

**IMAGE QUALITY EVALUATION OF ACR (RMI
156) PHANTOM USING MACHINE LEARNING IN
DIGITAL BREAST TOMOSYNTHESIS (DBT)**

NORSYAFIQAH ANDRIANA BINTI OTHMAN

SCHOOL OF HEALTH SCIENCES

UNIVERSITI SAINS MALAYSIA

2025

**IMAGE QUALITY EVALUATION OF ACR (RMI
156) PHANTOM USING MACHINE LEARNING IN
DIGITAL BREAST TOMOSYNTHESIS (DBT)**

by

NORSYAFIQAH ANDRIANA BINTI OTHMAN

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

JUNE 2025

CERTIFICATE

This is to certify that the dissertation entitled **“IMAGE QUALITY EVALUATION OF ACR (RMI 156) PHANTOM USING MACHINE LEARNING IN DIGITAL BREAST TOMOSYNTHESIS (DBT)”** is the bona fide record of research work done by **NORSYAFIQAH ANDRIANA BINTI OTHMAN** during the period of from October 2024 to June 2025 under my supervision. I have read this dissertation and that in my opinion it conforms to acceptable standards of scholarly presentation and it 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:

Field-Supervisor:

.....

Madam Siti Aishah Abd Aziz

Senior Lecturer

School of Health Sciences

Universiti Sains Malaysia

Health Campus

16150 Kubang Kerian

Kelantan, Malaysia

.....

Pn. Siti Salwa Hassan

Radiographer

Radiology Department

Hospital Pakar Universiti Sains Malaysia

16150 Kubang Kerian

Kelantan, Malaysia

Date: June 2025

DECLARATION

I, Norsyafiqah Andriana binti Othman hereby declare that the dissertatio entitled **“IMAGE QUALITY EVALUATION OF ACR (RMI 156) PHANTOM USING MACHINE LEARNING IN DIGITAL BREAST TOMOSYNTHESIS (DBT)”** is the result of my own investigations, except where otherwise sated 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.

.....

NORSYAFIQA ANDRIANA BINTI OTHMAN

Date: June 2025

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. Undertaking this research has been one of the most challenging and rewarding experiences of my academic journey. Completing this work would not have been possible without the support, encouragement, and guidance of many people, to whom I am deeply grateful. I owe a great debt of gratitude to my family. To my parents, Othman bin Yatim and Na Azura binti Abd Latif, thank you for your unconditional love, sacrifices, and unwavering belief in my potential.

I would like to express my heartfelt gratitude to my supervisor, Madam Siti Aishah Binti Abdul Aziz, for her continuous guidance, insightful feedback, and patience over the two semester of my study. Her mentorship not only improved the quality of this research but also helped shape my approach to inquiry, critical thinking, and scholarly work. My sincere thanks go to the staff of the Radiology Department at Hospital Pakar Universiti Sains Malaysia (HPUSM) whose assistance with equipment, and data collection was invaluable. Special appreciation goes to Dr. Muhammad Akmal bin Remli, Director of Institute For Artificial Intelligence and Big Data (AIBIG), UMK, and their PHD's students, Ainin Sofia and Meor Muhammad Muaz for the guidance in creating the code for this project.

I am deeply thankful to fellow friends, Amirah Atikah, Muhammad Haziq, and Nurul Syuhaida for helping me in this project, my close friends, Ahmad Amsyari, Aimi Fatihah, Azmin Nur Aina, and Akma Fatini, for their love, emotional, financial, and mental support throughout this thoughtful journey. I could not have completed this study without them.

TABLE OF CONTENTS

CERTIFICATE.....	iii
DECLARATION	iv
ACKNOWLEDGEMENT	v
LIST OF FIGURES	xi
LIST OF TABLES	xiii
LIST OF EQUATION.....	xiv
LIST OF ABBREVIATIONS	xv
ABSTRAK.....	xvii
ABSTRACT.....	xix
CHAPTER 1	1
INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement.....	3
1.3 Objective.....	4
1.3.1 General Objective	4
1.3.2 Specific Objectives	4
1.4 Hypothesis	4
1.4.1 Null Hypothesis	4
1.4.2 Alternative Hypothesis	4
1.5 Significant of Study	4

CHAPTER 2	6
LITERATURE REVIEW	6
2.1 Importance of Image Quality in Digital Breast Tomosynthesis (DBT)	6
2.2 Applications of Machine Learning to DBT Quality Control	7
2.3 Machine Learning Algorithms for Image Quality Evaluation	8
2.3.1 Support Vector Machines (SVM)	8
2.3.2 K-Nearest Neighbors (KNN)	9
2.3.3 Random Forest (RF)	9
2.4 Previous Study	10
2.5 Comparison of Machine Learning Algorithms for Image Quality Assessment	
11	
2.6 Image Preprocessing to Improve Image Quality	12
2.5.1 Noise Reduction	12
2.5.2 Contrast Enhancement	12
2.5.3 Region-of-Interest (ROI) Segmentation & Feature Extraction	13
2.7 Image Quality Metrics and Assessment Methods	13
2.6.1 Spatial Resolution	14
2.6.2 Signal-to-Noise Ratio (SNR)	14
2.6.3 Contrast-to-Noise Ratio (CNR)	14
2.6.4 Detectability	15
2.6.5 Accuracy	15
2.6.6 Loss	15

2.6.7	Precision.....	16
2.6.8	Recall (Sensitivity)	16
2.6.9	Specificity	16
2.6.10	F1 Score	17
2.8	ROC curve and AUC analysis	17
2.9	Comparison of Manual and Automated Evaluation Methods.....	19
CHAPTER 3		20
METHODOLOGY		20
3.1	Study Design.....	20
3.2	Study Location.....	20
3.3	Selection Criteria	20
3.3.1	Inclusion.....	20
3.3.2	Exclusion	20
3.4	Data Collection	21
3.5	Study Instruments	21
3.5.1	ACR (RMI 156) Phantom.....	21
3.5.2	Siemens Digital Breast Tomosynthesis (DBT) System	22
3.5.3	Pictures Archiving and Communication System (PACS).....	23
3.5.4	MATLAB Software.....	24
3.6	Method	25
3.6.1	Data Collection	26
3.6.2	Image Processing	26

