# A Multi-Stage Deep Learning Framework for the Automated Assessment of Mammographic Microcalcifications

Ana-Maria Simion[0009−0004−1598−7909], Adina Magda Florea[0000−0001−7249−1871], and Irina Mocanu[0000−0001−5176−9344]

Department of Computer Science, National University of Science and Technology Politehnica Bucharest, Bucharest, Romania
{ana_maria.simion,adina.florea,irina.mocanu}@upb.ro

**Abstract.** Microcalcifications are among the first signs of breast cancer and early detection can completely change the course of treatment and recovery of a patient. However, they are extremely small, often only a few pixels, and very hard to detect against dense breast tissue, so they are easily overlooked during routine checks, and such human mistakes can delay diagnosis and treatment. Automatically localizing them is very difficult, given the severe tumor to background imbalance. We propose a unified multi-stage framework for automated microcalcification analysis that combines semantic segmentation, patch level false-positive suppression, detection based false-negative recovery, and image-level BI-RADS assessment within a single workflow. We further investigate the impact of different segmentation architectures, domain-specific foundation model initialization, and detection strategies on the overall pipeline performance. Our results on the INbreast dataset show that the proposed multi-stage design reduces both false positives and missed lesions, substantially increasing the proportion of lesions the complete pipeline recovers, while pixel-level segmentation remains on par with strong domain-pretrained baselines.

**Keywords:** Breast Cancer · Microcalcifications · BI-RADS · Semantic Segmentation · Mammo-CLIP · Transfer Learning.

## 1 Introduction

Breast cancer is one of the leading causes of cancer-related mortality among women [1]. The treatments that we currently have available seem to be very efficient when they are applied in the early stages of the developing cancer [2]. Therefore, it is very important to be able to detect breast cancer as fast as possible with convenient and accessible tools [3]. In some parts of the world, screening is exceptionally hard to access, therefore, developing a medical tool based on an automated artificial intelligence system would be highly valuable [4].

Moreover, a comprehensive survey on deep learning techniques for the detection and classification of mammography abnormalities [5] highlights two critical

weaknesses in current approaches. First, these methods tend not to materialize in clinical practice due to a lack of standardized information, specifically, the failure to include standard clinical terminology and metrics used by medical specialists, such as the Breast Imaging-Reporting and Data System (BI-RADS) score. Second, existing solutions rarely focus on discovering precursor abnormalities. While automatically detecting large tumor masses is beneficial, it is not as thorough or time-consuming a task as finding microcalcifications.

Microcalcifications are tiny deposits of calcium which can easily be missed on a mammogram [6]. While the presence of microcalcifications is not a definitive sign of carcinoma and it depends heavily on the morphology and dispersion of the calcium deposits, they remain among the most critical radiologic indicators of early-stage, non-palpable breast malignancies. Roughly 30% to 50% of non-palpable breast cancers are caught purely because these tiny spots appear on a mammogram [7]. Microcalcifications most commonly signal specific types of pre-invading and early invasive breast cancers, notably Ductal Carcinoma In Situ, Invasive Ductal Carcinoma and Lobular Carcinoma In Situ [8, 9].

There are two major challenges to this task. The first is data scarcity and privacy constraints, a common problem for most medical imaging tasks. The second is extreme class imbalance. Because the INbreast dataset is currently the only public dataset providing precise, semantic pixel-level segmentation masks for individual microcalcifications, it was chosen for conducting our experiments.

In this paper, we propose a pipeline to find most microcalcifications, either by doing pixel wise semantic segmentation or by determining a bounding box localization. Finally we also assign a BI-RADS score to each. The purpose of this system is to find all microcalcifications, without having too many false-positives (FP), a common problem for a task based on limited medical datasets with small regions of interest. The main contributions of this work are summarized as follows: We introduce a multi-stage mammography analysis pipeline that combines segmentation, false-positive filtering, false-negative recovery, and BI-RADS prediction for comprehensive microcalcification assessment. We systematically investigate the contribution of each stage and quantify how segmentation, classification, and detection interact to improve the overall clinical performance.

