engineering proceedings

MDPI

Proceeding Paper

A Comparative Study of Deep CNN Architectures for Breast Lesion Classification: Performance, Efficiency, and Explainability Analysis †

Ghassane Matloub [1,*], Samira Douzi [2] and Khadija Douzi [1]

[1] LIM, Faculty of Science and Technology, University of Hassan II, P.O. Box 146, Mohammedia 28830, Morocco; khadija.douzi@fstm.ac.ma
[2] Faculty of Sciences, Mohammed V University in Rabat, P.O. Box 1014, Rabat 10000, Morocco; s.douzi@um5r.ac.ma
* Correspondence: ghassane.matloub1-etu@etu.univh2c.ma
† Presented at the 8th edition of the International Conference on Advanced Technologies for Humanity (ICATH 2026), Rabat, Morocco, 15–17 July 2026.

Abstract

This paper evaluates the effectiveness of several deep learning approaches for classifying breast lesions using the CBIS-DDSM dataset. The task is formulated as a binary classification problem, aiming to distinguish benign cases from malignant ones. Three state-of-the-art convolutional architectures are analyzed within a transfer learning framework: ResNet-50, EfficientNet-B0, and MobileNetV3-Large. Experiments show that ResNet-50 delivers strong overall performance, with an AUC of approximately 0.70 and the highest F1 score, reflecting its ability to capture complex visual features. EfficientNet-B0 achieves a favorable trade-off between predictive performance and computational efficiency, whereas MobileNetV3 Large, although highly sensitive, exhibits limited specificity and generates more false positives. Furthermore, analysis by lesion type indicates that calcifications are more easily identified than masses due to more distinct visual patterns. These observations highlight the importance of architectural choice in medical imaging tasks and provide useful insights for the deployment of our models in clinical diagnostic systems.

Keywords: breast lesion classification; mammographic Imaging; deep neural networks; convolutional neural networks; CBIS-DDSM; medical image analysis; transfer learning

1. Introduction

Breast cancer [1] represents one of the key reasons why women die across the globe. Therefore, the early diagnosis of the disease plays a critical role in improving survival chances. Mammography serves as the best approach for detecting the presence of the disease. However, this methodology faces a set of challenges and demands a considerable amount of expertise from radiologists. In turn, these issues imply the need to design efficient CAD [2] systems that will make the diagnostic procedure easier. Deep learning technologies can serve as an effective means of handling medical images because they allow automatic feature extraction. For example, the classification of breast abnormalities has been accomplished using CNNs [3]. However, although most researchers give priority to the efficiency of the models, other parameters, such as computation speed and interpretability, have been overlooked.

To overcome these weaknesses, the following research proposes a comparative study of the ResNet-50 [4], EfficientNet-B0 [5], and MobileNetV3-Large [6] architectures on

Check for updates

Academic Editors: Brahim El Bhiri, Ayman S. Mosallam and Younes Karfa Bekali

Published: 16 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the CBIS-DDSM dataset [7]. Different from previous studies, the proposed methodology is grounded on a multi-dimensional assessment approach that takes into account not only classification accuracy but also computational complexity and model interpretability through Grad-CAM interpretation. Also, the lesion types (mass and calcification) are considered during the analysis to investigate how well each of the architectures works for different medical conditions. This study contributes to the existing literature in three ways. Firstly, the comprehensive comparison of the selected neural networks under identical settings is carried out. Secondly, an explanation element is added to the evaluation process to help with the visualization of decision regions and improve the interpretability of the models. Finally, the balance between sensitivity and specificity is analyzed to provide guidance for the choice of suitable architectures considering the clinical application and computational requirements.

Whereas the prior literature focuses on only one aspect, the present work is unique in giving an integrated evaluation of performance, efficiency, and interpretability. It can be observed from the results obtained that ResNet-50 has the highest performance, mainly attributed to its high depth. EfficientNet-B0 is efficient and gives a high level of performance, but MobileNetV3-Large, which performs well due to its sensitivity, has low specificity. Another interesting observation is that calcification lesions are relatively more visible than masses, implying differences in their visualization. Overall, the current study is unique in that it builds upon the existing research work on deep learning to diagnose medical conditions and integrates different metrics such as performance, efficiency, and interpretability.

The remainder of this article is organized as follows. Section 2 describes the materials and methods, including the convolutional neural network (CNN) architectures evaluated—ResNet-50, EfficientNet-B0, and MobileNetV3-Large—as well as the training procedure and the CBIS-DDSM dataset. Section 3 presents the experimental results, including overall classification performance, specific evaluations of masses and calcifications, ROC curves, and confusion matrices. It also examines the key trade-offs between accuracy and computational efficiency. Finally, Section 4 summarizes the conclusions, discusses practical implications, and proposes future research directions.