3.6.3	Data Augmentation	26
3.6.4	Segmentation and Feature Extraction	27
3.6.5	ML Classification.....	29
3.6.6	ML Evaluation	31
3.7	Study Flowchart.....	33
3.8	Data Analysis	35
CHAPTER 4		36
RESULT AND DISCUSSION.....		36
4.1	Mammogram Raw Dataset Images.....	36
4.2	Output Images After Preprocessing	37
4.3	Segmentation and Feature Extraction	39
4.4	Augmentation Techniques	41
4.5	Model Performance Evaluation	41
4.5.1	Model Training	42
4.5.2	Model Training	45
4.6	Statistical Analysis.....	51
4.7	Limitation of Study.....	53
4.8	Future Recommendations	54
CHAPTER 5		55
CONCLUSION.....		55
REFERENCES		56
APPENDICES		61

APPENDIX A: Phantom Mammogram	61
APPENDIX B: Augmentation of The Preprocessed Images	62
APPENDIX C: MATLAB Code	63

LIST OF FIGURES

Figure 1 ROC Curve (Pierian Training, 2022).	18
Figure 2 Examples of augmented images with the data scaling process applied (Seo, Kim and Kim, 2025).	27
Figure 3 Example of segmentation of the RMI 156 Phantom mammography (Sundell et al., 2022).	28
Figure 4 Illustration of GLCM construction process (Gomede, 2024).....	29
Figure 5: 10-fold cross-validation method process (Andrade et al., 2020).	32
Figure 6 QC in DBT using machine learning classification pipeline.	33
Figure 7 Confusion matrix for SVM model training	43
Figure 8 Confusion matrix for KNN model training.	44
Figure 9 Confusion matrix for RF model training.....	44
Figure 10 Comparison of mean accuracy for SVM, KNN, and RF model.....	45
Figure 11 Confusion matrix for SVM model testing	46
Figure 12 Confusion matrix for KNN model testing.....	46
Figure 13 Confusion matrix for RF model testing.....	47
Figure 14 ROC and AUC analysis of of SVM model.....	47
Figure 15 ROC and AUC analysis of of KNN model.....	48
Figure 16 ROC and AUC analysis of of RF model.	48
Figure 17 Positioning of the RMI 156 Phantom.....	61
Figure 18 RMI 156 Phantom	61
Figure 19 Augmented images on slices 33 th	62
Figure 20 Augmented images on slices 34 th	62
Figure 21 Augmented images on slice 35 th	62
Figure 22 Augmented images on slice 36 th	62

Figure 23 Augmented images on slice 37 th	62
Figure 24 Augmented images on slice 38 th	62

LIST OF TABLES

Table 1 Raw dataset of mammogram QC phantom images.....	36
Table 2 Sample of phantom mammogram image before and after result.	37
Table 3 Feature extraction from preprocessed images for a few fiber fiber, specks, and masses.	39
Table 4 Summary of accuracy, precision, recall, specificity, F1 score, and training time for SVM, KNN, and RF model for model testing.	42
Table 5 Summary of accuracy, recall, specificity, F1 score, and training time for SVM, KNN, and RF model for model testing.	50
Table 6 Accuracy of SVM, KNN, and RF across 10-fold cross-validation.....	51
Table 7 Friedman Test Result.....	52
Table 8 Wilcoxon Signed-Rank Test Results.....	53

LIST OF EQUATION

Equation 1 Calculation of signal noise ratio (SNR) (Leyton et al., 2011).....	14
Equation 2 Calculation of contrast noise ratio (CNR) (Leyton et al., 2011)	14
Equation 3 Calculation of Accuracy (James, 2024).....	15
Equation 4 Calculation of Precision (James, 2024).....	16
Equation 5 Calculation of Recall (James, 2024).....	16
Equation 6 Calculation of Specificity (Ebrary, 2020).....	16
Equation 7 Calculation of F1-Score (James, 2024).	17

LIST OF ABBREVIATIONS

SNR	Signal-to-Noise Ratio
CNR	Contrast-to-Noise Ratio
MPV	Mean Pixel Value
SD	Standard Deviation
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
AUC	Area Under the Curve
F1 Score	Harmonic mean of Precision and Recall
ROI	Region of Interest
GLCM	Gray-Level Co-occurrence Matrix
GPU	Graphics Processing Unit
GUI	Graphical User Interface
DBT	Digital Breast Tomosynthesis
DICOM	Digital Imaging and Communications in Medicine
FFDM	Full-Field Digital Mammography
CNN	Convolutional Neural Network
ML	Machine Learning
PACS	Picture Archiving and Communication System
QA	Quality Assurance

RF	Random Forest
RMI 156	Radiological Imaging Phantom 156
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
KNN	K-Nearest Neighbour

PENILAIAN KUALITI IMEJ FANTOM ACR (RMI 156)
MENGGUNAKAN PEMBELAJARAN MESIN DALAM TOMOSINTESIS
PAYUDARA DIGITAL (DBT)