## 2 Related Work

A review of current literature shows a significant gap between research development and the actual adoption of such a technique in hospitals. In many approaches, segmentation is utilized not as the final objective, but rather as an intermediate step to facilitate some other tasks. For instance, the DeepMiCa framework [10] leverages semantic segmentation on the INbreast dataset to generate refined pixel-wise masks for the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) dataset. Their ultimate goal is to improve patch-level classification of benign and malignant regions, thereby reducing the need for unnecessary biopsies. Improved segmentation directly drives the performance of their final clinical scope. Similarly, the work in [11] uses

segmentation primarily to eliminate artifacts from mammograms. Another critical challenge in this domain is preserving the structural integrity of extremely small anomalies during preprocessing. The work in [12] addresses this by proposing a Fully Convolutional Network (FCN) designed to process mammograms at their native, high-resolution scale. They show that down-sizing mammograms can lead to the loss of actual tumor pixels due to the scale of the microcalcifications. Unfortunately they don't report segmentation specific metrics like IoU or Dice score. Overall, there are few semantic segmentation studies on microcalcifications, as mentioned in this review [4]. Something very useful for our problem was the work in [13] that introduces Mammo-CLIP, a domain specific foundation model that we can use to initialize an EfficientNet-B5 encoder, which gave our base segmentation model an improvement over ImageNet weights initialization. Furthermore, its approach to overcoming data scarcity through large-scale mammogram-report pairs aligns perfectly with our pipeline's purpose of maximizing performance on small, highly imbalanced medical datasets. Mammo-CLIP was trained on digitalized mammograms just like our data, unfortunately there are no segmentation masks associated with that dataset.

## 3 Methodology

### 3.1 Pipeline

In Fig. 1 we can see the workflow of our proposed pipeline. The input mammograms are split in patches and fed to the classification and base segmentation network. Patches containing masks are forwarded directly, while empty patches are routed to the detection module for recovery of FN patches. The resulting masks and bounding boxes are unified with a global BI-RADS prediction.

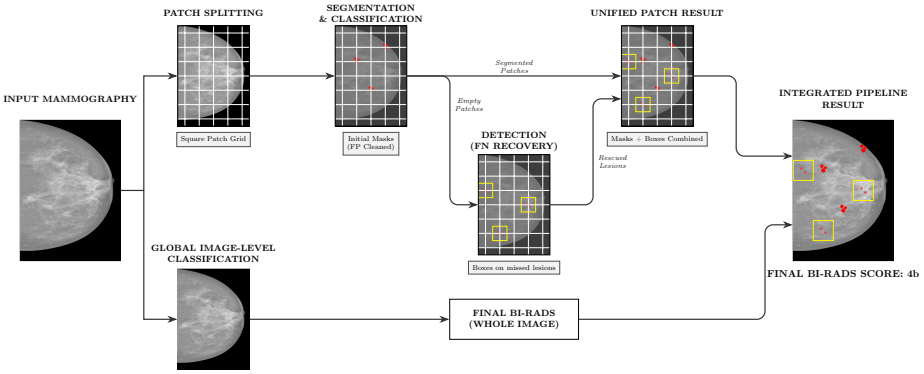


**Fig. 1.** Visual representation of our pipeline.

### 3.2 Datasets

**INbreast** We conducted our experiments on the INbreast dataset [14], which comprises a total of 410 mammograms, 313 of which contain calcifications. INbreast is particularly valuable for this task because it offers pixel-level annotations for microcalcifications and a BI-RADS score for each image. To our knowledge, it is the only public dataset of its kind (digital mammograms with pixel level microcalcifications annotations), similar repositories, such as the Weizmann/Sheba Medical Center dataset [12], have not been publicly released. The BI-RADS score directly correlates with potential malignancy. From a score of 4 and above, findings are clinically considered suspicious. In our dataset, 85% of images with a BI-RADS score $\geq 4$ contain microcalcifications, while only 15% contain mass-only tumors. Before splitting the image into patches, we removed the black background from the sides of the breast tissue, and then we padded the image until it reached sizes divisible by the size of the $256 \times 256$ patch. Image interpolation during resizing can erase small microcalcifications, so we avoid it.

