64

Table 4.6 Results of data training ……………………...………………………………………. 71

Table 4.7 Comparison between before and after using the filters ………………………………79

Table 4.8 : Performance Evaluation for all classifiers ………………………………………… 83

# LIST OF EQUATION


Equation 1: Accuracy ........ 18

Equation 2: F1 Score ........ 18

Equation 3: Recall ........ 19

Equation 4: Precision ........ 19

Equation 5: Specificity ........ 19

Equation 6: Sensitivity ........ 19

Equation 7: Sample Size Formula ........ 24

Equation 8: Sample Size Calculation ........ 24

Equation 9 : True Positive rate (Sensitivity) ........ 67

Equation 10 : True Negative rate (Specificity) ........ 67

Equation 11 : False Positive rate ........ 67

Equation 12 : False Negative rate ........ 67

Equation 13 : Gray conversion factor ........ 75

Equation 14 : Median Filter ........ 76

Equation 15 : White-top hat transforms ........ 77

Equation 16 : Top-hat filter syntax ........ 77

Equation 17 : Weiner filter syntax ........ 78

## LIST OF SYMBOLS

| | |
|---|---|
| *k* value | Number of mirrors neighbor |
| n | Sample size |
| z | Value representing the desired confidence level |
| Δ | Precision or true value |
| p | Anticipated dataset proportion |
| Z-score | Confidence level corresponds |
| C/C++ | Programming language |
| *R* | Red |
| *B* | Blue |
| *G* | Green |
| $I \circ B$ | Denotes the morphological opening |
| *T* | Top-hat transform |
| *I* | Image |

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ACC | Accuracy |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Networks |
| AUC | Area Under Curve |
| ASIC | Application-Specific Integrated Circuits |
| BI-RADS | Breast Imaging Reporting and Data System |
| BAC | Breast Artery Calcifications |
| CAD | Computer-Aided Diagnosis |
| CC | Cranio-Caudal |
| CLAHE | Contrast Limited Adaptive Histogram Equalization |
| CNN | Convolutional Neural Network |
| CUDA | Compute Unified Device Architecture |
| CT | Computed Tomography |
| DICOM | Digital Imaging and Communications in Medicine |
| DCIS | Ductal Carcinoma in Situ |
| DPI | Dots per Inch |
| DTCWT | Dual-Tree Complex Wavelet Transform |
| EHR | Electronic Health Record |
| GPU | Graphics Processing Unit |
| FN | False Negatives |
| FP | False Positives |
| FPGA | Field-Programmable Gate Arrays |

| | |
|---|---|
| FPR | False Positive Rate |
| HDL | Hardware Description Language |
| HOG | Histogram of Oriented Gradients |
| HUSM | Hospital Universiti Sains Malaysia |
| IT | Information Technology |
| K-NN | K-Nearest Neighbor |
| LIS | Laboratory Information System |
| LBP | Local Binary Patterns |
| ML | Machine Learning |
| MySCan | Malaysian Study on Cancer Survival |
| MLO | Mediolateral Oblique |
| MRI | Magnetic Resonance Imaging |
| NCR | National Cancer Registry |
| NLP | Natural Language Processing |
| OpComp | Optimal Compression |
| PACS | Picture Archiving and Communication System |
| RF | Random Forest |
| ROC | Receiver Operating Characteristic |
| ROI | Region Of Interest |
| SVM | Support Vector Machine |
| TN | True Negatives |
| TP | True Positives |
| TPR | True Positive Rate |

| | |
|---|---|
| WISH | Women Imaging Suite |
| 2D | 2-Dimensional |
| 3D | 3-Dimensional |

**PENGESANAN AUTOMATIK KALSIFIKASI DALAM DIAGNOSIS KANSER PAYUDARA BERASASKAN PENGKLASIFIKASI PEMBELAJARAN MESIN**

# ABSTRAK

Pengesanan awal kanser payudara melalui mammogram adalah penting, dengan kalsifikasi dalam mammogram sebagai penunjuk utama. Membezakan antara kalsifikasi malignan dan benigna adalah penting untuk diagnosis dan rawatan yang tepat. Kajian ini bertujuan untuk membangunkan sistem Diagnosis Berbantuan Komputer untuk mengenal pasti dan mengklasifikasikan kalsifikasi payudara. Data kes kanser payudara di *Women Imaging Suite (WISH),* Hospital Universiti Sains Malaysia (HUSM) dikumpulkan dari *Laboratory Information System (LIS)* dan ditentusahkan dengan Sistem Pengarkiban dan Komunikasi Gambar untuk memilih mammogram yang menunjukkan kalsifikasi dari September 2020 hingga Disember 2023. Prestasi model *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)* dan *Random Forest (RF)* dinilai menggunakan metrik seperti ketepatan, skor F1, *recall*, *precision*, kekhususan, sensitiviti dan *area under the curve (AUC)*. Model SVM menunjukkan prestasi yang seimbang dengan ketepatan 65.22% dan skor F1 0.6, yang menunjukkan kompromi antara *precision* (54.55%) dan *recall* (66.67%). Model KNN mempunyai prestasi terendah dengan ketepatan 47.83% dan skor F1 0.4, yang menyerlahkan cabaran klasifikasi. Model RF dinilai secara sederhana dengan ketepatan 60.87% dan skor F1 0.47, menunjukkan kekhususan yang tinggi (71.43%) tetapi sensitiviti yang lebih rendah (44.44%). Mencapai ketepatan 95% masih sukar kerana pengesanan bergantung kepada nilai piksel yang tinggi, kerumitan model terhad, dan kekangan data. Meningkatkan pengekstrakan ciri, data dan mengoptimumkan model boleh meningkatkan ketepatan. Menggabungkan pembelajaran mesin dengan pembelajaran mendalam atau kaedah *ensemble* menawarkan klasifikasi yang lebih baik dan pengurusan pesakit yang lebih baik.