ABSTRAK

Jaminan kualiti (QA) biasanya bergantung pada RMI 156 Phantom untuk menilai parameter seperti resolusi, kontras dan hingar (noise), tetapi penilaian manual semasa adalah subjektif, berbeza-beza antara penilai dan memakan masa, yang membawa kepada ketidakkonsistenan. Batasan ini menjejaskan kebolehpercayaan proses QA, yang berpotensi menjejaskan ketepatan diagnostik. Oleh itu, pembelajaran mesin, “machine learning” (ML) menawarkan peluang untuk mengautomasikan dan menyeragamkan penilaian kualiti imej DBT, meningkatkan kecekapan, konsistensi dan ketepatan. Kajian ini meneroka penggunaan pembelajaran mesin (machine learning, ML) dalam mengautomasikan penilaian kualiti imej DBT menggunakan fantom ACR (RMI 156). Objektif utama kajian adalah untuk membangunkan rangka kerja berasaskan ML yang mampu menilai kualiti imej dengan ketepatan, konsistensi, dan kecekapan yang lebih baik berbanding kaedah manual konvensional. Imej DBT yang diperolehi daripada pendedahan fantom telah diproses menggunakan MATLAB, termasuk prapemprosesan, segmentasi, pengekstrakan ciri, dan penggandaan data. Tiga model klasifikasi iaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Random Forest (RF) telah dilatih dan dinilai menggunakan pensahihan silang 10-lipat (10-fold cross-validation). Hasil kajian menunjukkan bahawa semua model mencapai ketepatan yang tinggi, dengan RF menunjukkan prestasi sedikit lebih baik berbanding model lain. SVM menunjukkan prestasi terbaik dari segi recall dan skor F1, terutamanya dalam mengesan kelas minoriti. Model KNN (0.10 Precision, 0.10 Recall, 0.10 Skor F1) dan RF (0 Precision, 0 Recall, 0 Skor F1) masing-masing mencapai ketepatan 93.89%, diikuti oleh SVM (0.033 Precision,

0.10 Recall, 0.05 Skor F1) dengan ketepatan 87.04%. Dari segi masa latihan, SVM (0.0149s) dan KNN (0.0216s) adalah lebih pantas, manakala model RF memerlukan lebih masa (0.9198s) kerana struktur kesatuannya. Walaupun keputusan yang diperoleh adalah memberangsangkan, kajian ini menghadapi beberapa kekangan seperti ketidakseimbangan set data dan pengecualian data klinikal. Dapatan kajian mencadangkan bahawa ML merupakan satu penyelesaian yang berpotensi dalam kawalan kualiti imej DBT, dan kajian lanjutan disyorkan dengan penggabungan set data yang lebih besar serta penggunaan teknik pembelajaran mendalam (deep learning) untuk meningkatkan kebolegunaan dan aplikasi dalam situasi sebenar.

IMAGE QUALITY EVALUATION OF ACR (RMI 156) PHANTOM USING MACHINE LEARNING IN DIGITAL BREAST TOMOSYNTHESIS (DBT)

ABSTRACT

Quality assurance (QA) typically relies on the RMI 156 Phantom to evaluate parameters like resolution, contrast, and noise, but current manual assessments are subjective, vary among evaluators, and are time-consuming, leading to inconsistencies. These limitations compromise the reliability of QA processes, potentially affecting diagnostic accuracy. This study explores the application of ML in automating image quality assessment for DBT using the ACR (RMI 156) phantom. The main objective was to develop an ML-based framework capable of evaluating image quality with improved accuracy, consistency, and efficiency compared to conventional manual methods. DBT images acquired from phantom exposures were processed using MATLAB, including preprocessing, segmentation, feature extraction, and data augmentation. Three classification models which are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) were trained and evaluated using 10-fold cross-validation. Results showed that all models achieved high accuracy, with RF slightly outperforming others. SVM demonstrated superior recall and F1 score, particularly in detecting minority class instances. Both KNN (0.10 Precision, 0.10 Recall, 0.1 F1 Score) and Rf (0 Precision, Recall, 0 F1 Score) achieved the high accuracy of 93.89%, followed by SVM (0.033 Precision, 0.10 Recall, 0.05 F1 Score), achieved the accuracy of 87.04%. In terms of training time, SVM (0.0149s) and KNN (0.0216s) were faster, while the RF model required more time (0.9198s) due to its ensemble structure. Despite achieving promising results, the study faced limitations such as dataset imbalance and the exclusion of clinical data. The findings suggest that ML offers a solution for DBT image quality

control and recommend further research incorporating larger datasets and deep learning techniques to enhance generalisability and real-world applicability.

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Breast cancer remains one of the leading causes of cancer-related morbidity and mortality among women worldwide, representing a major public health challenge across both developed and developing nations. According to the World Health Organization (WHO, 2021), breast cancer has surpassed lung cancer as the most commonly diagnosed cancer globally, accounting for approximately 2.3 million new cases annually. The burden of this disease has prompted continuous advancements in screening and diagnostic modalities aimed at improving early detection, which is essential for effective treatment and improved survival outcomes.

One such technological advancement is Digital Breast Tomosynthesis (DBT), which has revolutionised breast imaging practices by providing three-dimensional (3D) volumetric images of the breast. Unlike conventional two-dimensional (2D) full-field digital mammography (FFDM), which captures a single planar image, DBT acquires multiple low-dose projection images at different angles that are reconstructed into thin image slices. This tomographic approach significantly enhances lesion visibility, especially in dense breast tissue, while simultaneously reducing tissue overlap—a major limitation in 2D mammography. As a result, DBT has been associated with improved cancer detection rates and a reduction in recall rates due to false-positive findings (Rafferty et al., 2013). Consequently, DBT is now widely adopted in routine clinical practice for breast cancer screening and diagnostic workups.

However, the increased complexity of DBT systems and their image reconstruction processes necessitate rigorous and consistent quality assurance (QA)

protocols to ensure optimal performance and diagnostic accuracy. QA procedures are essential to monitor parameters such as image contrast, spatial resolution, noise characteristics, and artifact presence. A critical component in these procedures is the use of standardised phantoms, such as the American College of Radiology (ACR) RMI 156 phantom. This phantom mimics key anatomical and pathological features of the human breast, including embedded structures that simulate fibers, specks (calcifications), and masses. These structures serve as visual markers for evaluating system performance and are integral in detecting changes or degradations in image quality over time (American College of Radiology [ACR], 2018).

