



Detection of Tumors and Stages of *Mammae Carcinoma* on Mammography Imaging Using *Morphological Operations*

Nurul Auliyaa Hasbi^{1*}, Sultan Hamjar², Indah Musdalifah²

¹ Radiology Imaging Technology, Politeknik Muhammadiyah, Makassar, Indonesia.

² Radiology, Politeknik Muhammadiyah, Makassar, Indonesia.

Received: March 10, 2026

Revised: June 09, 2026

Accepted: July 25, 2026

Published: July 31, 2026

Corresponding Author:

Nurul Auliyaa Hasbi

nurulauliyaa@poltekkesmu.ac.id

DOI: [10.29303/jppipa.v12i7.15622](https://doi.org/10.29303/jppipa.v12i7.15622)

Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



Abstract: Carcinoma of the mammae is disease caused the changing breast cells and forming malignant tumors. Mammography plays an important role as an initial screening method, while histopathological results remain the gold standard for determining the stage of mammary carcinoma. And in the previous method, thresholding limitations did not separate noise, besides that Region Limitation Contrast technique was not good and Active Contour time was relatively long, therefore, Machine learning K-means clustering and Morphological operations became an accurate diagnosis solution through image algorithms, Machine learning have the ability to represent features that have similarities in expert results with Radiologists and even histopathology Results. Able to classify radiography results automatically detect the stage ca mammae and have the same results histopathology. Quasi experimental research on Post-test Only Control Group Design. Build machine learning Matlab program. The test measures accuracy, sensitivity, specificity, positive prediction value and NPV. Data analysis validity test and Wilcoxon statistical test. The study proved that 164 samples obtained good machine learning performance in detecting the stage mammary carcinoma with an accuracy value 97.35%, sensitivity 85.74%, specificity 96.92%, positive prediction value 88.04%, NPV 96.91%, there was similarity between Machine learning and histopathology results.

Keywords: Automatic detection; K-means clustering; Mammogram; Morphological surgery; Stage of mammary carcinoma

Introduction

Carcinoma mammary or breast cancer is a group of diseases that cause cells in the breast to change and grow uncontrollably. Many cancer cells eventually grow into abnormal masses known as malignant tumors (Bray et al., 2024). It is characterized by the formation of malignant cells within breast tissue that can invade nearby areas. Over the past decade, breast cancer has remained one of the major causes of mortality among women. If cancer can be detected early, treatment options and the chances of a complete cure will increase (Al-Antari et al., 2020; Al-Ayyoub et al., 2017).

The amount of prevalence shown from Carcinoma mammals that occurred is according to the Data and Information Center of the Ministry of Health of the

Republic of Indonesia, in 2020 there were 2,261,419 newly reported cases of breast cancer worldwide, resulting in 684,996 deaths. In Indonesia alone, 65,858 new cases were recorded, with 22,430 related deaths (Bustan et al., 2018; Michael et al., 2021). Breast cancer with malignant tumors according to the American College of Radiology (ACR) based on BI-RADS Descriptor, can be characterized by malignant tumor characteristics such as irregular shapes, lobulated, architectural distortion, and microlobulated, spiculated, obscured, ill-defined, and high density edges (Halls, 2019). The classification of breast cancer diagnosis can be divided into various stages from stages 0 to IV. Determining the stage of breast cancer is very necessary for treatment planning so that it can help improve the patient's survival rate (Bustan et al., 2018). According to

How to Cite:

Hasbi, N. A., Hamjar, S., & Musdalifah, I. (2026). Detection of Tumors and Stages of Mammae Carcinoma on Mammography Imaging Using Morphological Operations. *Jurnal Penelitian Pendidikan IPA*, 12(7), 813–822. <https://doi.org/10.29303/jppipa.v12i7.15622>

the American Joint Committee on Cancer (AJCC), the stage of breast cancer is determined based on tumor size, nodes, and metastases (Egner, 2010). So in the detection of breast cancer, various imaging techniques are used for early diagnosis of cancer, especially the commonly used imaging modality is mammography (AlKandari et al., 2024; Mohammed et al., 2024).

As a proven and effective imaging modality, Mammography is a special radiological screening and examination process using low-dose X-rays used to identify breast abnormalities such as cancer. Mammography provides high sensitivity to fatty breasts and excellent microcalcification demonstrations, it strongly indicates early detection of cancer (Gbenga et al., 2017; Milosevic et al., 2017; Shi et al., 2018). The growing complexity in analyzing mammogram results requires specialized radiology expertise, which has encouraged the advancement of automated detection systems using Machine Learning approaches. Therefore, Machine Learning models that are simple, accurate, and fast are expected to help address these challenges and provide timely support for patients (Caobelli, 2020). More recently, scientists increased the application of Machine Learning tools, namely, image segmentation, stratification and estimation for BC assessment. Image segmentation both identifies and extracts tumour and its adjoining normal tissue. Machine Learning tools, particularly, convolutional neural networks, are widely employed for automated BC segmentation (Fan et al., 2021; Guo et al., 2024).