**CBIS-DDSM** Because the INbreast dataset is limited we used the CBIS-DDSM [15] as a pre-training dataset. CBIS-DDSM consists of digitized screen-film mammograms and provides coarse, cluster-level annotations instead of the pixel-level microcalcification contours available in INbreast. It therefore differs from INbreast both in acquisition modality and in annotation granularity. Images and masks were passed through the same preprocessing stages as for INbreast.

### 3.3 Classification Models

**BI-RADS Classification Model** Due to the extreme class imbalance, models face severe rates of both false-negatives and false-positives. To address this, we use a classification model as a mechanism to determine whether to apply the segmentation model or the detection model. Beyond localizing the microcalcifications, we want to predict the associated BI-RADS score to bridge the communication gap between deep learning specialists and medical professionals highlighted by prior work [5].

We evaluated several architectures, including ResNet50 [16], EfficientNet [17], and InceptionV3 [18], training them with loss functions tailored for medical data. The combination of a ResNet50 backbone with Focal Loss outperformed all other configurations and we can see detailed results in Table 1.

**Table 1.** Classification report metrics of our best ResNet50 model.

| Metric | 1 | 2 | 3 | 4a | 4b | 4c | 5 | 6 | Acc. | Macro | W. Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Precision** | 0.91 | 0.92 | 1.00 | 1.00 | 1.00 | 0.94 | 0.98 | 1.00 | - | 0.97 | 0.94 |
| **Recall** | 0.88 | 0.99 | 0.83 | 1.00 | 0.88 | 0.77 | 0.90 | 1.00 | - | 0.90 | 0.94 |
| **F1-score** | 0.89 | 0.95 | 0.90 | 1.00 | 0.93 | 0.85 | 0.94 | 1.00 | 0.94 | 0.93 | 0.94 |

**False-Positive Gatekeeper Classification Model** For the gatekeeper against false-positives, we utilize an auxiliary segmentation patch-classifier to determine if an image patch consists of healthy tissue or a calcification. We conducted extended experiments with different augmentation techniques, CLAHE, patch sizes, balanced data, unbalanced data. The most important choice here proved to be the model. We used MedCLIP [19] and ResNet50 as before, but in this case MedCLIP obtained better results than ResNet50. The detailed results are available in Table 2 and we are particularly interested in class 1, which represents microcalcifications.

**Table 2.** Classification report: 0-healthy tissue, 1-microcalcification, 2-mass

| Model | Metric | 0 | 1 | 2 | Acc. | Macro | W. Avg |
|---|---|---|---|---|---|---|---|
| MedCLIP, | **Precision** | 0.86 | 0.94 | 0.85 | – | 0.88 | 0.91 |
| CLAHE, | **Recall** | 0.92 | 0.92 | 0.54 | – | 0.79 | 0.91 |
| Patches w/ calc. | **F1-score** | 0.89 | 0.93 | 0.66 | 0.91 | 0.83 | 0.91 |

### 3.4 Detection

For detection we trained two detectors on the same $256 \times 256$ patches, using cluster-level bounding boxes obtained from the ground truth segmentation masks: a RetinaNet [20] with an EfficientNet-B5 backbone initialized from Mammo-CLIP and a YOLOv11 Medium [21] initialized from its COCO pretrained weights. RetinaNet performs better and this is likely attributed to its focal loss, which is designed for exactly the extreme background imbalance and dense, small targets that characterize microcalcifications, and because its EfficientNet backbone lets us reuse the same Mammo-CLIP encoder as our segmentation models.

### 3.5 Segmentation

We tried to modify the existing AttentionUNet [22] to utilize a dual-encoder design to process at the same time a high-resolution central patch for fine details and a $3 \times 3$ grid of neighboring patches meant to capture surrounding tissue context. These two streams are merged at the bottleneck via a learned channel gate, ensuring the broader spatial context modulates the micro-level features without diluting them. Finally, an attention-gated decoder filters the skip connections to accurately pinpoint the tiny microcalcifications while effectively ignoring background artifacts. These changes were motivated by the fact that we split our mammograms into patches and we tried to keep the context of each. This is the only configuration trained from scratch, as no pretrained encoder is directly compatible with its 9-channel input. The comparison against the remaining models therefore confounds architecture with initialization, and we report it as a negative result.