Traditionally, the evaluation of phantom images is performed manually by trained radiographers or medical physicist, who visually inspect and score the visibility of test objects within the phantom. While this method is widely accepted, it has several limitations: it is inherently subjective, time-intensive, and susceptible to inter- and intra-observer variability. These limitations can compromise the consistency and reliability of QA assessments, especially in high-volume screening programs where standardization is paramount.

In response to these challenges, the integration of artificial intelligence (AI) and machine learning (ML) into medical image analysis has gained increasing attention. ML algorithms, particularly those trained for image recognition and pattern analysis, offer a promising avenue for automating the evaluation of DBT phantom images. These algorithms can be trained to detect and classify QA features such as fibers, specks, and masses, assess their visibility, and quantify image quality metrics with high precision. Automation not only enhances the objectivity and reproducibility of the QA process but also significantly reduces the time and human effort required.

This study aims to explore and develop ML-based techniques for the automated evaluation of ACR RMI 156 phantom images acquired using DBT systems. By leveraging image preprocessing, segmentation, and feature extraction methods in combination with classification model such as SVM, KNN, and RF. This research seeks to assess the viability of machine learning as a reliable and efficient tool for quality control in digital breast imaging. The outcomes of this study are expected to contribute toward more standardised QA protocols and pave the way for broader AI integration in radiological workflows.

1.2 Problem Statement

Breast cancer is a leading cause of death among women, and early detection through imaging is vital for effective treatment (WHO, 2021). Digital Breast Tomosynthesis (DBT) has proven more effective than standard 2D mammography in detecting cancer; however, ensuring consistent image quality remains a challenge. Quality assurance (QA) typically relies on the RMI 156 Phantom to evaluate parameters like resolution, contrast, and noise, but current manual assessments are subjective, vary among evaluators, and are time-consuming, leading to inconsistencies. These limitations compromise the reliability of QA processes, potentially affecting diagnostic accuracy. Recent developments in machine learning (ML) offer an opportunity to automate and standardize DBT image quality assessment, improving efficiency, consistency, and accuracy (Chen et al., 2020). This study investigates the application of ML in evaluating RMI 156 Phantom images to enhance QA procedures and support the delivery of high-quality diagnostic imaging for early breast cancer detection.

1.3 Objective

1.3.1 General Objective

To develop a machine learning framework for the automated evaluation of DBT image quality using the RMI 156 phantom.

1.3.2 Specific Objectives

1. To apply image processing for segmentation and improvement of image quality of RMI 156 phantom using the MATLAB software.
2. To compare the accuracy on different machine learning model such as SVM, K-NN, and Random Forest.
3. To verify the effectiveness of the machine learning performance in automating the DBT QC.

1.4 Hypothesis

1.4.1 Null Hypothesis

There is no significant difference in accuracy among the three models which are SVM, KNN, and RF

1.4.2 Alternative Hypothesis

There is a significant difference in the accuracy among the three models which are SVM, KNN, and RF.

1.5 Significant of Study

This study will employ an experimental approach to develop a machine learning model in assessing Digital Breast Tomosynthesis (DBT) image quality using the ACR (RMI 156) phantom. The significance of this research lies in addressing the inherent subjectivity and time consumption of manual QA evaluations, which are still widely used in clinical radiology settings. By automating the quality evaluation process, this study contributes

to improving diagnostic consistency, reducing human error, and streamlining radiology workflows. Furthermore, the findings are expected to promote the integration of AI-based tools in clinical QA, paving the way for scalable and standardised solutions in breast imaging technologies.

CHAPTER 2

LITERATURE REVIEW

2.1 Importance of Image Quality in Digital Breast Tomosynthesis (DBT)

Breast cancer remains a leading cause of mortality among women worldwide. Early detection and accurate diagnosis are crucial for effective treatment and improved prognosis. Digital Breast Tomosynthesis (DBT) has emerged as an advanced imaging modality that enhances cancer detection by producing a three-dimensional (3D) reconstruction of breast tissue, offering greater lesion visibility than traditional two-dimensional (2D) mammography (Mendelson et al., 2013). The primary advantage of DBT is its ability to minimize tissue overlap, which is often a limitation in conventional mammography, particularly for women with dense breasts.

The diagnostic utility of DBT is heavily dependent on the quality of the acquired images. Key image quality parameters include spatial resolution, signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and lesion detectability. High spatial resolution enables the detection of minute structures such as microcalcifications and architectural distortions, which may signify early-stage malignancies (Chen et al., 2020). Insufficient resolution can obscure critical details, leading to delayed diagnosis or misinterpretation.

Similarly, SNR and CNR are vital metrics that determine the clarity and distinguishability of breast tissue structures. A high SNR indicates a clear distinction between the signal (tissue structures) and background noise, which is essential for image interpretation. A high CNR ensures that different tissue types are adequately differentiated, thereby improving lesion detection accuracy (Sabel et al., 2015). These metrics are particularly relevant when imaging dense breast tissues where the contrast between normal and pathological areas is often subtle.

Image quality also influences radiographer confidence and diagnostic accuracy. Low-quality images may lead to false negatives or necessitate repeat imaging, increasing patient exposure to radiation and healthcare costs. Therefore, maintaining high image quality through routine quality assurance (QA) is essential to ensure the reliability and safety of DBT.

2.2 Applications of Machine Learning to DBT Quality Control

The integration of artificial intelligence (AI) and machine learning (ML) in medical imaging has revolutionised quality control processes. ML techniques, particularly deep learning, are increasingly employed to automate and standardize image quality assessment in DBT, thereby reducing subjectivity and improving consistency.

Convolutional Neural Networks (CNNs), a class of deep learning algorithms, are especially effective in image classification and pattern recognition tasks. CNNs can be trained to identify and quantify image artefacts such as motion blur, misalignment, and inadequate exposure. By analysing large datasets, these models learn to distinguish between high-quality and poor-quality images based on predefined features (Shin et al., 2018).

ML models also facilitate personalised imaging protocols by predicting optimal acquisition parameters tailored to patient-specific characteristics like breast density and size (Lee et al., 2019). This approach not only ensures consistent image quality but also minimizes radiation exposure by avoiding unnecessary retakes.