In medical image processing, segmentation is an essential step that isolates the Region of Interest (ROI) and its boundaries by dividing a digital image into multiple smaller regions (Vaka et al., 2020). There is a lack of one of the studies is that Thresholding has limitations that do not sufficiently separate noise, besides that there is a Region technique which has limitations by producing poor contrast and Active Contour techniques with relatively long time limitations (Bhalodiya et al., 2022). Meanwhile, studies comparing Fuzzy C-mean, Fuzzy SVM Algorithm, and K-mean segmentation with the results of the three segmentations are almost relatively similar in the detection of abnormal areas of breast cancer, but the operation of K-means segmentation is relatively much faster (Kamil et al., 2019). In previous research, the K-means clustering algorithm technique was able to detect automatic borders, assist doctors in analyzing lesions in a short time and prevent biopsies and contour shapes or borders can describe malignant and benign lesions (Ibrahim et al., 2021; Salloum, 2025).

From considerations based on previous segmentation research, K-means clustering is a better technique, where K-means clustering is an algorithm to group objects into a number of K clustering based on

features. K-means clustering has many advantages, namely in addition to partitioning abnormal areas well, it is also able to overcome the problem of unclear boundaries and gray scale levels, as well as relatively faster operation than other techniques, execution times 2 times faster than serial computing using CPUs and detects multiple overlapping networks and significant homogeneous areas. It's just that K-means clustering is still lacking in improving image quality, so there needs to be additional techniques, so in the literature study, it is stated that the weakness of the segmentation technique is that it is not able to improve high quality so it needs a technique that is able to remove segmentation imperfections, so there is a Morphological operations technique that has various features in producing quality images (He et al., 2017; Salloum, 2025).

Morphological operations are image processing methods that focus on the shapes of objects within an image. They are commonly used to correct imperfections produced during segmentation, helping enhance image quality and information while reducing noise, increase density, sharpness and detail of radiography results. by showing an appropriate anatomical picture and can provide a high diagnostic accuracy value so that organs and pathologies can be clearly seen. Mathematical morphological operations are used to structure elements and analyze the shape within an image. They also enhance image features and assist in eliminating undesired regions. Morphological closure operations are denoted by closing and opening. This surgery can be adapted for the application with the selection of the right structuring elements so that morphological surgery is performed to get a clear view of the tumor (Nelda. A et al., 2024; Salih et al., 2018).

Seeing the ever-increasing number of breast cancer incidences, however, manual work for this diagnosis requires precision, in addition to the weaknesses of manual methods requires skill accurately by selecting abnormal areas, which will take time. As well as in the structure of the human breast, which is very complex, it also presents its own difficulties in the identification of cancer, with gray levels, textures, colors, motions and uneven distribution of signals making it very difficult to detect. In addition, mammogram carcinoma images are often covered by dense tissue (stroma, epithelium, muscle and other tissues) so that they are not sensitive to less pronounced density differences between mammalian carcinoma and normal tissue (Baccouche et al., 2021). Therefore, this underlies the difficulty of radiologists in determining the size of the tumor and determining the stage of mammary carcinoma. The accuracy in early stage diagnosis is shifted to the examination of histopathology.

Therefore, there is a need to support the improvement of automatic detection software

development methods that will reduce subjectivity problems between the results of expertise between radiologists and each other, as a complement to the mammography modality in hospitals, in this study it provides a solution for a digital image processing software to measure tumors and detect early stages of Carcinoma mammals using a segmentation process, in this study the author will examine the results of the accuracy between the machine learning algorithm method compared to the results of the histopathology examination in diagnosing the early stages Breast Carcinoma. It is hoped that the results of this study using the machine learning algorithm method built can provide support for a diagnosis prediction similar to the results of histopathology in order to support the diagnosis of mammary carcinoma to be more accurate, fast, precise and time efficient.

Based on the explanation above, the proposed approach applies a machine learning algorithm for automatic early detection of mammary carcinoma using K-means clustering segmentation combined with morphological operations. The output produced is an early-stage classification. By integrating K-means clustering and morphological segmentation, the system is expected to automatically identify early signs of mammary carcinoma in digital mammogram images with reliable and optimal performance.

This combination is needed to balance processing speed (K-means) and shape accuracy (morphology), in order to overcome the challenge of uneven signal distribution in dense mammogram images. The segmentation results of the constructed system are used as a comparison to ensure that the prediction of the system's diagnosis has a strong and reliable correlation like the results of histopathological examination. This study explicitly validated machine learning results against the gold standard of histopathology to ensure the clinical relevance of the diagnosis results. The system not only detects tumors, but also contributes to the classification of early stages of carcinoma, which is crucial in determining the patient's treatment plan, thus providing an objective software solution to reduce subjectivity between radiologists, while speeding up the manual diagnosis process that is time-consuming and requires high precision due to the complexity of the human breast structure. This development is aimed at complementing mammography modalities in hospitals as a decision support system that makes the diagnosis process faster, more precise, and more time-saving in the clinical environment.

Method

This study employs an experimental method using a Post-test Only Control Group Design (Sugiyono, 2019).