**AUTOMATIC DETECTION OF CALCIFICATIONS IN BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING CLASSIFIERS**

## ABSTRACT

Early detection of breast cancer through mammography is vital, with calcifications in mammograms serving as key indicators. Distinguishing between benign and malignant calcifications is essential for accurate diagnosis and treatment. This study aims to develop a Computer-Aided Detection (CAD) system to identify and classify breast calcifications. Data from confirmed breast cancer cases were collected from the Laboratory Information System (LIS) at the Women Imaging Suite (WISH) of Hospital Universiti Sains Malaysia (HUSM) and cross-verified with the Picture Archiving and Communication System (PACS) to select mammograms showing calcifications that met the inclusion criteria from September 2020 to December 2023. The performance of Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) models was evaluated using metrics such as accuracy, F1 score, recall, precision, specificity, sensitivity, and area under the curve (AUC). The SVM model showed balanced performance with 65.22% accuracy and an F1 score of 0.6, indicating a trade-off between precision (54.55%) and recall (66.67%). The KNN model had the lowest performance with 47.83% accuracy and an F1 score of 0.4, highlighting classification challenges. The RF model performed moderately with 60.87% accuracy and an F1 score of 0.47, showing high specificity (71.43%) but lower sensitivity (44.44%). Achieving 95% accuracy remains difficult due to reliance on high pixel value detection, limited complexity of machine learning models, and data constraints. Enhancing feature extraction, data augmentation, and model optimization could improve accuracy. Combining machine learning with deep learning or using ensemble methods offers promise for better classification, ultimately improving patient management.

# CHAPTER 1

# INTRODUCTION

## 1.1 Background of the Study

Breast cancer stands as the most prevalent malignancy affecting women worldwide. With its profound impact on public health, efforts in early detection and intervention have become paramount. The Malaysian Study on Cancer Survival (MySCan) obtained breast cancer data from the National Cancer Registry (NCR) database, comprising a total of 18,206 entries. After rigorous processes to identify and remove duplicates, ensure data consistency, and verify morphology, 17,490 cases were considered eligible for analysis. However, the final study cohort consisted of only 17,009 cases diagnosed with breast cancer between 2007 and 2011, with follow-up data available until December 31, 2016 (National Cancer Registry, 2018).

Mammography, a specialized radiological technique aimed at diagnosing breast-related diseases, particularly breast cancer, is recommended for both symptomatic young women and those at high risk due to family history, as well as for asymptomatic women over 40 years old. This non-invasive screening tool serves as the gold standard in early breast cancer detection, utilizing low-dose x-rays, mammography produces images of the breast capable of identifying various lesions indicative of breast cancer including calcifications, masses, architectural distortions, and asymmetric densities (Coleman, 2017). Despite its effectiveness, interpreting mammograms can be challenging, particularly in cases of dense breast tissue, which may obscure abnormalities and lead to false negatives or missed diagnoses. Qualified radiologists may find interpretation difficult, necessitating careful analysis from different views, such as the Cranio-caudal (CC) and Mediolateral Oblique

(MLO) views, to distinguish true abnormalities. However, additional imaging modalities may be required to investigate suspected regions not adequately visualized on mammograms or obscured by dense tissue. Despite these challenges, mammography remains the most reliable screening tool due to its low cost, minimal radiation exposure, and high sensitivity in detecting breast cancer (Champaign & Cederbom, 2000).

Breast calcifications serve as indicators of potential carcinoma, often being the sole detectable sign of nonpalpable breast cancer during screening. Calcifications present as small, bright calcium deposits within the glandular tissue, typically measuring between 0.01 and 0.1 mm in size. They appear as localized areas of high intensity on radiographic films, with clinical studies suggesting the presence of five or more calcification points per centimeter. These calcifications are typically intramammary, situated within and around ducts, lobules, and vascular structures. Various characteristics such as size, morphology, number, distribution pattern, location, and density contribute to the determination of pathology (Nalawade, 2009).

Machine learning-based computer-aided diagnosis (CAD) systems have been created to automatically identify calcifications during breast cancer screening. These systems aim to aid radiologists in early breast cancer detection by spotting suspicious calcifications on mammograms (Jiménez-Gaona et al., 2020). CAD systems have demonstrated potential in enhancing the specificity of mammography interpretation and decreasing the need for biopsies in cases of benign lesions (Gao et al., 2023). Nevertheless, obstacles persist in creating CAD models with high accuracy, such as insufficient training data and less-than-optimal performance in dense breast tissue (Gao et al., 2019). Current research aims to enhance CAD by employing machine learning algorithms, which can autonomously extract pertinent features from mammographic images without necessitating manual feature engineering. Incorporating these advanced AI-driven CAD systems into clinical settings holds promise for augmenting radiologists' capacity to identify and analyze breast calcifications, potentially resulting in earlier cancer detection and better patient prognoses (Loizidou et al., 2023).

## 1.2 Problem Statement

The complex and diverse malignant characteristics found in calcifications is highly challenging due to the presence of noise. However, due to the presence of inadequate noise in medical images, there are numerous limitations related to the process of interpreting a mammogram image (Rehman et al., 2017). Common difficulty encountered in diagnosing breast cancer is the frequent occurrence of incorrect diagnoses. In order to prognosticate calcifications clusters, a highly discriminative classifier is developed after pertinent manually constructed features are extracted. Nevertheless, this method frequently produces unreliable outcomes (Leong et al., 2022). A challenge in differentiating between malignant and benign tumours, particularly when there are associated calcifications. While numerous methods have been suggested for identifying calcifications points, each tends to exhibit a high occurrence of false positives (Mahmood et al., 2021).