A study by You et al. (2020) demonstrated the feasibility of real-time DBT image quality evaluation using ML. The model successfully identified regions with low contrast or blurring, which could indicate suboptimal imaging conditions or equipment

malfunction. Such real-time feedback mechanisms enhance workflow efficiency and reduce the risk of diagnostic errors.

Moreover, ML applications extend to automated QA in phantom-based assessments. By leveraging standardised phantoms like the ACR (RMI 156), ML algorithms can evaluate essential image quality parameters, offering objective and repeatable assessments that align with regulatory standards.

2.3 Machine Learning Algorithms for Image Quality Evaluation

Various ML algorithms have been explored for automated image quality evaluation, with Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Random Forest (RF) being the most commonly utilized.

2.3.1 Support Vector Machines (SVM)

SVM is a supervised learning model that classifies data by finding the optimal hyperplane that separates different classes. In the context of image quality evaluation, SVMs can be trained on labeled phantom images to distinguish between acceptable and suboptimal quality based on features like texture, edge sharpness, and intensity distribution.

Ho et al. (2022) employed an SVM framework to assess ACR phantom images, using a dataset labeled by expert physicists. The model achieved high accuracy in identifying visibility patterns of fibers, specks, and masses, demonstrating its potential for routine QA applications. The use of SVMs in this context ensures that evaluations are based on objective criteria, reducing inter-observer variability.

2.3.2 K-Nearest Neighbors (KNN)

The KNN is a nonparametric method in machine learning used for classification and regression tasks. It involves storing training samples and computing the distances to find the k closest neighbors to make predictions for new data points (GeeksforGeeks, 2017). In medical image quality assessment, including Digital Breast Tomosynthesis (DBT), KNN can be utilized to evaluate image features and classify them into quality categories based on proximity to known labelled data points in a multidimensional feature space. Despite its simplicity, KNN's performance can be influenced by the choice of k , the presence of noisy or irrelevant features, and the curse of dimensionality. However, with proper feature selection and normalization, KNN can serve as an effective baseline classifier or a component in ensemble learning frameworks for quality assurance in DBT systems.

Recent studies have explored the use of KNN in radiological applications. For example, Zhou et al. (2021) applied KNN to classify mammographic image patches and reported promising accuracy when combined with preprocessing and dimensionality reduction techniques. Similarly, Singh et al. (2022) demonstrated that KNN performed competitively in breast cancer classification when applied to texture-based features.

2.3.3 Random Forest (RF)

Random Forest is an ensemble learning technique that constructs a multitude of decision trees during training and outputs the mode of the classes for classification. RFs are particularly robust to overfitting and noise in datasets, making them ideal for medical imaging applications where data variability is common. In DBT quality evaluation, RF has been used to classify image quality based on extracted radiomic and texture features

(Chen et al., 2021). The combination of feature importance ranking and decision boundaries enhances model transparency and interpretability.

2.4 Previous Study

Hejduk et al. (2023) developed an artificial intelligence (AI) platform utilising deep convolutional neural networks (dCNNs) to automate the quality assessment (QA) of Digital Breast Tomosynthesis (DBT) images. The study analysed 11,733 DBT images, as well as synthetic mammography images, to evaluate features such as breast positioning, sharpness, and image clarity. The AI platform was trained and validated on these images, showing its potential to automate routine QA tasks with greater consistency compared to manual methods. The results demonstrated that the platform could reduce inter-observer variability and enhance the efficiency of QA processes. However, a key limitation of the study was its focus on clinical images, as the platform was not tested with phantom-based images, which are typically used for standardised QA evaluations.

Chen et al. (2020) conducted a systematic review of machine learning applications in medical image quality assessment. The review covered various ML techniques, including deep learning, support vector machines, and random forests, and highlighted their potential to automate the evaluation of image quality metrics such as contrast-to-noise ratio (CNR) and spatial resolution in mammography and DBT. The authors emphasised the importance of standardising QA processes in medical imaging, which could be facilitated by the use of ML algorithms. However, the review lacked practical implementation details for phantom-based QA in DBT, which limits its applicability to the specific context of DBT image evaluation. Furthermore, while the review acknowledged the potential of ML to improve efficiency, it did not provide specific insights into how these algorithms could be adapted for different DBT systems or imaging conditions (Chen et al., 2020).

Sundell et al. (2022) developed a CNN-based model to score ACR phantom images, achieving accuracy levels above 95% under standardised imaging conditions. The model effectively mimicked human scoring, providing consistent and rapid evaluations. CNNs offer scalability and adaptability, making them suitable for implementation across different imaging systems and clinical environments. CNNs have become the cornerstone of deep learning in medical imaging due to their superior performance in visual data analysis. CNNs consist of multiple layers that automatically extract hierarchical features from input images. In QA tasks, CNNs can learn to recognise specific image features indicative of high or low quality.

2.5 Comparison of Machine Learning Algorithms for Image Quality Assessment

Algorithm	Strength	Limitation	References
SVM	-High performance in high-dimensional data -Effective in small sample size -Robust classification boundaries	- Sensitive to kernel and parameter tuning - May struggle with class imbalance without weighting	(Ho et al., 2022)
KNN	- Simple and easy to implement - Non-parametric and fast training - Effective in low-resource settings	- Performance sensitive to the value of k and feature scaling - Struggles with high-dimensional or imbalanced data - Computationally costly during prediction	(Zhou et al., 2021; Singh et al., 2022)
RF	- Handles large feature sets and noise - Provides feature importance	- Biased toward majority class in imbalanced datasets - Slower training due to	(Chen et al., 2021)

	- Reduces overfitting via ensemble learning	multiple trees - Less interpretable	
--	--	--	--

2.6 Image Preprocessing to Improve Image Quality

Preprocessing is a crucial step in enhancing image quality before ML-based evaluation. It involves techniques aimed at improving image clarity, reducing noise, and highlighting relevant features. It involves the use of mathematical or statistical operations to modify images for many applications, including and not limited to medical and satellite imagery and digital photography (Anuradha Mahato, 2023). The main objective of preprocessing is to normalise and remove variation for easier classification and recognize correctly. Image preprocessing plays a crucial role in enhancing image quality before applying machine learning algorithms for automated evaluation. In the context of ACR RMI 156 phantom images, preprocessing aims to improve the visibility of embedded objects and remove unwanted artefacts or noise that may interfere with analysis.