For training our models we split the data in the following way: 80% for training, 10% for validation and 10% for testing. To prevent data leakage across sets, we ensured that all patches from a given mammogram remain within a single fold. The network was trained using a compound objective function, specifically the DiceBCELoss function which combines two distinct metrics to evaluate the model's predictions (1). We evaluated all models with 5-fold cross-validation.

$$\mathcal{L}_{Total} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] + 1 - \frac{2\sum_{i=1}^{N} p_i y_i + 1}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + 1} \tag{1}$$

We used Adam optimizer (learning rate $1 \times 10^{-4}$, batch size 8) and twice as many negative patches as positive ones. Data augmentation included geometric manipulations (flips and 90° rotations coupled with a matching permutation of neighbor channels) and mild intensity jitter. Training ran for up to 300 epochs with an early stopping patience of 20 epochs based on validation foreground Dice. For the EfficientNet-based models (ImageNet and Mammo-CLIP-initialized U-Net, Attention-U-Net, and SegFormer [23]), grayscale patches were replicated to three channels and normalized with ImageNet statistics to match the pretrained encoder. We minimized the compound loss objective with the AdamW optimizer (weight decay $1 \times 10^{-4}$) using discriminative learning rates—$1 \times 10^{-4}$ for the decoder and $1 \times 10^{-5}$ for the pretrained encoder, which was frozen for the first 3 epochs so the decoder could adapt before the pretrained weights were updated.

**Pre-training on CBIS-DDSM** To overcome the severe limitations of training on a small dataset, we tried to pre-train the network on the CBIS-DDSM dataset. The pre-training stage focuses heavily on calcification localization, while fine boundary delineation is supposed to be left for the fine-tuning stage on INbreast. The best model weights from pre-training were used to initialize the INbreast model, which was fine-tuned with frozen encoders for a 5-epoch warm-up, followed by full-network optimization.

## 4 Results

We evaluated eight segmentation configurations (IDs 1–8 as we can see in Table 3). Model 1 is a custom Dual Encoder Attention U-Net trained from scratch using a 9-channel input. Models 2 (ResNet-50) and 3 (EfficientNet-B5) serve as generic U-Net baselines initialized with ImageNet weights. To assess domain-specific initialization, Models 4, 6, and 7 employ an EfficientNet-B5 backbone initialized with Mammo-CLIP weights, paired with standard U-Net, Attention U-Net, and SegFormer decoders, respectively. Model 5 extends Model 4 via a two-stage pretraining approach (Mammo-CLIP → CBIS-DDSM). Finally, Model 8 evaluates a cascade approach using RetinaNet bounding boxes to prompt a fine-tuned MedSAM [24]. Mammo-CLIP initialization gives the best segmentation score of 0.695, ahead of the ImageNet EfficientNet-U-Net 0.673 and ahead of the

context model trained from scratch 0.559. This gap reflects the dominance of pretraining rather than a deficiency of the dual-encoder design itself (Sect. 3.5). The classification gate adds $\approx 0.11$–$0.21$ Dice by filtering FPs, and the RetinaNet rescue consistently outperforms YOLOv11. We made an analysis about FP catching, of 45 FP patches the classifier suppressed 43 (95.6%), but it wrongly removed 10 of 116 true lesion patches. For FN rescue the detection found a total of 345/403 ($\approx 86\%$), including 42 lesions detected but never segmented. We can see our results in Table 3. Segmentation is evaluated using Dice score. Patches that reach the false negative rescue phase are evaluated using mAP@50. In Fig. 2 we can observe how the different models integrated in the pipeline performed.