## 1.3 Objective

### 1.3.1 General Objective

To develop a Computer Aided Diagnosis (CAD) system to distinguish between benign and malignant breast calcifications.

### 1.3.2 Specific Objectives

1. To apply image processing to specify details of calcifications.

2. To evaluate the image quality from machine learning based algorithms using MATLAB software.
3. To verify the effectiveness of machine learning classifiers for auto-detection of calcifications.

## 1.4 Hypothesis

### 1.4.1 Null Hypothesis

The accuracy of differentiation between benign and malignant breast calcification by using machine learning classifiers produces results that do not match the BI-RADS subjective assessments.

### 1.4.2 Alternative Hypothesis

The accuracy of differentiation between benign and malignant breast calcification by using machine learning classifiers produces results that are more accurate than BI-RADS subjective assessments.

## 1.5 Significant of Study

This research focused on using machine learning classifiers to automatically detect calcifications in breast cancer holds considerable significance in breast cancer diagnosis. By aiming to improve diagnostic accuracy, minimize unnecessary biopsies, and utilize advanced machine learning techniques and ensemble models, these studies have the potential to transform breast cancer screening and diagnosis. Creating computer-aided diagnosis systems capable of precisely identifying and categorizing calcifications as benign or malignant not only enhances patient care by enabling early cancer detection but also helps reduce patient anxiety and increase survival rates (Mahmood et al., 2021). Moreover, integrating these systems into clinical practice can assist radiologists in making more informed decisions, ultimately enhancing diagnostic capabilities, and improving patient outcomes through early cancer detection (Loizidou et al., 2023).

## 1.6. Conceptual Framework

The conceptual framework of this study, illustrated in Figure 1.1, is the automatic detection of calcifications in breast cancer using machine learning. The images obtained for both training and testing sets will go through image pre-processing, detection, and feature extraction to be divided into pixel values. The pixel value will then be classified into classifiers to distinguish between benign and malignant breast calcifications.

