

Analysis of Convolutional Autoencoder and U-Net for Breast Tumor Detection in Mammogram Images

Talitha Nirmala Neva^{1✉}, Zaenal Alamsyah², Nugraha³
(1,2,3) Informatics Engineering, Nusa Putra University, Sukabumi, Indonesia

Abstract: Breast cancer is one of the most frequently diagnosed cancers among women worldwide. Early detection through mammogram analysis is important, but manual interpretation can be challenging because of low image contrast and variability in radiological assessment. This study evaluates the implementation of Convolutional Autoencoder (CAE) and U-Net architectures for mammogram image analysis and integrates both models into a Streamlit-based web application. The CBIS-DDSM dataset was used, comprising 723 training images and 181 testing images. The research stages included preprocessing, model training, evaluation, result analysis, and web application implementation. The U-Net model achieved a Dice Score of 0.4186, an Intersection over Union (IoU) of 0.2647, a precision of 0.5254, a recall of 0.3479, an F1-score of 0.4186, and an accuracy of 0.8955. Meanwhile, the CAE model employed a reconstruction-error threshold for anomaly detection, producing a mean anomaly score of 1.4779% and identifying 31 anomalous images among the 181 test images, using a reported threshold of 1.7578%. These findings indicate that U-Net is suitable for supervised lesion segmentation and localization, whereas CAE provides an annotation-independent approach to reconstruction-based anomaly analysis. However, further evaluation is required to establish the reliability of anomaly detection and the clinical applicability of both models.

Article history:

Received: 15 July 2026

Revised: 09 October 2026

Accepted: 13 October 2026

Published: 19 October 2026

Keyword:

mammogram, convolutional autoencoder, U-Net, image segmentation, streamlit

This is an open-access article under the [CC-BY-SA License](#).



How to cite: Neva, T. N., Alamsyah, Z., & Nugraha. (2026). Analysis of Convolutional Autoencoder and U-Net for Breast Tumor Detection in Mammogram Images. *RESWARA: Jurnal Riset Ilmu Teknik*, 4(4), 1666–1673. <https://doi.org/10.70716/reswara.v4i4.750>

INTRODUCTION

Cancer is a disease characterized by the abnormal and uncontrolled proliferation of body cells. Genetic mutations disrupt the regulatory mechanisms of the cell cycle, impairing the natural process of programmed cell death, or apoptosis (What Is cancer, n.d.). These malignant cells have the potential to invade adjacent tissues and spread to distant organs through metastasis. Globally and within Indonesia, cancer has emerged as a critical public health concern. It ranks as the second leading cause of death worldwide, with breast cancer incidence showing a marked increase, particularly among women (Asmalinda et al., 2022).

Cancer may develop in various anatomical sites, including the lungs, breasts, and blood, with multifactorial etiologies encompassing genetic predisposition, environmental exposures, and lifestyle factors. Uncontrolled cellular growth often results in the formation of a mass known as a tumor (Saravana, 2023). Among these, breast tissue represents one of the most common sites for tumor development. According to GLOBOCAN 2022 data, breast cancer accounts for the second highest number of new cancer cases globally, totaling 2,296,840 cases. It is also the most frequently diagnosed cancer in women, representing 23.8% of all cases (Nursoleha & Utami, 2025). Although predominantly affecting women, the disease can also occur in men. Projections indicate that global mortality from

breast cancer will continue to rise, increasing from 666,000 deaths in 2022 to an estimated 715,000 by 2025 (Arnold et al., 2022).

These alarming mortality projections underscore the critical need for improved understanding of cellular growth mechanisms to support early detection and the development of more effective therapeutic strategies (Bhat et al., 2024). One of the most impactful approaches to reducing breast cancer mortality is early detection through mammography screening. Common methods employed for early detection include breast self-examination (SADARI), clinical breast examination (SADANIS), and mammography (Nomiko et al., 2025).

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide. Early detection through mammography screening has been proven to significantly improve patient prognosis and survival rates. However, manual interpretation of mammograms by radiologists presents substantial challenges, including subjective variability, visual fatigue, and the high visual complexity of dense breast tissue, which often obscures subtle tumor masses or small microcalcifications (Brown et al., 2023).

To overcome the limitations of human visual diagnosis, prior studies have increasingly adopted deep learning techniques, particularly supervised learning architectures such as U-Net. U-Net has demonstrated remarkable performance in semantic lesion segmentation due to its encoder-decoder structure equipped with skip connections that preserve high-level spatial details. Nevertheless, supervised models depend entirely on the availability of precise pixel-level annotations made manually by medical experts—a process that is exceptionally time-consuming, costly, and prone to inter-annotator bias (Chatzara et al., 2025).