This study will be analyzed by comparing assessments using digital image processing software in to detect early-stage mammary carcinoma and obtain sensitivity, specificity, positive predictive value, negative predictive value, and accuracy metrics. The source of the research data used secondary data from dicom images and medical records that had clinical mammary carcinoma (Hasbi et al., 2022). The subject of this study was mammogram images with clinical Susp. mammae carcinoma by comparing the results of the combination of K-means clustering and Morphological operations segmentation software with the results of histopathology. The target population in the study was mammogram images with Susp. carcinoma mammals at the Radiology Installation of Dr. Wahidin Sudirohusodo Hospital Makassar, while the affordable population was the target population with Craniocaudal (CC) projection mammogram images from November 2021-March 2022 with Susp. carcinoma mammae cases and had histopathology results at Dr. Wahidin Sudirohusodo Hospital Makassar. The sample selection in this study is included in the group of non-probability sampling samples, using the purposive sampling technique where the researcher selects the research sample based on subjective and practical considerations in accordance with the desired research objectives (Hasbi et al., 2022). A total of 164 digital mammogram images will be used as samples in this study. Data analysis will be performed using cross-validation and evaluated through a confusion matrix to obtain sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. For statistical analysis, the Wilcoxon test will be applied to determine whether a significant difference exists.

This study uses DICOM-format mammogram images that go through a pre-processing stage in the form of downscaling for computational efficiency, light intensity normalization to standardize contrast, and spatial filtering to reduce noise. The segmentation process is performed using an unsupervised K-means algorithm to distinguish the background, normal tissue, and lesion area, which is initialized through iterative optimization. The results of segmentation are then refined using morphological operations with disk-shaped arrangement elements; The closing process is applied to close the internal gap of the tumor, while the opening process is used to eliminate external artifacts. Model performance evaluation is carried out using 5-Fold Cross Validation to ensure stable data generalization and minimize the risk of overfitting. The system pipeline then extracts the tumor diameter features from the final segmentation results, where the system will classify the stage of Mammaan Carcinoma if the diameter meets the clinical threshold, or assign the outcome as a normal condition if no lesions are detected.

Here's a Flowchart of the K-Means Clustering and Morphological Operations process using Matlab:

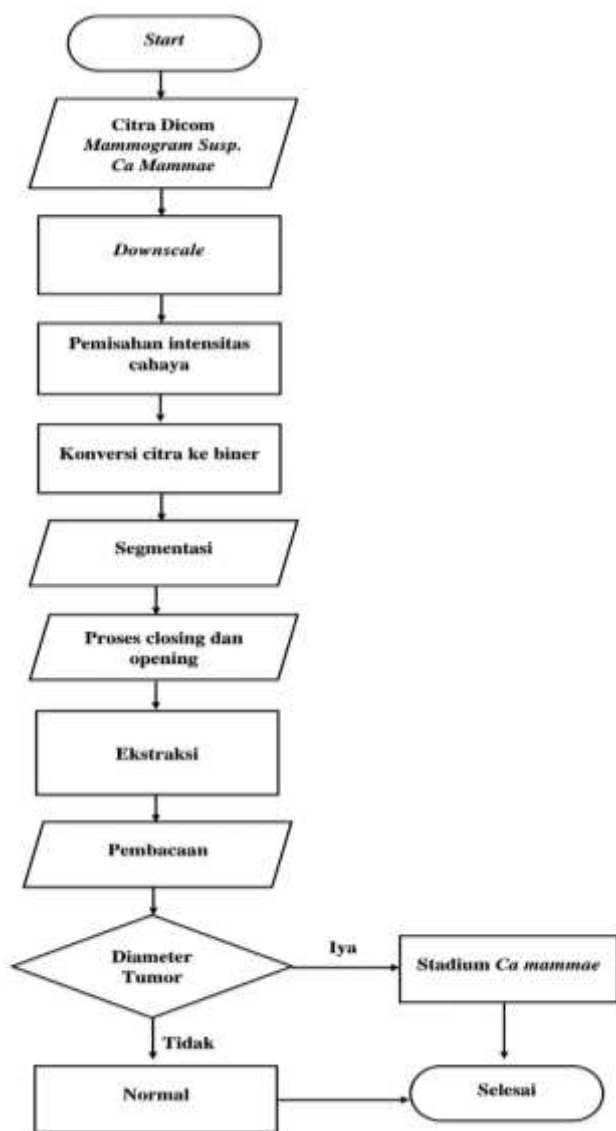


Figure 1. Flowchart of the k-means clustering and morphological operations

Result and Discussion

Result

This research was carried out on mammogram images using a Machine Learning model based on K-means clustering and morphological operations implemented in the MATLAB program. The image samples used amounted to 164 mammogram images consisting of 155 mammogram data from mammogram carcinoma and 9 normal mammogram data. Mammogram images are dominated in the age range of 25-75 years. Based on this, it refers to the results of a study based on the prevalence of a large study of Global Cancer Statistics which states that the average age of mammary carcinoma cases is 50 years (interquartile

range of 35-58 years), 87% of cases are 30-75 years old and 13% are ≥ 75 years old, and the number of female patients is 98% (Bray et al., 2024).