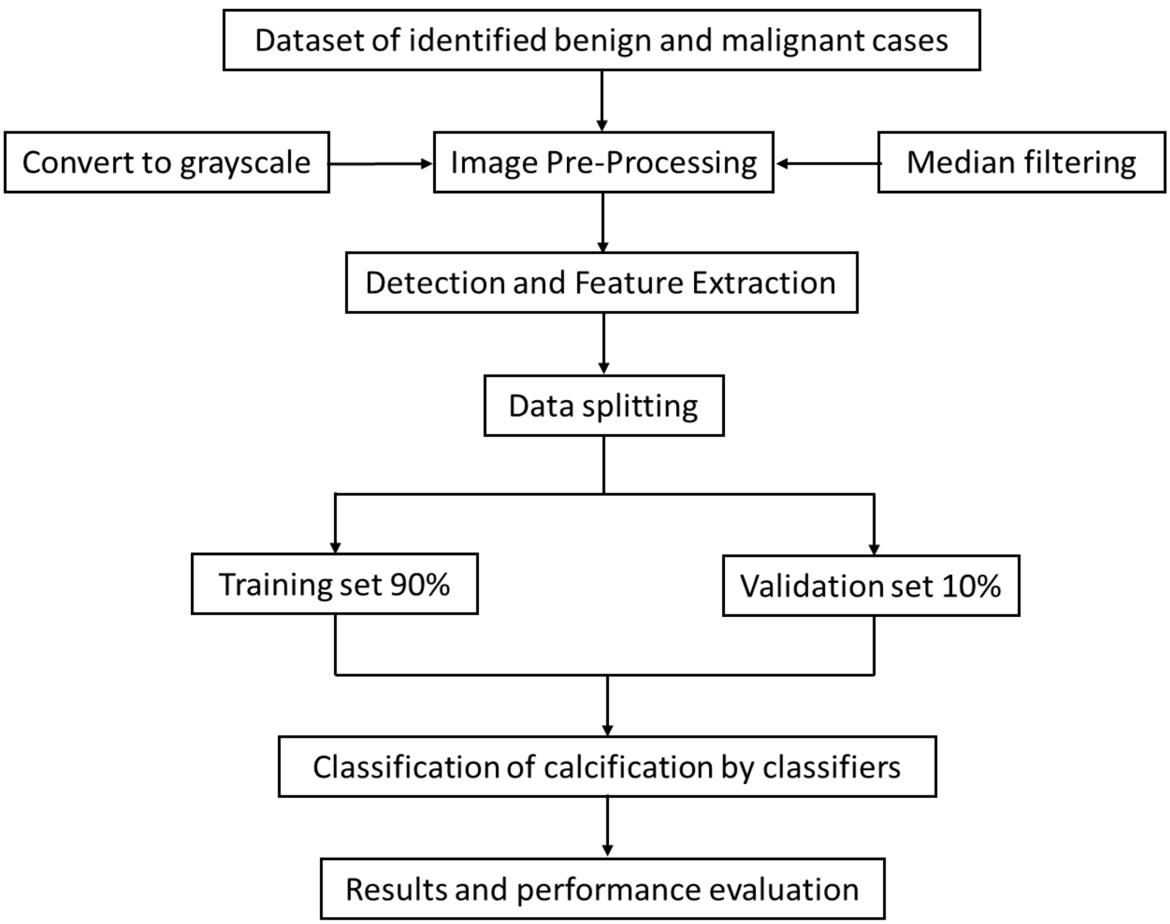


Figure 1.1 Conceptual Framework

# CHAPTER 2

## LITERATURE REVIEW

### 2.1 Mammography Imaging and Calcification Detection

Mammography remains the foremost method for detecting and characterizing breast calcifications. Calcifications serve as a crucial signal for potential breast malignancy. Using the Breast Imaging Reporting and Data System (BI-RADS) lexicon, calcifications can be categorized as either benign or suspicious, with calcifications being classified as "suspicious” depending on their morphology and distribution. Calcifications, particularly those displaying suspicious morphology and distribution, may indicate early non-palpable breast cancers such as Ductal Carcinoma in Situ (DCIS) (Kameswari et al., 2021).

One of the key challenges associated with manual interpretation is the subjective aspect of evaluating mammographic calcifications. The assessment of calcifications can greatly depend on the experience, training, and cognitive biases of individual radiologists. Despite the aim of the BI-RADS lexicon to standardize mammographic findings reporting, there remains only moderate agreement among radiologists in using its terms to describe masses and calcifications.

Additionally, there is considerable diversity among radiologists in interpreting mammographic results, including calcifications. Research has shown moderate concordance among radiologists in utilizing BI-RADS terminology to characterize masses and calcifications, underscoring the difficulties in maintaining uniform interpretation. Radiologists' levels of experience, their training, and their familiarity with reporting frameworks such as BI-RADS all play roles in this interpretative variability (Kerlikowske et al., 1998).

The quality of mammographic images is also essential for radiologists to accurately interpret calcifications. Reduced spatial resolution and increased quantum noise can notably impede radiologists' perceived ability to interpret calcification cases, making it challenging to differentiate between benign and malignant calcifications. Additionally, effects related to post-acquisition image processing can influence radiologists' interpretations, with certain issues affecting calcification cases compared to soft tissue cases (Boita et al., 2021).

Researchers have developed deep learning algorithms to identify calcifications in mammograms automatically. A study showcased a convolutional neural network that attained impressive accuracy rates (95%), F1-scores (76%), and AUC-ROC (76%) in detecting breast calcifications, linked to increased cardiovascular risks. This algorithm precisely pinpointed calcifications, including tiny ones, through gradient-weighted class activation mapping (Grad-CAM++) visualizations (Sakaida et al., 2023).

## 2.2 Preprocessing Techniques

Preprocessing techniques play a crucial role in detecting calcifications related to breast cancer, which helps with early diagnosis and treatment planning. Several research works have focused on optimizing algorithms and methods to improve the precision and effectiveness of calcification detection in mammography images.

A fundamental step in these pre-processing techniques involves converting input images into grayscale. This conversion simplifies the data and prepares it for further analysis. By transforming the input image data into grayscale, the complexity of the image is reduced, facilitating segmentation, feature extraction, and classification tasks. Grayscale conversion eliminates colour information, focusing solely on intensity variations within the image, which is crucial for identifying subtle features such as calcifications in breast tissue (Hirra et al., 2021). Various techniques can be employed for grayscale conversion, with Otsu's method being a popular choice for automatically selecting the optimal grayscale threshold for segmenting calcifications in breast cancer images. This method aids in generating binary images that highlight regions of interest, and also assisting in subsequent analysis and detection process (Guzmán-Cabrera et al., 2012).

Beyond grayscale conversion, it's crucial to address noise in mammography images, as it can obscure important details and hinder accurate analysis. This noise often originates from the imaging equipment itself and can manifest as random fluctuations in pixel values. One effective method to combat this noise while preserving the sharpness of edges and important features.is median filtering. Median filtering emerges as a powerful solution in this regard. Unlike other noise reduction methods that might blur or distort edges, median filtering maintains the integrity of these edges while effectively reducing noise.

For each pixel in the image, the median filter evaluates the values of its neighboring pixels within a defined window or kernel. Within this window, arrange the values in ascending order, instead of averaging these values, it selects the median value. This median value is then assigned to the pixel under consideration. Following this, discard the old value, acquire new samples, and repeat the aforementioned calculation procedure. By replacing each pixel's value with the median of its neighbors, median filtering effectively reduces the impact of random noise while preserving the sharpness of edges. This preservation of edge details is crucial in mammography, where even minor distortions or blurring can obscure important features indicative of breast abnormalities (AlSalman, 2020). Figure 2.1 illustrates an example of a median filter calculation.