On the other hand, unsupervised learning approaches using Convolutional Autoencoders (CAE) have emerged as an efficient alternative. CAEs are trained exclusively on normal mammogram images to learn the spatial reconstruction of healthy tissue without requiring annotated lesion masks. When abnormal images are processed, the model produces a high reconstruction error or anomaly score in the lesion areas (Mienye & Swart, 2025). However, the fundamental limitation of CAEs lies in their inability to localize exact lesion boundaries with the high spatial precision offered by supervised segmentation models.

A fundamental research gap exists because existing literature predominantly evaluates U-Net and CAE in isolation, lacking a systematic comparative analysis of the trade-off between supervised localization accuracy and unsupervised annotation efficiency on the exact same dataset (Penedo et al., 2026). No comprehensive study has explicitly evaluated the performance boundary between supervised segmentation masks and unsupervised reconstruction error scores on the CBIS-DDSM mammography dataset. Comparing U-Net and CAE is scientifically justified as they represent two contrasting yet complementary paradigms in medical image analysis. This comparison is critical for evaluating ground-truth annotation dependency, output evaluation mechanisms (pixel-level precision vs error ratio), and data availability constraints in real-world clinical scenarios (Guo et al., 2026).

The novelty of this study lies in the establishment of a dual-paradigm evaluation framework that directly integrates supervised pixel localization metrics (U-Net Dice and IoU) with unsupervised reconstruction error thresholding metrics (CAE MSE and Anomaly Score) on the CBIS-DDSM dataset. Furthermore, this research formulates a standard deviation thresholding mechanism for pixel-wise reconstruction errors in CAE to convert continuous reconstruction maps into quantitative anomaly detections comparable to supervised segmentation, while integrating these findings into an interactive

Streamlit-based web application. The expected scientific contribution is to provide a comprehensive guiding benchmark on when AI-assisted diagnostic systems should leverage supervised U-Net models for precise lesion boundaries versus when unsupervised CAE models are more optimal for large-scale initial screening without annotated data dependencies (Lepcha et al., 2025).

Based on this background, this research is formulated to evaluate the performance of U-Net in semantic lesion segmentation using Dice Score, IoU, Precision, Recall, and Accuracy, to evaluate the performance of CAE in anomaly detection based on MSE reconstruction errors and Anomaly Score percentages, and to perform a cross-paradigm comparative analysis to balance the trade-off between localization precision and annotation efficiency. The primary objectives encompass the implementation and testing of U-Net and CAE architectures on the CBIS-DDSM dataset, the analysis of their technical performance trade-offs to inform clinical deployment, and the development of an interactive Streamlit application (Parra-Cabrera et al., 2026).

Academically, this study contributes to the scientific literature by providing a cross-paradigm comparative analysis of supervised versus unsupervised deep learning in medical mammography processing. Practically, the outcomes yield a prototype for an interactive web-based Clinical Decision Support System that can assist radiologists in streamlining initial screening and lesion analysis in healthcare settings.

To address the limitations of manual mammographic interpretation, artificial intelligence approaches particularly deep learning have gained increasing adoption in medical image analysis. One widely applied technique is the Convolutional Autoencoder (CAE), which learns image feature representations through encoding and decoding processes (Mienye & Swart, 2025). This architecture has proven effective for feature extraction and dimensionality reduction in complex image data. However, conventional autoencoder models exhibit limitations in preserving fine spatial details during the reconstruction process, particularly for small lesions or regions with indistinct boundaries. These limitations may affect the model's ability to accurately reconstruct abnormal areas, making the reconstruction error an important indicator for identifying potential anomalies in mammogram images (Swiecicki et al., 2021). Therefore, further analysis using different deep learning architectures is required to better understand their capabilities in mammogram image analysis (Luo et al., 2025).

A more suitable approach for enhancing segmentation performance in medical images involves the use of deep learning architectures specifically engineered to capture spatial structures. In medical image segmentation, the U-Net architecture has become one of the most extensively utilized models due to its encoder-decoder design combined with skip connections that link corresponding layers between the encoder and decoder (Azad et al., 2024). This design enables the model to retain important image details throughout the network, resulting in more precise segmentation outputs.