**Table 3.** Performance metrics across the pipeline stages. **(a)** Segmentation Dice scores for the base segmentation models and after the classification gatekeeper. **(b)** Standalone detection performance (mAP@50) for the rescue models.

| | | **(a) Segmentation (Dice)** | | **(b) Detection (mAP@50)** | |
|---|---|---|---|---|---|
| **ID** | **Model** | **Base** | **+ Classifier** | **RetinaNet** | **YOLOv11** |
| 1 | DualEncoderUNet (ours) | $0.5590 \pm 0.0210$ | 0.7657 | 0.7641 | 0.4426 |
| 2 | U-Net (ResNet-50, ImageNet) | $0.6521 \pm 0.0254$ | 0.7824 | 0.7794 | 0.3676 |
| 3 | EfficientNet-B5 U-Net (ImageNet) | $0.6728 \pm 0.0243$ | 0.7967 | 0.7834 | 0.4066 |
| 4 | Mammo-CLIP U-Net (EfficientNet-B5) | $\mathbf{0.6950} \pm 0.0253$ | 0.8165 | 0.7770 | 0.4046 |
| 5 | Mammo-CLIP $\rightarrow$ CBIS $\rightarrow$ INbreast | $0.6832 \pm 0.0297$ | 0.8143 | 0.7227 | 0.3254 |
| 6 | Mammo-CLIP Attention-U-Net | $0.6880 \pm 0.0259$ | **0.8303** | 0.8018 | 0.4478 |
| 7 | Mammo-CLIP SegFormer-MLP | $0.6738 \pm 0.0329$ | 0.7920 | **0.8195** | **0.5035** |
| 8 | MedSAM cascade (RetinaNet boxes) | $0.6406 \pm 0.0244$ | 0.7954 | 0.4154 | 0.1987 |
| | MC-GenRef [25] (SoTA) | 0.80 | | - | |

**Domain gap analysis.** In the base segmentation stage, the two-stage model (ID 5) reaches 0.6832 against 0.6950 for direct Mammo-CLIP initialization (ID 4). The difference is less than the fold-to-fold variability of either configuration ($\pm 0.0297$ and $\pm 0.0253$). Rather than a degradation, this indicates that the additional pretraining stage did not produce measurable benefit, for which we identified three likely causes. First, the annotation granularity: cluster-level regions reward coarse blob localization, whereas the target task requires pixel-precise delineation of individual deposits, so the decoder is pretrained towards an objective that partially conflicts with the downstream one. Second, the non-linear film response and digitization noise of screen-film acquisitions differ from full-field digital images precisely in the high-frequency band where microcalcifications reside. Third, redundancy rather than mismatch may also contribute: the encoder already carries Mammo-CLIP weights learned in-domain on digital mammograms, so a second pretraining stage on out-of-domain data can mainly overwrite representations that were already well matched to the target distribution. Consistent with this reading, the two-stage model converges faster but does not generalize better, suggesting that a pretext task decoupled from mask granularity (contrastive or reconstruction-based) would be a more productive use of CBIS-DDSM than supervised segmentation.