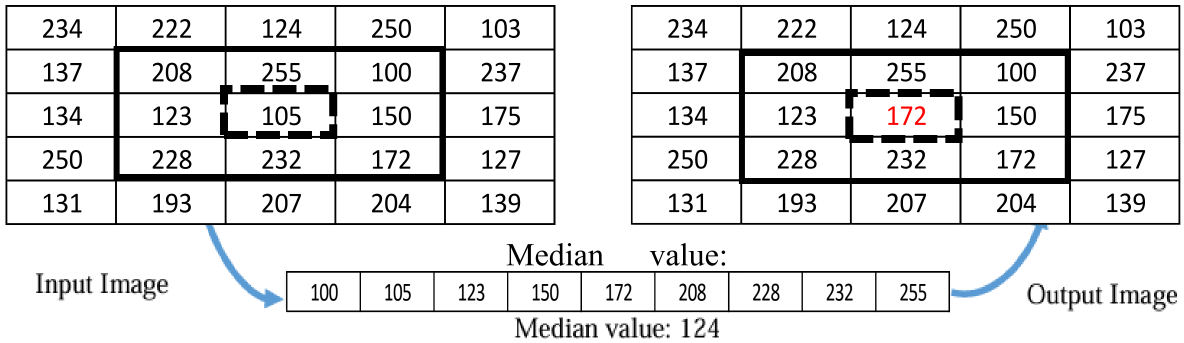


Figure 2.1 Median Filter Calculation

## 2.3 Feature Extraction Methods: Thresholding

Mammography is a crucial imaging technique for the early detection of breast cancer, as it provides detailed images of breast tissues to identify abnormalities such as masses and calcifications. A major challenge in mammography analysis is accurately separating these features from the surrounding breast tissue, which often exhibits varying densities and contrast levels (Yassir Edrees Almalki et al., 2022). Traditional thresholding methods, which apply a single threshold value across the entire image, may not effectively capture these variations, leading to suboptimal segmentation results. Adaptive thresholding techniques have shown promise in addressing these issues by improving the accuracy of feature extraction and classification in mammograms through the adjustment of thresholds based on local image characteristics (Adaptive Threshold-Based Tumor Detection Algorithm For Mammograms Images, 2023).

Adaptive thresholding offers a significant advantage over global thresholding by allowing for local threshold adjustment, which is crucial for handling the heterogeneous nature of breast tissue (Yu, Wang and Zhang, 2023). Due to factors including tissue type and imaging settings, distinct sections of a mammogram may show notable changes in density and contrast. Adaptive thresholding improves the ability to distinguish between normal and aberrant tissues by dynamically modifying the threshold for each location according to its unique local characteristics (Yassir Edrees Almalki et al., 2022). This local modification is especially helpful in differentiating surrounding tissue structures from calcifications, which frequently show up in mammograms as tiny, high-contrast patches (Hanife Avcı and Jale Karakaya, 2023). Techniques such as the adaptive mean, adaptive Gaussian, and locally adaptive methods have been developed to implement this localized thresholding approach, leading to enhanced segmentation accuracy (Yu, Wang and Zhang, 2023).

Adaptive thresholding improves segmentation accuracy in mammography analysis, as numerous studies have shown. As an example, it has been demonstrated that an adaptive thresholding technique based on local image statistics performs better than global thresholding in the correct identification of microcalcifications, which are important markers of early breast cancer. By leveraging local contrast and intensity information, adaptive thresholding reduces false positives and increases the reliability of computer-aided diagnostic (CAD) systems by more precisely distinguishing calcifications from the background (Yu, Wang and Zhang, 2023). Furthermore, adaptive thresholding approaches have been effectively integrated with other image processing techniques, like edge detection and texture analysis, to boost the overall diagnostic performance of mammography analysis systems and improve segmentation outcomes.

One of the notable advantages of adaptive thresholding techniques is their potential for real-time applications due to their computational efficiency (Yu, Wang and Zhang, 2023). Since many adaptive thresholding techniques are computationally light weight by nature, they can be used in real-time processing contexts and embedded devices (Yassir Edrees Almalki et al., 2022). This is especially crucial in healthcare situations when prompt diagnosis and analysis are required (Hanife Avcı and Jale Karakaya, 2023). Adaptive thresholding can help radiologists make timely and well-informed judgements by quickly and accurately segmenting mammograms, which will ultimately enhance patient outcomes. Furthermore, adaptive thresholding algorithms' scalability makes it possible to include them into a variety of CAD systems, expanding their usefulness in a range of healthcare contexts.

Adaptive thresholding greatly improves mammography analysis, although it works best when paired with more sophisticated feature extraction methods (Adaptive Threshold-Based Tumor Detection Algorithm For Mammograms Images, 2023). Combining adaptive thresholding with other methods like texture analysis, edge detection, and deep learning might result in a feature set that is more comprehensive and has a greater classification accuracy (Hanife Avcı and Jale Karakaya, 2023). It has been shown, for example, that integrating adaptive thresholding with convolutional neural networks (CNNs) enhances the identification and classification of breast lesions by gathering complex patterns and textures in mammograms. These hybrid techniques offer a solid basis for accurate and efficient mammography analysis by fusing the best features of adaptive thresholding's segmentation capabilities with the best parts of deep learning feature extraction.

In conclusion, adaptive thresholding is a valuable technique for enhancing mammogram analysis by allowing for local threshold adjustments that address variations in tissue density and contrast. This improves the segmentation accuracy of important features like calcifications, making breast cancer detection more reliable and efficient. Its real-time potential and computational efficiency make it highly applicable in clinical settings. When combined with other advanced feature extraction techniques, adaptive thresholding significantly enhances the detection and classification of suspicious lesions, supporting early breast cancer diagnosis and treatment planning. Future research should aim to develop more advanced algorithms and explore their integration with new technologies to further improve mammogram analysis.

## 2.4 Machine Learning Algorithms

Breast cancer is still a common health issue in the world, which drives ongoing research into cutting-edge technology to improve detection and diagnosis. The field of medical imaging has witnessed the rise of machine learning algorithms as potent instruments in recent times. These algorithms hold great potential for enhancing the precision and effectiveness of breast cancer detection, especially in the identification of calcifications, which are crucial markers of cancerous growth (Chaudhury et al., 2021).

Support Vector Machines (SVM) have attracted a lot of interest because of its adaptability and efficiency in categorization applications. SVM algorithms have proven to be quite accurate in the field of breast cancer detection, especially when used with thermographic pictures (Khalid et al., 2023). Furthermore, SVM has demonstrated its efficacy in detecting the presence and severity of breast artery calcifications (BACs) from mammograms, demonstrating its usefulness in assisting with early diagnosis and treatment planning (Khan and Giovanni Luca Masala, 2023).