Model Machine Learning K-Means Clustering dan Morphological Operations

The image sample was obtained retrospectively from CR at the location of the Hospital, namely at Dr. Wahidin Sudirohusodo Hospital Makassar. The data collection process began in the period of January 2022-April 2022. A total of 818 digital mammogram images were obtained which were divided into 654 training data sets consisting of 618 mammograms of mammary carcinoma and 36 normal mammograms, while for the confirmed test dataset of 155 mammogram data of mammary carcinoma and 9 normal mammograms, so that the total sample used was 164 samples. A confirmed sample of mammary carcinoma is based on the results of the histopathology Laboratory, then the normal sample is validated by an anatomical pathologist. The K-means clustering and Morphological Operations machine learning model is built using unsupervised machine learning algorithms.

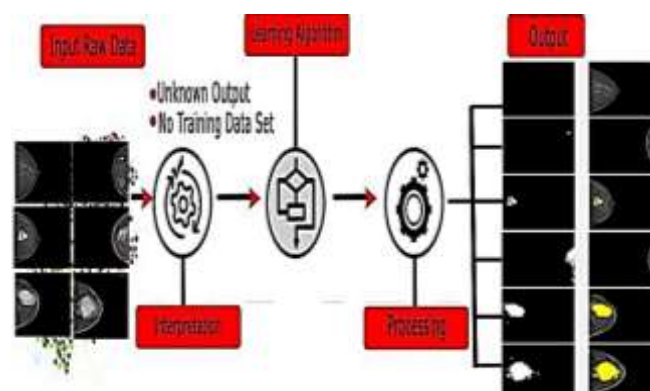


Figure 2. Model machine learning K-means clustering dan morphological operations

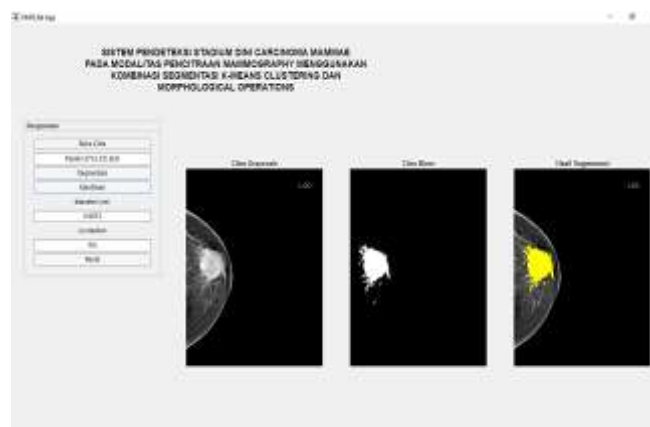


Figure 3. Results of machine learning K-means clustering and morphological operations detection

Performance Validation Analysis of K-means Clustering and Morphological Operations Machine Learning Models

Machine learning models are applied to improve the accuracy of image classification and from various machine learning studies no one has ever achieved 100% accuracy (Mawaddah, 2021). The existence of an effective performance measurement tool for machine learning, namely with K Fold Cross Validation in the confusion matrix, based on diagnostic tests from dichotomous classifications and compared to gold standards.

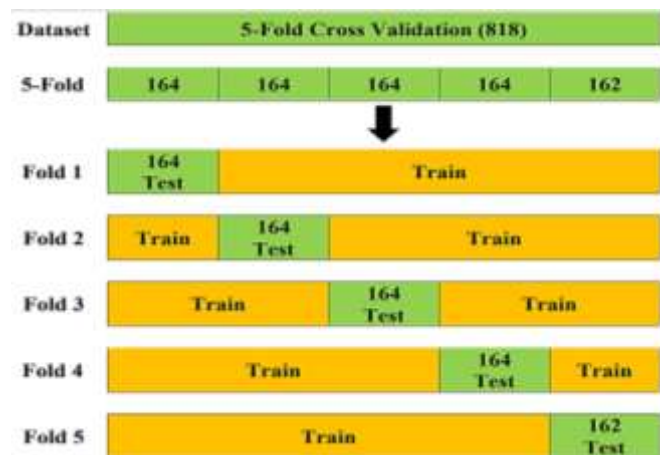


Figure 4. Folding cross validation

Testing is a range of values that determine the performance ability of a machine learning model to be good or fail, the higher the percentage of results to the values of accuracy, sensitivity, specificity, positive prediction value and NPV, the better the machine learning model. In assessing Machine Learning Performance. Where performance with a score of 0.90-1.00 is classified as very good, a score of 0.80-0.90 is classified as good, a score of 0.70-0.80 is classified as fair or equal, a score of 0.60-0.70 is classified as Low, and a score of 0.50-0.60 can be classified as Failed (Arry et al., 2018; Lathifah et al., 2025).