2. Materials and Methods

The current paper considers the problem of classifying breast lesions as a binary task of separating benign and malignant cases according to the feature set obtained from mammogram regions of interest (ROI) [8]. The experiments are conducted using CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography dataset, which is freely available in TCIA [9,10]). The following CNN models are considered in a transfer learning setting: ResNet-50, EfficientNet-B0, MobileNetV3-Large.

2.1. Architecture CNN

The proposed method for CBIS-DDSM database analysis involves three CNN architectures within the transfer learning system: ResNet-50, EfficientNet-B0, and MobileNet-V3. ResNet-50 [11] is a CNN architecture that enables the extraction and interpretation of visual elements such as edges, textures, and shapes for object recognition. The model, proposed in 2015 by Microsoft Research, rapidly gained popularity in image classification tasks due to its ability to enhance the network depth without losing valuable information. This is achieved through an efficient and effective recurrent architecture, as illustrated in Figure 1.

https://doi.org/10.3390/engproc2026168013

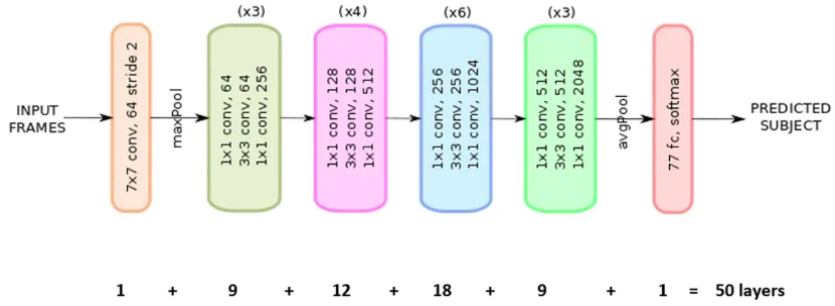


Figure 1. The architecture of the ResNet-50 with residual connections enables deep feature learning.

EfficientNet-B0 [12] relies on a uniform scaling principle, which simultaneously increases/decreases the depth, width, and resolution of the architecture. In contrast to conventional approaches that change each parameter independently, the EfficientNet technique scales them by applying a combined scaling factor. The proposed solution not only enhances the efficiency of the model but also significantly improves its effectiveness in fields like object detection, medical imaging, and embedded applications, as demonstrated in Figure 2.

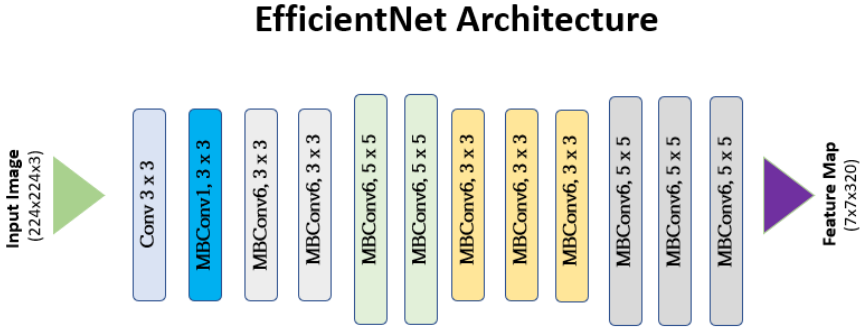


Figure 2. EfficientNet-B0 architecture illustrating compound scaling for improved accuracy and efficiency.

MobileNet-V3 [13] is available in two variants, namely MobileNet-V3 Large and Small, designed respectively for environments with sufficient computational resources and those with limited capacity. Both variants leverage Neural Architecture Search (NAS) and NetAdapt techniques to identify optimal network configurations, incorporating architecture-specific optimizations. In this study, we adopt the MobileNet-V3 Large model, as demonstrated Figure 3.

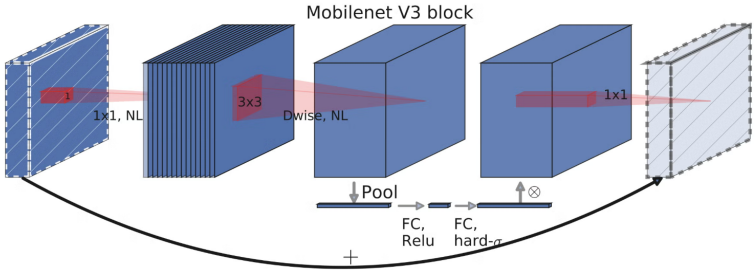


Figure 3. MobileNetV3 is designed to be efficient in resource-limited settings. The colorful blocks show the layers in the model, while the dotted lines show the residual links between layers.