Artificial Neural Networks (ANN), including their versions like the Back-propagation neural network, have demonstrated impressive performance in the breast calcification categorization. ANN algorithms have demonstrated good accuracy rates by utilizing textural data collected from fractal-based and Gabor wavelet-based approaches, highlighting their potential for accurate diagnostic assessment. Moreover, early detection and diagnosis of breast cancer by ANN has the potential to improve patient outcomes by enabling prompt intervention and individualized treatment plans (Khalid et al., 2023).

Random forests have become essential tools for identifying breast calcifications and detecting breast cancer because of their exceptional adaptability and efficiency. Notably, with reported accuracy rates reaching up to 98%, these algorithms have proven superior to other machine learning competitors in accurately categorizing breast tumours as benign or malignant. They are exceptionally skilled at working with complex datasets, like mammography scans, because they can handle a large number of features and variables without becoming overfit (Anisha et al., 2021). In addition, Random Forests are very good at identifying nonlinear correlations and interactions between variables, which improves their ability to extract relevant information from complicated mammography data. Researchers have successfully isolated the most pertinent features for breast cancer detection by utilizing sophisticated feature selection approaches in conjunction with the Random Forest algorithm. This has greatly improved the identification and characterization of breast calcifications (Hasan, Sihem Chaabouni and Fakhfakh, 2023).

Convolutional Neural Networks (CNNs) have emerged as a revolutionary tool in breast cancer detection and characterization of breast calcifications, as revealed in the search results. Firstly, CNN algorithms have showcased promising efficacy in detecting breast arterial calcifications from mammographic images, alongside demonstrating high accuracy and specificity in distinguishing between benign and malignant breast calcifications based on textural features (Guevara-Ponce et al., 2023). Their transformative potential in medical imaging analysis is highlighted by their unmatched powers in image recognition and classification, which outperform conventional machine learning techniques. CNN-based techniques provide a compelling answer for breast cancer diagnosis by improving diagnostic accuracy and clinical decision-making with their precision. CNN models also offer the benefit of automatically identifying pertinent features from mammography pictures, which gets around the drawbacks of expert-driven analysis and manual feature engineering (Masud, Eldin and M Shamim Hossain, 2020). When taken as a whole, these results highlight the important role that CNN algorithms have had in improving the identification of breast cancer and characterizing breast calcifications, which represents a critical development in medical imaging technology.

A wide range of advantages and skills are shown when machine learning algorithms for calcification classification and breast cancer diagnosis are compared. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) emerge as major participants, exhibiting remarkable sensitivity and accuracy in both tasks (Ragab et al., 2019). SVM predicts breast cancer risk with high accuracy rates, but ANN successfully discriminates between breast tissues that are normal and pathological. While they have not received much attention, Random Forests demonstrate promise in classification tasks, which is consistent with their potential utility in the identification of breast cancer (Ebrahim Edriss Ebrahim and Feng, 2016). Convolutional Neural Networks (CNN) are a promising technology that can improve patient outcomes and diagnostic capacities due to their excellent specificity and accuracy. CNNs have significant promise in the detection of breast cancer because of their capacity to identify breast artery calcifications and distinguish between benign and malignant calcifications. When combined, these algorithms offer a wide range of tools that can help improve breast cancer management by enhancing diagnostic skills (Ozcan, Aydin and Cetinkaya, 2022).

## 2.5 ROC curve and AUC analysis

The receiver operating characteristic (ROC) curve and the area under the curve (AUC) are additional metrics used to assess the effectiveness of the proposed model across various classification thresholds. The ROC curve represents a probability curve, while the AUC indicates the level of separability, quantifying the model's ability to differentiate between classes (Chan, 2022).

Yang & Berdine (2017) explained that a ROC curve visually illustrates the performance of a classifier system as the discrimination cut-off value varies across the range of predictor variables. As depicted in Figure 2.2, the x-axis, or independent variable, denotes the False Positive Rate (FPR) for predictive data, reflecting the model's 1-specificity. Meanwhile, the y-axis, or dependent variable, signifies the True Positive Rate (TPR) for predictive data, indicating the model's sensitivity.

Yang & Berdine (2017) suggested that each point within ROC space corresponds to a pair of true positive and false positive data instances at a particular discrimination cut-off value for a prediction test. By adjusting this threshold, multiple sets of TPR and FPR values can be generated, forming the basis for constructing a ROC curve (MathWorks, 2023). Typically, in practical scenarios, each choice of discrimination cut-off results in a single point on the ROC graph. Classifiers with curves closer to the top-left corner demonstrate superior performance. Ideally, achieving coordinates (0, 1) indicates perfect results, signifying no false positives and only true positives (Nahm, 2021).

When the discrimination cut-off value for the predictive variable is smaller than the lowest observed value, it forms the point (0, 0) in the ROC space. As this cut-off value increases to encompass more data points, a series of points within the ROC space is generated, potentially connected by a curve. Conversely, the point (1, 1) arises when the discrimination cut-off value surpasses the highest observed value. The diagonal line linking the (0, 0) and (1, 1) points signifies that test predictions are no more accurate than random guesses. The further a point extends from this diagonal line within the ROC space, the stronger the predictive capability of the test. However, if the curve approaches closer to the 45-degree diagonal of the ROC space, the test's accuracy is less accurate (Park et al., 2004).