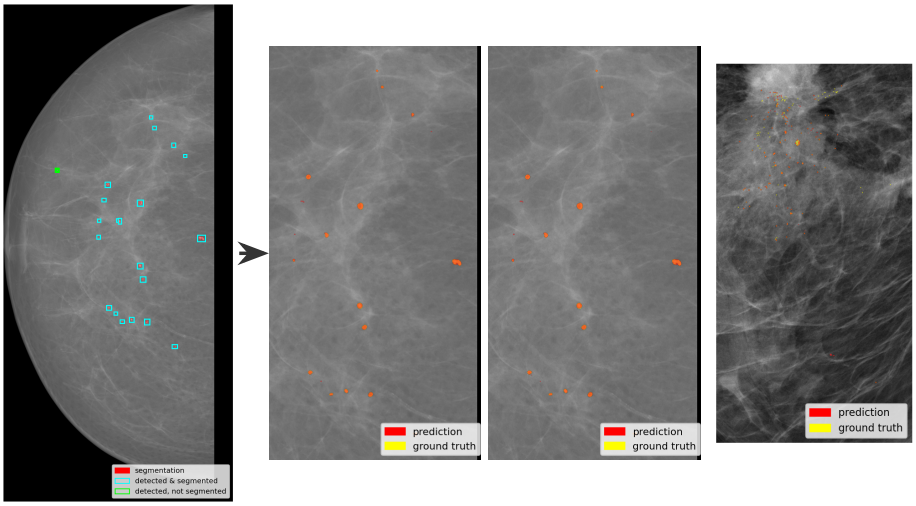


**Fig. 2.** Full mammogram view (left) and two corresponding zoomed regions of interest demonstrating different model predictions, ID for model from left to right 6, 7. The rightmost image is showcasing an example where our segmentation identifies most, but not all, microcalcifications.

## 5 Conclusions and Future Work

We presented an integrated framework for microcalcification segmentation and classification, demonstrating that pipeline success relies heavily on auxiliary stages and transfer learning rather than architectural complexity. Domain-specific pretraining (Mammo-CLIP) outperformed generic ImageNet weights and models trained from scratch, while decoder choice proved marginal. At the lesion level, the auxiliary stages dominate: a classification gatekeeper substantially reduced false positives, and a RetinaNet detector recovered lesions the segmentation branch had missed, bringing all architectures to comparable coverage. Moreover, our solution also provides a BI-RADS score.

**Limitations.** The full pipeline is validated on INbreast alone. It remains, to our knowledge, the only public dataset with pixel-level masks for individual microcalcifications, but with 313 calcification-bearing images it is small, and in the absence of an external test set the values reported here are indicative rather than generalization estimates. Our splits are also constructed at image level, so multiple views of the same breast may fall in different folds. The gatekeeper trades sensitivity for specificity, discarding 8.6% of true lesion patches, which is tolerable here only because the detection branch recovers part of them; the operating point would require explicit tuning before clinical use.

Future work will evaluate our framework using patient-level splits on larger, multi-source datasets with explicit histopathological outcomes. We will also explore pretext tasks for CBIS-DDSM that are decoupled from annotation gran-

ularity, in order to make cross-domain pretraining beneficial rather than neutral. We further plan to inflate the pretrained stem convolution to the required number of input channels, which would allow the spatial context design to be evaluated independently of initialization. Finally, we aim to investigate multimodal foundation models like MedCLIP-SAMv2 and interactive prompting methods like point-wise annotations to develop a human-in-the-loop system. Providing specialists with this level of interactive control is critical for clinical adoption.

**Acknowledgments.** This research was supported by the project "Romanian Hub for Artificial Intelligence – HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021–2027, MySMIS no. 351416.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Dessie, Z.G., Zewotir, T.: Global determinants of breast cancer mortality: a comprehensive meta-analysis of clinical, demographic, and lifestyle risk factors. BMC public health **25**(1), 2640 (2025). https://doi.org/10.1186/s12889-025-24036-w
2. Wang, L.: Early Diagnosis of Breast Cancer. Sensors **17**(7), 1572 (2017). https://doi.org/10.3390/s17071572
3. Barrios, C.H.: Global challenges in breast cancer detection and treatment. Breast **62**(Suppl 1), S3–S6 (2022). https://doi.org/10.1016/j.breast.2022.02.003
4. Ochoa Domínguez, H.d.J., Luna Lozoya, R.S., Cruz Sánchez, V.G., Vergara Villegas, O.O., Sossa Azuela, J.H., Santiago Ramirez, E.: AI-Driven Microcalcification Detection in Digital Mammography for Early Breast Cancer Diagnosis: A Scoping Review, Challenges, Limitations, and Future Perspectives. Mathematics **14**, 2367 (2026). https://doi.org/10.3390/math14132367
5. Flores, C.G.R., Ortega, J.C.P., Salazar-Licea, M.C.L.A., Fernandez, M.A.A., Osti, M.S.: A survey of approaches in Deep Learning techniques for the detection and classification of mammography abnormalities. In: 2022 19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–6. IEEE (2022). https://doi.org/10.1109/CCE56709.2022.9975900
6. Henrot, P., Leroux, A., Barlier, C., Génin, P.: Breast microcalcifications: The lesions in anatomical pathology. Diagnostic and Interventional Imaging **95**(2), 141–152 (2014). https://doi.org/10.1016/j.diii.2013.12.011
7. Logullo, A.F., Prigenzi, K.C.K., Nimir, C.C.B.A., Franco, A.F.V., Campos, M.S.D.A.: Breast microcalcifications: Past, present and future (Review). Molecular and clinical oncology **16**(4), 81 (2022). https://doi.org/10.3892/mco.2022.2514
8. Bragg, A., Candelaria, R., Adrada, B., Huang, M., Rauch, G., Santiago, L., Scoggins, M., Whitman, G.: Imaging of Noncalcified Ductal Carcinoma In Situ. Journal of clinical imaging science **11**, 34 (2021). https://doi.org/10.25259/JCIS_48_2021
9. Evans, A.: The diagnosis and management of pre-invasive breast disease: radiological diagnosis. Breast Cancer Res. **5**(5), 250–253 (2003). https://doi.org/10.1186/bcr621
10. Gerbasi, A., Clementi, G., Corsi, F., Albasini, S., Malovini, A., Quaglini, S., Bellazzi, R.: DeepMiCa: Automatic segmentation and classification of breast MIcroCAlcifications from mammograms. Computer Methods and Programs in Biomedicine **235**, 107483 (2023). https://doi.org/10.1016/j.cmpb.2023.107483