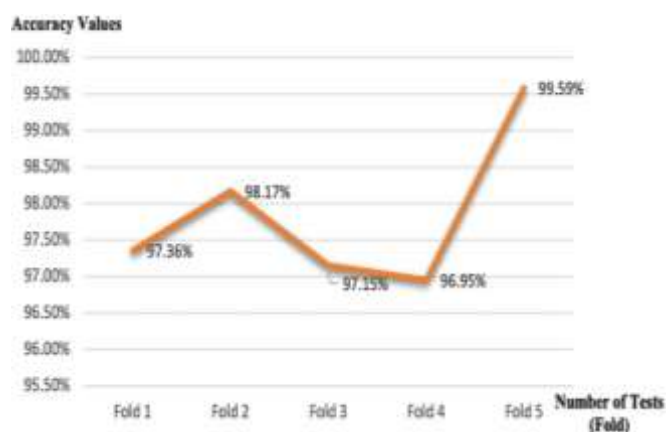


Figure 5. Test data accuracy

Table 1. Performance of K-Means & Morphological operation algorithm

Category	Value (%)	Indicator
Accuracy	97.35	Excellent
Sensitivity	85.74	Good
Specificity	96.92	Excellent
NPP	88.04	Good
NPN	96.91	Excellent

The overall performance of the model showed excellent performance with an accuracy of 97.35% (Table 1). The model has a very high ability to identify normal images, but there is room for improvement in the sensitivity of tumor detection.

Table 2. Performance Comparison with Previous Studies

Researcher (Year)	Method	Accuracy	Notes
Researcher (2019) (Kamil et al., 2019)	K-Means	63%	No Morphology
Researcher (2020) (Vaka et al., 2020)	HA-BiRNN	82.50%	No Morphology
This Research	K-Means + Morphology	97.35%	High efficiency

This table presents a comparison of the accuracy performance of the machine learning model proposed in this study against the methods used in previous studies for segmentation of mammogram images. This comparison aims to highlight the significant contribution of morphological operation integration to improved system accuracy compared to methods that rely solely on single clustering or deep learning methods without additional morphological image processing stages. The data show that the combination of K-Means and morphological surgery developed in this study results in a more optimal and efficient 97.35% detection performance for clinical diagnostic needs.

Classification Speed Analysis

The learning stage to validation, the machine learning model with an image size of 50 x 50 takes 10 to 15 minutes, the image size of 100 x 100 takes 15 to 30 minutes, and the image size of 200 x 200 is 60 to 90 minutes (Polat et al., 2019). Meanwhile, in this study, the K-means clustering and Morphological Operations machine learning process was validated to determine the accuracy, sensitivity, specificity, positive prediction value, NPV and error rate in two stages, namely learning (Training) and test (Tasting). At the learning (Training) and test (Tasting) stages it takes 37.676 seconds against 818 data.

At the learning (Training) and test (Tasting) stages it takes 37.676 seconds against 818 data. That at the learning stage (Training) it takes 30.140 seconds against 654 data. Meanwhile, the test stage (tasting) takes 7.535

seconds for 164 data. So the speed of classification of the K-means clustering and Morphological Operations machine learning process takes 46 seconds for 1 data, in the form of classification of Carcinoma mammary stages Ia, Ib, IIa, IIb, IIIa and normal.

Comparison of Machine Learning Performance of K-Means Clustering and Morphological Operations with Histopathology Results

The results of the histopathology laboratory which is the gold standard as a comparison with the performance results of the K-means clustering and Morphological Operations Machine Learning models to find out whether the machine learning model created can be a support for radiologists in hospitals.

The results of the study with the Wilcoxon comparative test showed that there was no significant difference between the performance results of the K-means clustering and Morphological Operations Machine Learning model with the results of histopathology, so the conclusion was drawn based on the hypothesis which means the alternative hypothesis was accepted that there was no difference in the reading results of mammogram images in detecting the early stage of mammary carcinoma between machine learning segmentation K-means clustering and Morphological operations with histopathology results. From the results of the Wilcoxon test, machine learning segmentation K-means clustering and Morphological operations with histopathology results with a p-value ($p > 0.05$) of 0.982.

Discussion

The results of the research were obtained by running the K-means clustering and Morphological Operations machine learning model through the Matlab program. Then a mammogram image processing page will appear. Running a machine learning model program starts by inputting a training data set that has been previously converted into a greyscale image so that it can be processed by Matlab. Furthermore, the training data set that has been carried out in the training process is then saved, followed by the testing process by entering the test data set in the prepared folder. After the test, the test classification results will appear.

Model Machine Learning K-Means Clustering and Morphological Operations

The performance assessment of the classification of the K-means clustering and Morphological Operations machine learning model was carried out after running the machine learning model program and calculated using K Fold Cross Validation and confusion matrix for further diagnostic tests with indicators of monitoring the performance of a machine learning model namely accuracy, sensitivity, specificity, positive prediction

value, and NPV (Auliya, 2022). In addition to the assessment of the performance of the machine learning model, a comparative assessment of K-means clustering and Morphological Operations machine learning was also carried out on histopathology results. histopathology results are the gold standard in the detection of the Mammary Carcinoma stage, so that to determine the acceptance of the machine learning model, the model performance is compared with the results of histopathology because it has the highest performance value. Therefore, to find out the proportion of K-means clustering and Morphological Operation machine learning models, it is necessary to do it comparatively with the results of histopathology.