According to Lalkhen & McCluskey (2008), elevating the cut-off point leads to a reduction in false positives but an increase in false negatives, resulting in a highly specific yet not particularly sensitive. Conversely, lowering the cut-off point decreases false negatives but increases false positives, yielding a highly sensitive yet not particularly specific.

Additionally, a reliable method for assessing the performance of various classifiers involves calculating the area under the ROC curve (AUC). This metric serves multiple purposes, including determining the optimal cut-off value for a specific test and evaluating the performance of multiple alternative tests. A high AUC value, approaching one, indicates excellent separability in the model. The higher the AUC, the more accurately the model distinguishes between classes, effectively predicting zeros as zeros and ones as ones. Analogously, a greater AUC signifies a more precise differentiation between positive and negative classes. Conversely, when AUC is 0.5, the model lacks discrimination ability to distinguish between positive and negative classes. An AUC close to zero suggests that the model is essentially reversing the classes, predicting negative as positive and vice versa (Narkhede, 2018).

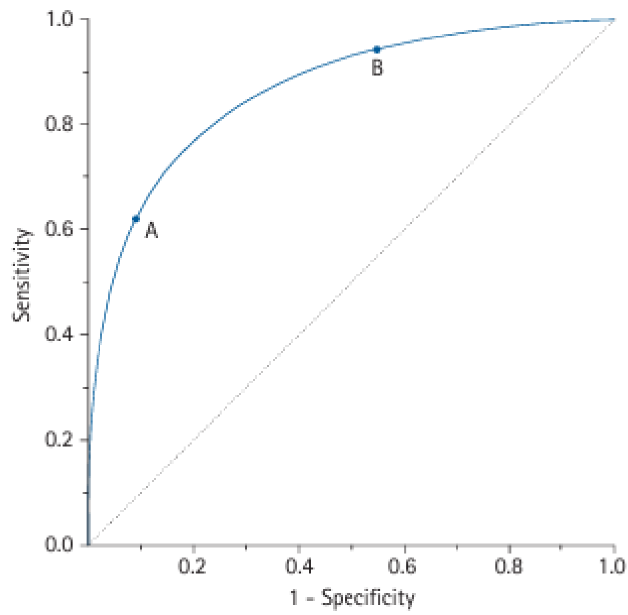


Figure 2.2 ROC Curve (Nahm, 2021)

## 2.6 Performance Evaluation Metrics

Performance evaluation metrics play a vital role in machine learning, serving as essential tools for gauging the accuracy of a model's predictions. These metrics offer quantitative insights into a model's performance on specific datasets, enabling developers to iteratively refine and enhance their models. This study utilized metrics such as accuracy, loss, F-1 score, recall, precision, specificity, sensitivity, and AUC.

Accuracy stands as the most commonly used metric for assessing machine learning model performance. It quantifies the ratio of correct predictions to all predictions made by the model. For instance, if a model accurately predicts 90 out of 100 instances, its accuracy is 90%. Accuracy serves well when dealing with balanced datasets, where instances across classes are approximately equal. However, in the case of imbalanced datasets, accuracy can be misleading, as a model might achieve high accuracy by simply predicting the majority class. It is defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{FP} + \mathrm{FN} + \mathrm{TP} + \mathrm{TN}} \tag{1}$$

where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively, as determined by the classifier's predictions.

Loss, also referred to as error or cost, gauges a model's performance by assessing the disparity between its predictions and the real outcomes. To train the model, a loss function is employed to tweak its parameters, aiming to minimize the gap between predictions and actual results. Typical loss functions comprise mean squared error, mean absolute error, and cross-entropy. The selection of a loss function hinges on the particular problem and the nature of the data under examination. For instance, mean squared error is frequently employed for regression tasks, while cross-entropy is favored for classification challenges (Alake, 2023).

The F1-score assesses a model's performance by considering both precision and recall, computed as the harmonic mean of these two metrics. This metric is used independently to determine the accuracy of test datasets.

$$\mathrm{F1-Score} = 2 \times \frac{\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}}{\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} + \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}} \tag{2}$$

Recall evaluates a model's ability to accurately detect all positive instances. It is determined by dividing the number of true positives by the sum of true positives and false negatives. Recall holds significance in scenarios where overlooking a positive instance can yield notable repercussions, such as in medical diagnosis or fraud detection.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

Precision gauges a model's effectiveness in accurately identifying positive instances while minimizing false positives. This metric is computed by dividing the number of true positives by the sum of true positives and false positives. Precision holds relevance in contexts where false positives can lead to notable repercussions, such as in medical diagnosis or spam detection.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

Specificity, alternatively termed as the true negative rate, evaluates a model's ability to accurately identify negative instances. This metric is determined by dividing the number of true negatives by the sum of true negatives and false positives. Specificity holds significance, particularly in medical diagnosis scenarios, where high specificity guarantees that patients are not misdiagnosed with a condition they do not have.

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{5}$$

Sensitivity, also referred to as the true positive rate, assesses a model's ability to accurately identify positive instances. This metric is computed by dividing the number of true positives by the sum of true positives and false negatives. Sensitivity holds paramount importance in medical diagnosis, where achieving high sensitivity ensures that patients are not overlooked if they indeed have a condition.

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

The AUC evaluates a model's performance in terms of both sensitivity and specificity. This measure is obtained by plotting the true positive rate against the false positive rate across various thresholds and then calculating the area under the curve. AUC proves valuable in scenarios where both sensitivity and specificity hold significance, such as in medical diagnosis or credit risk assessment.