RESEARCH METHOD

The research methodology applied in this study is systematically designed to evaluate the performance of two primary Deep Learning architectures, namely U-Net and Convolutional Autoencoder (CAE), in analyzing mammogram images. The methodological workflow encompasses the stages of data acquisition, preprocessing, model training, evaluation using pixel-based quantitative metrics, and final system integration into a Streamlit-based interactive web application (Añez et al., 2025).

Dataset and Preprocessing Workflow

This study utilizes the comprehensive secondary dataset, the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). The total dataset is partitioned into two main cohorts: 723 mammogram images for training data and 181 mammogram images for testing data. Each mammogram image within this dataset is accompanied by a binary segmentation mask that serves as the absolute ground truth. Prior to feeding the images into the neural networks, all samples undergo standardized preprocessing steps, which include image resizing to a uniform resolution, normalization of pixel intensity values to an optimal contrast range, and local contrast enhancement to clarify subtle microcalcifications or tumor mass structures.

Quantitative Evaluation Formulas for the U-Net Model

The U-Net architecture is specifically implemented to perform the task of supervised semantic segmentation of breast lesions (Li et al., 2021). The pixel-level segmentation performance is evaluated by directly comparing the predicted image matrix (Y) against the reference ground truth mask (X). The quantitative evaluation is built upon pixel-level classification components, namely True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The formal mathematical formulas used are delineated below:

1. Dice Score / F1-Score

The Dice Score is employed to measure the degree of proximity, similarity, or spatial overlap ratio between the model-predicted lesion region and the actual region on the reference image. The mathematical formula is defined as:

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|} = (2 \times TP) / ((2 \times TP) + FP + FN) \quad (1)$$

2. Intersection over Union (IoU) / Jaccard Index

The IoU computes the ratio of the intersection area to the union area (total extent) between the prediction and the ground truth. This metric offers a rigorous linear assessment of the outer boundary matching of the segmented lesion object:

$$IoU = |X \cap Y| / |X \cup Y| = TP / (TP + FP + FN) \quad (2)$$

3. Precision

Precision measures the accuracy or reliability of the model's positive predictions, representing the ratio of correctly predicted lesion pixels to the total pixels claimed as a lesion by the model (including false positive errors):

$$Precision = TP / (TP + FP) \quad (3)$$

4. Recall / Sensitivity

Recall, or sensitivity, assesses the model's capacity to detect and recover all lesion pixels that should ideally be found from the reference image, thereby minimizing the chance of missing critical malignant lesions (false negatives):

$$Recall = TP / (TP + FN) \quad (4)$$

5. Accuracy

Accuracy provides a global overview of the binary pixel classification across the entire mammogram matrix area, evaluating both true lesion pixel discoveries (TP) and correct background/normal tissue identifications (TN):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

Quantitative Evaluation Formulas for the Convolutional Autoencoder (CAE)

Unlike U-Net, the Convolutional Autoencoder (CAE) is trained in an unsupervised manner to comprehend the spatial compression and reconstruction characteristics of normal breast tissue structures (Zadeh et al., 2020). The evaluation of abnormality determination is derived from the magnitude of the error generated when the model attempts to reconstruct the tested image.

1. Reconstruction Error

MSE is utilized to calculate the average squared difference between the original input image pixel grayscale levels (x_i) and the reconstructed output image pixels ($\hat{x}_i$), where n represents the total number of pixels in the image:

$$MSE = (1 / n) \times \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (6)$$

2. Anomaly Score (%)

The indicator determining the severity of an anomaly or the presence of a foreign lesion is calculated via the percentage of individual pixels whose reconstruction error exceeds a predefined standard deviation tolerance threshold:

$$Anomaly\ Score = (Count(MSE > Threshold) / Total\ Pixels) \times 100\% \quad (7)$$

RESULTS AND DISCUSSION

The Convolutional Autoencoder (CAE) and U-Net models were trained independently for 50 epochs. The complete architecture of the CAE model is described in Table 1.