The process of creating a machine learning model program uses parameters to control the flow of machine learning model functions in order to produce accurate classification predictions. The main purpose of unsupervised learning algorithms is to find previously unknown patterns in the data such as high-level features of local distinctive signs such as edges, motifs, lines, and other visual elements (Auliya, 2022). In this study, the introduction of irregular, lobulated, architectural distortion, and microlobulated, spiculated, obscured, ill-defined, and high density edges was also used. Meanwhile, the determination of the stage is based on Tumor size (T), based on Node (N) and Metastasis (M). The purpose of classification in the Mammary Stage Carcinoma detection study is used to detect the Mammary Stage Carcinoma, which means that if the stage detection is more similar to the histopathology (gold standard) results, then it is good.

Performance Validation Analysis of K-Means Clustering and Morphological Operations Machine Learning Models

The classification performance of the machine learning model using K-means clustering and morphological operations shows that accuracy, specificity, and negative predictive value all fall within the 0.90–1.00 range. This indicates that the model is highly capable of classifying mammary carcinoma stages and distinguishing normal digital mammography images, including correctly identifying and labeling normal cases. Therefore, the machine learning approach can be categorized as "Excellent" for detecting the stage of mammary carcinoma.

The sensitivity and positive prediction value are above the range of 0.80–0.90 which means that the proposed model's ability to detect mammogram carcinoma mammary and the proposed model's ability to detect Susp carcinoma mammary and classified as mammary carcinoma is said to be "good".

Based on the accuracy value obtained as much as 97.35%, it means that there are still errors of 2.65% caused by several obstacles previously described from

the application of the K-means clustering and Morphological Operations machine learning models in detecting mammary carcinoma, but supported by the basic results of the test can still be concluded that Machine learning is still categorized as very good used to detect the stage of mammary carcinoma.

The integration of K-means allows for the separation of contrasting background and breast tissue, while morphological operations act as filters that eliminate spatial noise. This makes tumor features (such as spiculation) more prominent than if only one method was used. The lower sensitivity suggests that the model is slightly more conservative in detecting tumors that are very small in size or that are camouflaged by high tissue density. The higher specificity indicates that the model is very reliable in identifying normal tissue (minimal false positive). Meanwhile, classification failure (2.65% error) mainly occurred due to truncation artifact (image of the tumor cut at the limit of the mammography field of view). This cuts off the connectivity of morphological features that the algorithm should recognize, so that the tumor shape becomes incomplete (not due to K-means failure in clustering).

Comparison of the performance between the model proposed in this study and several previous methods. This comparative analysis shows that the integration of machine learning and image processing techniques has a significant impact on detection accuracy. The Synergy of Methods is advantageous, which significantly surpasses the conventional K-Means method which only reaches 63%. This confirms that the use of K-Means alone without advanced stages tends to result in segmentation that is less than optimal and prone to noise. Effectiveness of Morphological Operations, compared to the HA-BiRNN method which achieved 82.50% accuracy without the support of morphological operations, the model proposed in this study proves that the addition of morphological operations after the clustering process is able to improve the quality of segmentation. The morphology stages effectively smooth out the contours and remove artifacts that are often the factors that cause the low accuracy of other methods. The achievement of 97.35% accuracy with high efficiency indicates that the combination of K-Means and morphological surgery is a more reliable solution for clinical practice. In addition to improving accuracy, this method maintains the crucial classification time efficiency to support rapid decision-making by radiologists in detecting the stage of mammary carcinoma.

Classification Speed Analysis

At the learning (Training) and test (Tasting) stages it takes 37,676 seconds against 818 data. So that at the learning stage (Training) it takes 30,140 seconds against

654 data. Meanwhile, the test stage (tasting) takes 7,535 seconds for 164 data. So the speed of classification of the K-means clustering and Morphological Operations machine learning process takes 46 seconds for 1 data, in the form of classification of Carcinoma mammary stages Ia, Ib, IIa, IIb, IIIa and normal.

The speed of classification in K-means clustering and Morphological Operations machine learning will be influenced by the size and number of images used. The higher the size and number the more it will take longer. But it's just that if the image size is getting smaller, it will reduce the number of pixels and decrease the resolution in the mammogram image which causes the image information on the digital mammogram, this is one of the reasons for this study to directly use the original image.

In contrast to Convolutional Neural Networks (CNN) models that often require high-performance Graphic Processing Units (GPUs) and much longer training times, these K-Means methods and morphological operations offer advantages in terms of lighter hardware requirements. With a classification speed of 46 seconds per data using standard computer specifications, the model is highly applicable for use in hospitals or healthcare centers with limited computing infrastructure, without sacrificing radiology workflows. The decision to use the original image to downscale is a deliberate form of trade-off. While downscaling can speed up processing times, it risks reducing pixel count and resolution, which can eliminate important texture and microcalcification information in digital mammogram images. Therefore, the use of the original image is maintained to ensure that diagnostic accuracy remains optimal, even with the consequence of longer processing time than if the image is scaled down.