## 2.7 Ensemble Learning

Ensemble learning is a sophisticated machine learning technique that enhances predictive performance by amalgamating the outputs from multiple models. These separate models, also known as base learners or base models, each add to the final prediction, resulting in more reliable and accurate outcomes overall (Keita, 2018). The fundamental tenet of ensemble learning is that the strengths of various models can be used to offset the shortcomings of individual models. This technique serves to reduce the overall generalizability of the ensemble and smooths out individual model biases, which shows to be especially useful when working with noisy and complicated datasets (Brownlee, 2021).

There are several ensemble techniques, each intended to take advantage of unique advantages. Bagging, also known as Bootstrap Aggregating, is the process of averaging the results of many models that have been trained on different subsets of the training data. The primary way that this strategy lowers variance is by averaging the errors of individual models (Brownlee, 2021). Boosting, on the other hand, progressively trains models to correct the mistakes made by their forebears in an effort to reduce bias and variance. A strong aggregate model is produced by models in a boosting setup learning from the errors of previous models. Another strategy is stacking, which is training a meta-model to combine the predictions of multiple base models. This method frequently produces better results by identifying intricate relationships that individual models would miss (Keita, 2018).

Group learning offers numerous advantages. By combining the strengths of multiple algorithms, ensemble models often achieve higher prediction accuracy than single models. This combination results in more robust models that handle noise better and generalize more effectively to new data. Additionally, ensemble methods are highly versatile, applicable to both classification and regression problems (Simplilearn, 2021). However, there are some downsides to ensemble learning. One major drawback is interpretability; it can be challenging to understand the contributions of each base model within the ensemble. Furthermore, the computational cost of training multiple models and merging their predictions can be substantial, especially with large datasets. Overfitting is another potential issue, particularly if the base models are overly complex or if the ensemble is too large (Mwiti, 2022).

Ensemble learning is applied across various fields, leveraging multiple models to enhance prediction performance. In image classification, ensemble methods improve accuracy by combining outputs from several models, each capturing unique details and subtleties, leading to a more comprehensive understanding and better predictions (Brownlee, 2021). Similarly, in natural language processing (NLP), ensemble techniques enhance performance in tasks like sentiment analysis, language translation, and text categorization by using diverse linguistic patterns identified by different models. This diversity allows for a better grasp of complex linguistic structures and semantics. For time series forecasting, ensemble approaches generate more accurate and reliable predictions by merging forecasts from different models, effectively handling the inherent variability and complexity of time series data (Simplilearn, 2021).

The implementation of ensemble learning in Python is facilitated by numerous tools and packages, making it accessible and practical for practitioners. Scikit-Learn, a versatile machine learning framework, provides robust implementations of various ensemble techniques, such as boosting, stacking, and bagging. XGBoost is favored for its efficiency and superior performance in gradient boosting, while LightGBM is known for its speed and effectiveness, and CatBoost excels in processing categorical features, broadening the scope of ensemble learning applications. These tools empower data scientists and machine learning engineers to experiment with and apply ensemble methods easily, harnessing their full potential to tackle complex predictive modelling challenges across different domains. Ensemble learning is a vital technique in modern machine learning, combining the strengths of multiple models to enhance prediction accuracy, robustness, and generalizability (Keita, 2018).

# CHAPTER 3

## METHODOLOGY

### 3.1 Study Design

This is a retrospective study to develop a Computer Aided Diagnosis (CAD) system to distinguish between benign and malignant breast calcifications. 384 mammogram images were obtained from HUSM PACS system from 2021 until 2023 based on the inclusion and exclusion criteria. The system evaluation in this study is divided into four major components which are cancer breast image data input, pre-processing, detection and feature extraction, and classification.

### 3.2 Study Location

The image dataset will be collected from picture archiving and communication system (PACS) in Hospital Universiti Sains Malaysia (HUSM), Kubang Kerian.

### 3.3 Selection Criteria

#### 3.3.1 Inclusion Criteria

Women diagnosed with breast cancer who have undergone mammogram in standard mediolateral oblique (MLO) view and cranial caudal (CC) view.

#### 3.3.2 Exclusion Criteria

i. Women with previous breast surgery.
ii. Breast cancer women post neoadjuvant chemotherapy.
iii. Women with breast implant or breast filler injection.
iv. Mammogram images without calcifications.

### 3.4 Sample Size Estimation

The sample size will measure the number of breast cancer images dataset used in this study. Cochran's formula (1977) is used to determine the sample size of this research with a 95% confidence level within 0.5 of true value. The equation used is as follow:

$$n = \left[\frac{z}{\Delta}\right]^2 p(1-p) \tag{7}$$

$n =$ sample size

$z =$ value representing the desired confidence level

$\Delta =$ precision or true value

$p =$ anticipated dataset proportion

The confidence level corresponds to a Z-score. The constant value for the 95% confidence level is 1.96. As there is no previous study yet, the true value of the sample size is assumed as 0.05. The estimated dataset proportion p=0.5, which will produce the largest possible sample size that would be required. The calculation is as follow:

$$n = \left[\frac{1.96}{0.05}\right]^2 0.5(1-0.5) \tag{8}$$

$$n = 384$$

### 3.5 Data Collection

The data collection for this study will be conducted at the Women Imaging Suite (WISH) of Hospital Universiti Sains Malaysia (HUSM). Initially, data will be extracted from the Laboratory Information System (LIS) for cases of confirmed breast cancer. This data will then be cross verified with the Picture Archiving and Communication System (PACS) at HUSM to identify mammograms that show calcification and adhere to the inclusion criteria. The selection process will focus on mammogram images taken within the past four years, specifically from September 2020 to December 2023. Only mammograms in PACS that