2.5.1 Noise Reduction

Noise in medical images can obscure diagnostic details and reduce overall image quality. Common noise reduction techniques include median filtering, Gaussian smoothing, and bilateral filtering. These methods help eliminate random variations while preserving edges and structural details. This is essential in low-dose DBT images where noise levels can be significant.

2.5.2 Contrast Enhancement

Enhancing image contrast is essential for improving the visibility of structures with subtle intensity differences. Histogram equalisation, contrast stretching, and adaptive contrast enhancement are widely used methods to achieve this. Enhanced contrast improves feature detection and facilitates accurate evaluation. By improving the

contrast between the image's background, fibrous tissue, dense tissue, and sick tissue, which includes microcalcifications and masses, the mammography histogram is modified using these procedures (Alshamrani et al., 2022).

2.5.3 Region-of-Interest (ROI) Segmentation & Feature Extraction

ROI segmentation focuses the analysis on diagnostically relevant areas, such as fibers, specks, and masses in ACR phantom images. Segmentation techniques include thresholding, edge detection, and region growing. Accurate ROI extraction ensures that the ML model concentrates on pertinent features, improving evaluation precision.

Oh et al. (2022) developed a deep learning model that incorporates preprocessing steps to detect and segment phantom regions. The model outlined ROIs using bounding boxes and subsequently evaluated the visibility of test objects, demonstrating the importance of preprocessing in automated QA workflows.

Feature extraction techniques are used to identify and extract relevant features from an image. These features can be used in object recognition and image classification. The standard techniques used are edge detection, corner detection, and texture analysis (Mahato, 2023). Edge detection technique used to enhance boundaries of phantom objects, aiding in segmentation and detection tasks.

2.7 Image Quality Metrics and Assessment Methods

Quantitative metrics provide standardised criteria for evaluating image quality. These metrics are essential for both manual and automated assessments. To evaluate the effectiveness of ML algorithms in image quality assessment, several statistical performance metrics are commonly used. These metrics provide a comprehensive view of how well the model identifies image quality attributes such as detectability, sharpness, and lesion visibility.

2.6.1 Spatial Resolution

Spatial resolution refers to the ability of an imaging system to distinguish small objects that are close together. It is typically assessed using line pair phantoms or edge spread functions. In ACR phantom images, spatial resolution is evaluated by observing the clarity of fiber structures.

2.6.2 Signal-to-Noise Ratio (SNR)

SNR quantifies the proportion of meaningful signal relative to background noise. High SNR values indicate clearer images, which are crucial for detecting low-contrast lesions. SNR can be calculated using mean pixel values and standard deviation within defined ROIs.

$$SNR = \frac{MPV - offset}{sd} \quad (1)$$

Equation 1 Calculation of signal noise ratio (SNR) (Leyton et al., 2011).

Where MPV is the mean value of pixel in the region of interest, sd is the value of the standard deviation of the region of interest.

2.6.3 Contrast-to-Noise Ratio (CNR)

CNR measures the contrast between two regions relative to the noise level. It is particularly useful in evaluating the detectability of subtle lesions. High CNR values correlate with better diagnostic performance.

$$CNR = \frac{MPV_1 - MPV_2}{\sqrt{\frac{sd_1^2 + sd_2^2}{2}}} \quad (2)$$

Equation 2 Calculation of contrast noise ratio (CNR) (Leyton et al., 2011)

Where MPV is the mean value of the pixel in the region of interest 1 or 2, sd is the value of the standard deviation of the region of interest 1 or 2.

2.6.4 Detectability

Detectability refers to the visibility of test objects such as fibers, specks, and masses in phantom images. It is often scored using a visual grading scale by experts or calculated using automated feature detection algorithms.

Kim et al. (2024) employed phantom studies to investigate lesion detectability across different DBT configurations. They calculated SNR and CNR for various phantom patterns and found strong correlations between these metrics and diagnostic confidence.

2.6.5 Accuracy

Accuracy measures the overall correctness of the model's predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

Equation 3 Calculation of Accuracy (James, 2024).

Where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives. It reflects the proportion of correctly classified instances over the total number of instances.

2.6.6 Loss

Loss quantifies the error made by the model in its predictions. Common loss functions include Cross-Entropy Loss for classification tasks and Mean Squared Error (MSE) for regression. A lower loss value indicates better model performance.

2.6.7 Precision

Precision assesses the proportion of true positive results among all positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Equation 4 Calculation of Precision (James, 2024).

Where TP is True Positives, FP is False Positives. High precision indicates that the model has a low false-positive rate.

2.6.8 Recall (Sensitivity)

Recall, or sensitivity, measures the proportion of actual positives that were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

Equation 5 Calculation of Recall (James, 2024).

Where TP is True Positives, FN is False Negatives. A high recall means the model effectively identifies all relevant instances.

2.6.9 Specificity

Specificity is the proportion of actual negatives correctly identified by the model:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

Equation 6 Calculation of Specificity (Ebrary, 2020).

Where TN is True Negatives, FP is False Positives. It complements sensitivity by showing how well the model avoids false positives.

2.6.10 F1 Score

The F1 Score is the harmonic mean of precision and recall, providing a balance between them:

$$\text{F1-Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

Equation 7 Calculation of F1-Score (James, 2024).

It is particularly useful when the dataset is imbalanced, as it accounts for both false positives and false negatives.