Comparison of Machine Learning Performance of K-Means Clustering and Morphological Operations with Histopathology Results

The Wilcoxon Signed Rank test is used to compare the external system with the gold standard (histopathology). The results showed a p-value of 0.982 ($p > 0.05$), which indicated that no statistically significant difference was found between the distribution of system diagnosis results and histopathological diagnosis. Although statistically indifferent, this value suggests that the model has high potential as a decision support system that is aligned with the pathological diagnosis, but not as an absolute substitute for the role of the radiologist in the final verification.

Therefore, the alternative to this method to minimize the errors that occur is that the mammogram population is much longer than this study which is able to reach one year or more, and uses test data that meets the image criteria which structurally shows the overall

tumor picture is not cut and adds many clustering in the K-means clustering test, because there are more clustering more and more are also able to be identified by Machine Learning. So by paying attention to it, it is able to minimize the level of errors that occur.

The K-means clustering and Morphological operations machine learning model is a software to assist radiology specialists in determining diagnoses amid a lack of accuracy in detecting the stage of mammary carcinoma so that machine learning software is needed to ease the work, improve the diagnostic process and still get fast and accurate results. For this reason, the K-means clustering and Morphological operations machine learning model can be used on the condition that the verification of the diagnosis decision remains referenced based on the confirmation and justification of the radiologist specialist.

Conclusion

This study successfully proves that the integration of K-means clustering algorithm and morphological surgery is an effective and reliable method in detecting the stage of Mammogram Carcinoma in mammography images. The main contribution of this research lies in the development of segmentation software that is able to achieve a high level of accuracy with optimal computational time efficiency, making it very feasible to implement in healthcare facilities with limited computing infrastructure. Clinically, this system functions as an objective decision support system, helping to reduce subjectivity between examiners in the early diagnosis of breast cancer. However, this study has limitations in imaging cases with artifact truncation, where tumors that are cut at the boundary of the field of view can affect segmentation performance. Therefore, future research directions are suggested to expand the variety of datasets with more complete image quality and increase the complexity of the model by increasing the number of clusters or integrating with more adaptive deep learning methods. Although the results show strong statistical alignment with the gold standard of histopathology, the use of this system is still recommended as a support for clinical decisions that require final verification and justification by a radiologist to ensure diagnostic accuracy for patients.

Acknowledgments

Thank you to all those involved in this research in making important contributions to the implementation of the research until it is completed.

Author Contributions

S.H and I.M. analyzed and interpreted patient data regarding the disease. S.H and I.M. conceived the study, contributed to

sample preparation. All authors discussed the results and contributed to the final manuscript.

Funding

All sources of funding for the research report should be declared, for example: This research was partially supported by Kementerian Pendidikan Tinggi, Sains, dan Technology Republik Indonesia with grant number: 0488/C/DT.06.01/2025.

Conflicts of Interest

The authors state there is no conflict of interest with the parties involved in this study.

References

- Al-Antari, M. A., Al-Masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2020). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics*, 137, 104114. <https://doi.org/10.1016/j.ijmedinf.2020.104114>
- Al-Ayyoub, M., Alzu'Bi, S. M., Jararweh, Y., & Alsmirat, M. A. (2017). A GPU-based breast cancer detection system using Single Pass Fuzzy C-Means clustering algorithm. *International Conference on Multimedia Computing and Systems -Proceedings*, 0, 650–654. <https://doi.org/10.1109/ICMCS.2016.7905595>
- AlKandari, M., & Ahmad, I. (2024). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*, 20(3/4), 231–250. <https://doi.org/10.1016/j.aci.2019.11.002>
- Arry, & Widdo. (2018). Pemanfaatan Ciri Gray Level Co-Occurrence Matrix (GLCM). *The Scientific World Journal*, 5769–5776. Retrieved from <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3420>
- Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. S. (2021). Connected-UNets: a deep learning architecture for breast mass segmentation. *Npj Breast Cancer*, 7(1). <https://doi.org/10.1038/s41523-021-00358-x>
- Bhalodiya, J. M., Lim Choi Keung, S. N., & Arvanitis, T. N. (2022). Magnetic resonance image-based brain tumour segmentation methods: A systematic review. *Digital Health*, 8. <https://doi.org/10.1177/20552076221074122>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263.