11. Leong, Y.S., Hasikin, K., Lai, K.W., Mohd Zain, N., Azizan, M.M.: Microcalcification Discrimination in Mammography Using Deep Convolutional Neural Network: Towards Rapid and Early Breast Cancer Diagnosis. Frontiers in Public Health **10**, 875305 (2022). https://doi.org/10.3389/fpubh.2022.875305
12. Zamir, R., Bagon, S., Samocha, D., Yagil, Y., Basri, R., Sklair-Levy, M., Galun, M.: Segmenting Microcalcifications in Mammograms and its Applications. SPIE 11596, Medical Imaging 2021: Image Processing (2021)
13. Ghosh, S., Poynton, C.B., Visweswaran, S., Batmanghelich, K.: MammoCLIP: A vision language foundation model to enhance data efficiency and robustness in mammography. arXiv preprint arXiv:2405.12255 (2024). https://doi.org/10.48550/arXiv.2405.12255
14. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Academic radiology **19**(2), 236–248 (2012). https://doi.org/10.1016/j.acra.2011.09.014
15. Sawyer-Lee, R., Gimenez, F., Hoogi, A., Rubin, D.: Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]. The Cancer Imaging Archive (2016). https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385 (2015). https://arxiv.org/abs/1512.03385
17. Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946 (2020). https://arxiv.org/abs/1905.11946
18. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions. arXiv preprint arXiv:1409.4842 (2014). https://arxiv.org/abs/1409.4842
19. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: MedCLIP: Contrastive Learning from Unpaired Medical Images and Text. arXiv preprint arXiv:2210.10163 (2022). https://arxiv.org/abs/2210.10163
20. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal Loss for Dense Object Detection. arXiv preprint arXiv:1708.02002 (2018). https://arxiv.org/abs/1708.02002
21. Khanam, R., Hussain, M.: YOLOv11: An Overview of the Key Architectural Enhancements. arXiv preprint arXiv:2410.17725 (2024). https://arxiv.org/abs/2410.17725
22. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas. arXiv preprint arXiv:1804.03999 (2018). https://arxiv.org/abs/1804.03999
23. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv preprint arXiv:2105.15203 (2021). https://arxiv.org/abs/2105.15203
24. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1) (2024). https://doi.org/10.1038/s41467-024-44824-z
25. Cho, H., Kwon, Y., Kim, M.J., Yoo, Y.: MC-GenRef: Annotation-free mammography microcalcification segmentation with generative posterior refinement. arXiv preprint arXiv:2604.04470 (2026). https://arxiv.org/abs/2604.04470