These metrics are critical for comparing manual and ML-based QA methods. For instance, a high F1 score indicates that the automated system is reliable and suitable for deployment in clinical or QA settings. Evaluation results are typically presented in tabular or graphical form to illustrate performance across different DBT systems or phantom image sets.

2.8 ROC curve and AUC analysis

Receiver Operating Characteristic (ROC) curve analysis is a fundamental tool used to evaluate the diagnostic performance of machine learning models, particularly in binary classification tasks related to medical image assessment (Nahm, 2022). It provides insight into the trade-off between sensitivity (true positive rate or TPR) and 1-specificity (false positive rate or FPR) across different threshold values (Wang et al., 2022).

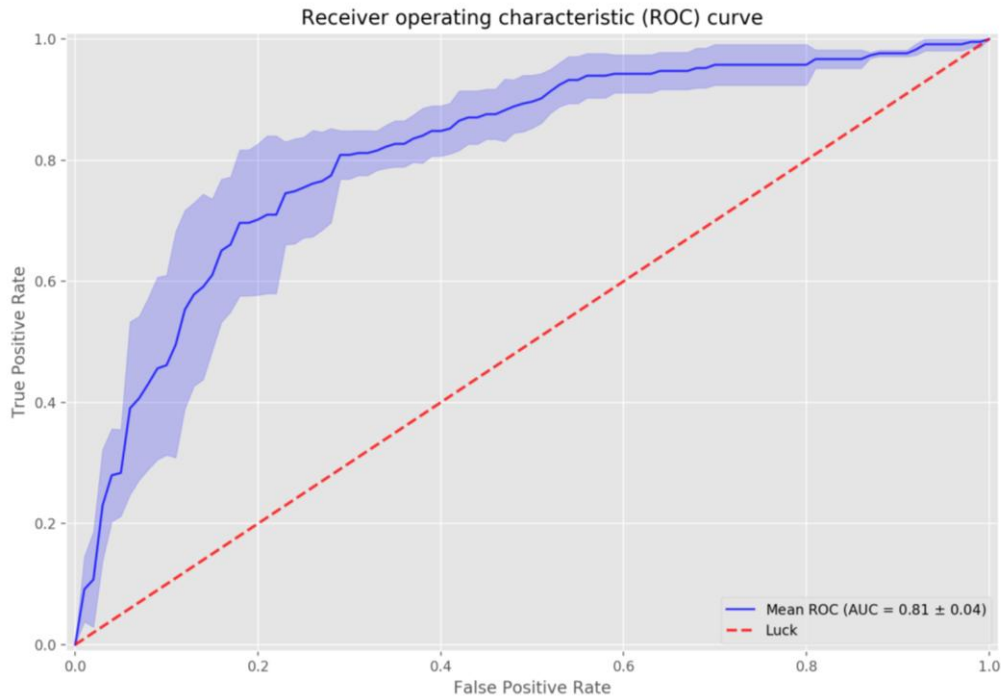


Figure 1 ROC Curve (Pierian Training, 2022).

The ROC curve plots the TPR against the FPR at various threshold settings. This takes place in the ROC space that is defined by TPR and FPR as x and y axes and shows the trade-off between sensitivity (TPR) and specificity ($1 - \text{FPR}$). A diagonal divides the ROC space and results above the diagonal represent good results while results under the diagonal represent bad results (Pierian Training, 2022).

Area Under the Curve (AUC) is a fundamental performance metric primarily used in machine learning for evaluating binary classification models. It quantifies a model's ability to distinguish between positive and negative classes. AUC of 1.0 represents a better model performance, while an AUC of 0.5 indicates no discriminative ability (Ultralytics, 2025).

AUC values can be interpreted as follows (Dash, 2023)

- **0.90 – 1.00:** High
- **0.70 – 0.90:** Moderate

- **0.00 – 0.70:** Low

2.9 Comparison of Manual and Automated Evaluation Methods

Manual evaluation of ACR phantom images has long been the standard in QA processes. Manual evaluation, performed by radiographers or medical physicists, is the traditional method for assessing phantom images. However, it is inherently subjective and time-consuming. Differences in training, experience, and fatigue levels among evaluators can lead to variability in scoring.

Automated ML-based methods address these limitations by providing objective, consistent, and scalable evaluations. Hejduk et al. (2023) highlighted that deep learning models could perform quality assessment of DBT images with greater consistency and reduced turnaround time compared to human observers. They found significant reduction in inter-observer variability when AI-assisted tools were implemented.

Kim et al. (2020) conducted a comparative study between radiographers and ML models in interpreting mammographic images. The ML models achieved comparable sensitivity and specificity, supporting the feasibility of ML in diagnostic and QA applications. Automated methods also enable real-time image quality feedback, facilitating immediate corrective actions and reducing the need for repeat imaging. This capability is particularly valuable in high-volume clinical settings where efficiency is paramount.

While automated systems offer numerous advantages, their implementation requires careful validation. Comparative studies must ensure that ML models perform equivalently or better than human experts under diverse imaging conditions. Regulatory compliance and interoperability with existing imaging systems are also critical considerations.

CHAPTER 3

METHODOLOGY

3.1 Study Design

This is a quantitative, experimental study designed to assess and analyze the image quality of Digital Breast Tomosynthesis (DBT) images acquired using the ACR (RMI 156) mammographic phantom. The images were obtained from HPUSM PACS system on March 2025 based on the inclusion and exclusion criteria. The study workflow consists of phantom imaging, preprocessing, augmentation, segmentation, feature extraction, and machine learning classification.

3.2 Study Location

The study is conducted at the Radiology Department, Hospital Pakar Universiti Sains Malaysia, for image acquisition using DBT. Image processing, feature analysis, and machine learning modeling are conducted in the Institute for Artificial Intelligence and Big Data (AIBIG), Universiti Malaysia Kelantan, using MATLAB with the Image Processing Toolbox and Statistics and Machine Learning Toolbox.

3.3 Selection Criteria

3.3.1 Inclusion

- i. DBT images acquired using validated ACR (RMI 156) phantoms.
- ii. Images containing quality metrics such as fibers, specks, and masses.
- iii. Images generated using DBT systems with standardised acquisition settings.