Table 1. Convolutional Autoencoder Architecture

Layer	Operation	Filter Size	Number of Filters	Output Size
Input	Input Image	-	-	$128 \times 128 \times 1$
Conv2D_1	Convolution + ReLU	3×3	32	$128 \times 128 \times 32$
MaxPooling_1	Max Pooling	2×2	-	$64 \times 64 \times 32$
Conv2D_2	Convolution + ReLU	3×3	64	$64 \times 64 \times 64$
MaxPooling_2	Max Pooling	2×2	-	$32 \times 32 \times 64$
Conv2D_3	Convolution + ReLU	3×3	128	$32 \times 32 \times 128$
MaxPooling_3	Max Pooling	2×2	-	$16 \times 16 \times 128$

Layer	Operation	Filter Size	Number of Filters	Output Size
Conv2D_4	Convolution + ReLU	3×3	256	$16 \times 16 \times 256$
UpSampling_1	Upsampling	2×2	-	$32 \times 32 \times 256$
Conv2D_5	Convolution + ReLU	3×3	128	$32 \times 32 \times 128$
UpSampling_2	Upsampling	2×2	-	$64 \times 64 \times 128$
Conv2D_6	Convolution + ReLU	3×3	64	$64 \times 64 \times 64$
UpSampling_3	Upsampling	2×2	-	$128 \times 128 \times 64$
Conv2D_7	Convolution + ReLU	3×3	32	$128 \times 128 \times 32$
Output	Convolution + Sigmoid	3×3	1	$128 \times 128 \times 1$

The evaluation results and a performance comparison between the two models are presented below.

Table 2. Comparison of Evaluation Results

No	Aspect	Convolutional Autoencoder	U-Net
1	Dice Score	-	0.4186
2	IoU	-	0.2647
3	Precision	-	0.5254
4	Recall	-	0.3479
5	F1-Score	-	0.4186
6	Accuracy	-	0.8955
7	Mean Anomaly Score	1.4779%	-
8	Threshold	1.7578%	-
9	Detected Images	31/181	-

CONCLUSION

This study implemented Convolutional Autoencoder (CAE) and U-Net architectures for mammogram image analysis using the CBIS-DDSM dataset and integrated both models into a Streamlit-based web application. The U-Net model achieved a Dice Score of 0.4186, an IoU of 0.2647, a precision of 0.5254, a recall of 0.3479, an F1-score of 0.4186, and an accuracy of 0.8955,

demonstrating its applicability to supervised breast lesion segmentation and localization. Meanwhile, the CAE model provided a reconstruction-based approach to anomaly analysis without requiring lesion annotations during model training. Using the reported anomaly threshold of 1.7578%, the CAE model identified 31 anomalous images among the 181 test images, with a mean anomaly score of 1.4779%.

The findings suggest that U-Net and CAE serve different purposes in mammogram image analysis. U-Net is more directly suited to pixel-level lesion segmentation, whereas CAE offers an alternative for reconstruction-based anomaly detection. However, their overall performance cannot be directly ranked because they use different evaluation approaches and metrics.

This study has several limitations. The U-Net model achieved a relatively low Dice Score, indicating that lesion segmentation accuracy requires further improvement. The CAE evaluation was limited to reconstruction-error-based anomaly scores and threshold-based detection counts, without a comprehensive assessment of image-level classification performance using metrics such as sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). In addition, the study used a limited dataset partition and did not report external validation on an independent dataset or clinical validation involving radiologists. Future research should investigate threshold calibration using a clearly defined validation procedure, improve segmentation performance, and conduct external and clinical validation before considering practical clinical implementation.

ACKNOWLEDGMENT (OPTIONAL)

The authors would like to express their deepest gratitude and appreciation to Nusa Putra University, the advisors Zaenal Alamsyah, M.Kom, and Nugraha, M.Kom, as well as family and colleagues who supported this research from inception to completion.

REFERENCES

- Añez, D., Conti, G., Uriarte, J. J., Serrano-Olmedo, J. J., Martínez-Murillo, R., & Casanova-Carvajal, O. (2025). Artificial intelligence pipeline for mammography-based breast cancer detection: An integrated systematic review and large-scale experimental validation. *Medicina*, 61(12), Article 2237. <https://doi.org/10.3390/medicina61122237>
- Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66, 15–23. <https://doi.org/10.1016/j.breast.2022.08.010>
- Asmalinda, W., Jasmi, J., Setiawati, D., Khatimah, K., & Sapada, E. (2022). Deteksi dini kanker payudara menggunakan pemeriksaan payudara sendiri (SADARI). *Jurnal Pengabdian Kepada Masyarakat (ABDIKEMAS)*, 4(1), 10–17. <https://doi.org/10.36086/j.abdikemas.v4i1>
- Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical image segmentation review: The success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 10076–10095. <https://doi.org/10.1109/TPAMI.2024.3435571>
- Bhat, G. R., Sethi, I., Sadida, H. Q., Rah, B., Mir, R., Algehainy, N., ... Bhat, A. A. (2024). Cancer cell plasticity: From cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. *Cancer and Metastasis Reviews*, 43(1), 197–228. <https://doi.org/10.1007/s10555-024-10172-z>