- <https://doi.org/10.3322/caac.21834>
- Bustan, M. N., Tiro, M. A., & Adiatma. (2018). Modeling of Breast Cancer Diagnosis Classification Based on Hospital Medical Records. *Journal of Physics: Conference Series*, 1028(1), 1–7. <https://doi.org/10.1088/1742-6596/1028/1/012229>
- Caobelli, F. (2020). Artificial intelligence in medical imaging: Game over for radiologists? *European Journal of Radiology*, 126, 108940. <https://doi.org/10.1016/j.ejrad.2020.108940>
- Egner, J. R. (2010). AJCC Cancer Staging Manual. *Jama*, 304(15), 1726. <https://doi.org/10.1001/jama.2010.1525>
- Fan, J., Wang, J., Zhang, C., & Zhang, Y. (2021). Mammogram segmentation using a modified U-Net model. *IEEE Access*, 9, 99120–99130. <https://doi.org/10.1109/ACCESS.2021.3096535>
- Gbenga, D. E., Christopher, N., & Yetunde, D. C. (2017). Performance Comparison of Machine Learning Techniques for Breast Cancer Detection. *Nova Journal of Engineering and Applied Sciences*, 6(1), 1–8. Retrieved from www.novaexplore.com
- Guo, Y., Zhang, H., Yuan, L., Chen, W., Zhao, H., Yu, Q. Q., & Shi, W. (2024). Machine learning and new insights for breast cancer diagnosis. *Journal of International Medical Research*, 52(4), 1–29. <https://doi.org/10.1177/03000605241237867>
- Halls, S. B. (2019). Potentially malignant microcalcification by texture. *Journal of Applied Clinical Medical Physics*, 24. Retrieved from <https://breast-cancer.ca/malcalctex/>
- Hasbi, N. A., Adi, K., & Setiawan, A. N. (2022). Early Stadium Detection of Mammar Carcinoma on Mammography Imaging Modality Using K-Means Clustering and Morphological Operations Segmentation Combination. *Journal of Medical Imaging and Radiation Sciences*, 53(4), S52–S53. <https://doi.org/10.1016/j.jmir.2022.10.172>
- He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>
- Ibrahim, A., Elshennawy, N. M., & Sarhan, A. M. (2021). Deep learning approach for breast cancer identification in mammograms. *IEEE Access*, 9, 165507–165520. <https://doi.org/10.1109/ACCESS.2021.3134604>
- Kamil, M. Y., & Salih, A. M. (2019). Mammography images segmentation via Fuzzy C-mean and K-mean. *International Journal of Intelligent Engineering and Systems*, 12(1), 22–29. <https://doi.org/10.22266/IJIES2019.0228.03>
- Lathifah, A., & Fendriani, Y. (2025). Klasifikasi Penyakit Kanker Payudara pada Citra Mammogram Menggunakan CNN dan Random Forest. *Jurnal Online UNJA*, 10(3), 95–103. Retrieved from <https://www.kaggle.com/datasets/hayder17>
- Mawaddah, L. (2021). Classification of Benign Tumors and Malignant Tumors On Mammogram Imagery Using Gray Level Co Occurrence Matrix (GLCM) and Support Vector Machine (SVM). *The Scientific World Journal*. Retrieved from <https://www.researchgate.net/publication/356352467>
- Michael, E., Ma, H., Li, H., Kulwa, F., & Li, J. (2021). Breast Cancer Segmentation Methods: Current Status and Future Potentials. *BioMed Research International*, 2021, 1–29. <https://doi.org/10.1155/2021/9962109>
- Milosevic, M., Jovanovic, Z., & Jankovic, D. (2017). A comparison of methods for three-class mammograms classification. *Technology and Health Care*, 25(4), 657–670. <https://doi.org/10.3233/THC-160805>
- Mohammed, A. D., & Ekmekci, D. (2024). Breast Cancer Diagnosis Using YOLO-Based Multiscale Parallel CNN and Flattened Threshold Swish. *Applied Sciences*, 14(7), 2680. <https://doi.org/10.3390/app14072680>
- Nelda, A. A. R., Setiawan, R., Hendrawan, M. A., Nugroho, D. B., & Hakim, A. N. R. (2024). Deteksi Kanker Payudara Hasil Citra Mammografi menggunakan Metode Convolutional Neural Network (CNN) Arsitektur ResNet-50 Breast Cancer Detection from Mammography Images using Convolutional Neural Network (CNN) Method ResNet-50 Architecture. *Jiitu*, 01(01), 25–32. Retrieved from <https://share.google/sdq2wcj82x0YIht7L>
- Polat, H., & Mehr, H. D. (2019). Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture. *Applied Sciences (Switzerland)*, 9(5). <https://doi.org/10.3390/app9050940>
- Salih, A. M., & Kamil, M. Y. (2018). Mammography image segmentation based on fuzzy morphological operations. *Proceedings - 2018 1st Annual International Conference on Information and Sciences, AiCIS 2018*, 40–44. <https://doi.org/10.1109/AiCIS.2018.00020>
- Salloum, S. (2025). K-Means Clustering and Classification of Breast Cancer Images Using Histogram of Oriented Gradients Features and Convolutional Neural Network Models: Diagnostic Image Analysis Study. *JMIR Formative Research*, 9, e71974–e71974. <https://doi.org/10.2196/71974>
- Shi, P., Zhong, J., Rampun, A., & Wang, H. (2018). A hierarchical pipeline for breast boundary

- segmentation and calcification detection in mammograms. *Computers in Biology and Medicine*, 96, 178-188. <https://doi.org/10.1016/j.compbiomed.2018.03.011>
- Sugiyono, D. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan Tindakan*. Alfabeta.
- Vaka, A. R., Soni, B., & K., S. R. (2020). Breast cancer detection by leveraging Machine Learning. *ICT Express*, 6(4), 320-324. <https://doi.org/10.1016/j.icte.2020.04.009>