3.3.2 Exclusion

- i. DBT images with incomplete or missing quality assurance metrics.
- ii. Phantom images obtained from non-ACR compliant systems.

- iii. Low-resolution images unsuitable for machine learning processing.

3.4 Data Collection

The data collection for this study involved primary data acquisition using the ACR (RMI 156) phantom, which was imaged twice using a Digital Breast Tomosynthesis (DBT) system on 24 February 2025 and 12 March 2025 at the Radiology Department, Hospital Pakar Universiti Sains Malaysia. A total of six original DBT slices (three from each of two exposures on 24 February and 12 March 2025) were selected for analysis based on visibility of QA features (fibers, specks, and masses). These six images were then used for data augmentation to expand the dataset to 180 images. The images were exported in DICOM format. These images were imported into MATLAB for preprocessing, segmentation, and feature extraction. Data augmentation was applied to increase the dataset size for machine learning analysis. Extracted features along with imaging parameters were used to train and evaluate classifiers including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). The performance of these models was assessed using metrics such as accuracy, precision, recall, F1 score, specificity, sensitivity, and area under the curve (AUC) to determine their effectiveness in image quality assessment.

3.5 Study Instruments

3.5.1 ACR (RMI 156) Phantom

The Mammographic Accreditation Phantom RMI 156 is designed to test the performance of a mammographic system by a quantitative evaluation of the system's ability to image small structures similar to those found clinically. Objects within the phantom simulate micro-calcifications, fibrous structures in ducts, and tumor-like masses. The Mammographic Accreditation Phantom can determine if your mammographic

system can detect small structures that are important in the early detection of breast cancer. This phantom simulates a compressed human breast of average density, typically 4.2 cm thick, and contains embedded test objects that mimic anatomical structures found in breast tissue. These test objects include six nylon fibers of varying diameters to represent fibrous tissue, five groups of calcium carbonate specks to simulate microcalcifications, and five simulated masses to represent soft tissue lesions (Sunnuclear.com, 2022).

The phantom is widely used for quality control, performance verification, and accreditation of imaging systems, ensuring consistency and diagnostic reliability across clinical sites. In DBT, the phantom is scanned using standard clinical parameters, allowing radiographers and medical physicists to assess the detectability of critical structures across tomographic slices.

3.5.2 Siemens Digital Breast Tomosynthesis (DBT) System

The Siemens MAMMOMAT Revelation is an advanced digital mammography system engineered to enhance breast cancer detection, improve patient comfort, and streamline clinical workflows. It's supporting the Tomosynthesis mode. A 3D information can be obtained by taking exposures from different angles. This significantly reduces the tissue overlap problem which limits diagnosis in conventional mammography. During tomosynthesis acquisition, the MAMMOMAT Revelation can covers an angular range from $+25^{\circ}$ to -25° while 25 views are acquired (Siemens-healthineers.com, 2023). The system also supports Insight 2D and Insight 3D synthetic imaging options, which provide comprehensive visualization without additional radiation exposure (Cassling, 2024).

To improve patient comfort, the system features Personalized Soft Compression, which adapts compression force based on individual anatomy, thereby reducing discomfort without compromising image quality. The OpComp and SoftSpeed were used to prevent unnecessary pain by adjusting to the anatomy of each woman's breast. OpComp calculates and applies optimal compression while SoftSpeed slows the compression paddle down as soon as it reaches the breast (Siemens-healthineers.com, 2024). The MAMMOMAT Revelation is also compatible with AI tools such as Transpara® and iCAD, which provide decision support and enhance diagnostic accuracy (Cassling, 2024).

From a technical perspective, the system is built around an amorphous selenium (aSe) detector with a 24 cm × 30 cm field of view, a tungsten anode X-ray tube, and motorized isocentric rotation that enables 180° tomosynthesis acquisition. The system is designed with scalability in mind, ensuring adaptability to future clinical needs and technological developments (Siemens Healthineers, 2023).

3.5.3 Pictures Archiving and Communication System (PACS)

PACS is a medical imaging technology used primarily in healthcare organizations to securely store and digitally transmit electronic images and clinically-relevant reports. It serves as a replacement for traditional film-based workflows by using the Digital Imaging and Communications in Medicine (DICOM) standard to acquire, transmit, and store imaging data from various modalities, including X-ray, CT, MRI, ultrasound, and digital mammography systems such as the Siemens MAMMOMAT Revelation. PACS allows radiographers and clinicians to view and analyze images through dedicated workstations or web-based interfaces, improving diagnostic efficiency, remote access, and interdepartmental collaboration. In breast imaging, particularly in Digital Breast

Tomosynthesis (DBT), PACS is essential for handling the substantial volume of image slices generated during each scan.

PACS supports machine learning workflows by offering structured access to annotated image datasets, which can be extracted using DICOM interfaces and integrated into image analysis pipelines using software tools like MATLAB or Python. PACS ensures that all imaging data are securely stored, encrypted, and accessible only to authorized personnel, thereby complying with healthcare privacy regulations such as HIPAA or local data protection acts (Petrova, 2025).

A Cloud-based Picture Archiving and Communications System (Cloud PACS) combines the benefits of traditional PACS with the power of cloud infrastructure management. Native cloud PACS includes services and resources built in the cloud, which offer faster development of software, elasticity, data redundancy, plus data & resources reliability (Intelrad, 2022). Cloud-based PACS hosting offers a transformative solution by enhancing medical imaging security, ensuring HIPAA compliance, and improving patient care (Petrova, 2025).

3.5.4 MATLAB Software

MATLAB is the proprietary software app and programming language by MathWorks, which facilitates complex data analysis tasks such as algorithm implementation, interacting with other apps and manipulating a data matrix (BasuMallick, 2022). MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages. MATLAB combines a desktop environment tuned for iterative analysis and design processes with a programming language that expresses