- Brown, A. L., Vijapura, C., Patel, M., De La Cruz, A., & Wahab, R. (2023). Breast cancer in dense breasts: Detection challenges and supplemental screening opportunities. *RadioGraphics*, 43(10), Article e230024. <https://doi.org/10.1148/rg.230024>
- Chatzara, A., Maliagkani, E., Mitsopoulou, D., Katsimpris, A., Apostolopoulos, I. D., Papageorgiou, E., & Georgalas, I. (2025). Artificial intelligence approaches for geographic atrophy segmentation: A systematic review and meta-analysis. *Bioengineering*, 12(5), Article 475. <https://doi.org/10.3390/bioengineering12050475>
- Guo, S., Zhao, X., Ren, J., Shaw, M., Wan, J., & Su, J. (2026). Efficient cardiac MRI multi-structure segmentation for cardiovascular assessment with limited annotation by integrating data-level and network-level consistency. *npj Digital Medicine*, 9(1), 1–11. <https://doi.org/10.1038/s41746-026-02475-y>
- Lepcha, D. C., Goyal, B., Dogra, A., Alkhayyat, A., Sahu, P. K., Ali, A., & Kukreja, V. (2025). Deep learning in medical image analysis: A comprehensive review of algorithms, trends, applications, and challenges. *CMES—Computer Modeling in Engineering & Sciences*, 145(2), 1487–1573. <https://doi.org/10.32604/cmcs.2025.070964>
- Li, J., Cheng, L., Xia, T., Ni, H., & Li, J. (2021). Multi-scale fusion U-Net for the segmentation of breast lesions. *IEEE Access*, 9, 137125–137139. <https://doi.org/10.1109/ACCESS.2021.3117578>
- Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., Chu, W. C. W., Cheng, K. T., & Chen, H. (2025). Deep learning in breast cancer imaging: A decade of progress and future directions. *IEEE Reviews in Biomedical Engineering*, 18, 130–151. <https://doi.org/10.1109/RBME.2024.3357877>
- Mienye, I. D., & Swart, T. G. (2025). Deep autoencoder neural networks: A comprehensive review and new perspectives. *Archives of Computational Methods in Engineering*, 32(7), 3981–4000. <https://doi.org/10.1007/s11831-025-10260-5>
- Nomiko, D., Bar, A., Monalisa, M., & Eliezer, B. (2025). Effectiveness of risk detection application in breast cancer prevention in women of childbearing age. *Health Education and Health Promotion*, 13(1), 55–60. <https://doi.org/10.58209/hehp.13.1.55>
- Nursoleha, E., & Utami, W. (2025). Bibliometric analysis of drug resistance biomarkers in breast cancer (2020–2025). *International Journal of Applied Pharmaceutics*, 17(4), 10–18. <https://doi.org/10.22159/ijap.2025v17i4.54227>
- Parra-Cabrera, G., Jiménez-Delgado, J. J., & Pérez-Cano, F. D. (2026). Artificial intelligence for pulmonary abnormality detection in chest X-ray imaging: A detailed review of methods, datasets and future directions. *Technologies*, 14(3), Article 147. <https://doi.org/10.3390/technologies14030147>
- Penedo, P., Machado, J., Anjos, R., Marta, A., Silva, A. C., & Cunha, A. (2026). Weakly supervised deep learning for ocular image segmentation: A systematic review of fundus and OCT methods. *Applied Sciences*, 16(5), Article 2241. <https://doi.org/10.3390/app16052241>
- Saravana, S. K. (2023). *Fundamentals of cancer biology*. ResearchGate. <https://www.researchgate.net/publication/371165300>
- Swiecicki, A., Konz, N., Buda, M., & Mazurowski, M. A. (2021). A generative adversarial network-based abnormality detection using only normal images for model training with application to digital breast tomosynthesis. *Scientific Reports*, 11(1), Article 1–13. <https://doi.org/10.1038/s41598-021-89626-1>
- American Cancer Society. (n.d.). *What is cancer?* <https://www.cancer.org/cancer/understanding-cancer/what-is-cancer.html>
- Zadeh, H. G., Fayazi, A., Binazir, B., & Yargholi, M. (2021). Breast cancer diagnosis based on feature extraction using dynamic models of thermal imaging and deep autoencoder neural networks. *Journal of Testing and Evaluation*, 49(3). <https://doi.org/10.1520/JTE20200044>