2.2. Training Procedure

The primary hyperparameters of the three deep learning models [14] (ResNet50, EfficientNet-B0, and MobileNetV3-Large) utilized for lesion classification are shown in

Table 1. Using a batch size of 32 for 60 epochs, all models are trained on 224 × 224 input images using the AdamW optimizer with a cosine cooling schedule and a learning rate of 1×10^{-4}. To enhance generalization, regularization techniques such as MixUp augmentation ($\alpha = 0.4$) and dropout (which ranges from 0.3 to 0.5) are used. Furthermore, focal loss ($\gamma = 2.0$, $\alpha = 0.75$) is employed to improve the robustness of the model and solve class imbalance.

Table 1. Unified hyperparameters of the three deep learning models.

Setting	ResNet50	EfficientNet-B0	MobileNetV3-Large
IMG_SIZE	224	224	224
BATCH_SIZE	32	32	32
NUM_CLASSES	2	2	2
EPOCHS	60	60	60
BASE_LR	1×10^{-4}	1×10^{-4}	1×10^{-4}
optimizer	AdamW	AdamW	AdamW
scheduler	CosineAnnealingLR	CosineAnnealingLR	CosineAnnealingLR
dropout	0.5	0.4	0.3
MIXUP_ALPHA	0.4	0.4	0.4
FocalLoss γ	2.0	2.0	2.0
FocalLoss α	0.75	0.75	0.75

2.3. Dataset Description

The proposed framework relies on CBIS-DDSM [15,16], which is the standardized subset of the DDSM database intended for screening mammograms. Introduced by Lee et al., this dataset includes 2620 pixel-wise labeled images that contain ROI masks for both masses and microcalcifications and also the BI-RADS labels. In the current study, two kinds of lesion types are considered in order to calculate the performance of the developed algorithm under conditions of heterogeneity. It can be argued that such an approach better reflects the real-world medical practice since models have to deal with diverse visual characteristics and structures. The data set is split based on pre-established training and test subsets, whereas the DICOM format allows reproducing results and performing cross-study comparisons.

This research used the standard CBIS-DDSM partitions of training and testing. After the preprocessing step, 3568 images were kept, with 2864 images (80%) in the training dataset and 704 images (20%) in the test dataset. The training data were further divided into two parts, with 90% used for model fitting and 10% for the validation process. The validation data were used to monitor convergence and facilitate model selection, while the test data were used only for model evaluation.

3. Results and Discussion

3.1. Results

We conduct model comparison using the CBIS-DDSM dataset. The different models are compared in this section. Figures 4–6 show the variation in the accuracy and loss of the models with respect to the number of epochs.

All networks demonstrate good convergence with minimal overfitting, as indicated by the small gap between training and test losses. Among the evaluated models, ResNet-50 achieves the best performance (AUC ≈ 0.70), followed by EfficientNet-B0 (≈0.65) and MobileNetV3-Large (≈0.63). The overall moderate performance can be attributed to the limited dataset size, class imbalance, and the intrinsic complexity of mass lesions, which are characterized by high heterogeneity and ill-defined boundaries. Despite its lower performance, MobileNetV3-Large shows early convergence, suggesting that its limitations are primarily related to its lightweight architectural design rather than insufficient training.

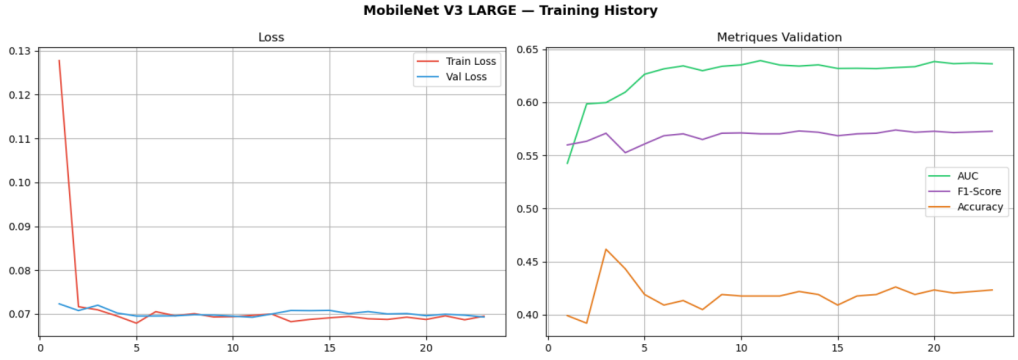


Figure 4. Training and validation curves of MobileNetV3-Large showing stable convergence with lower performance.



Figure 5. Training and validation curves of ResNet-50 demonstrating strong convergence and improved performance.



Figure 6. Training and validation curves of EfficientNet-B0 showing stable learning behavior.

ResNet-50 outperforms all other networks in terms of AUC and F1 scores due to its optimal balance of sensitivity and specificity. The choice of EfficientNet-B0 is a suitable one, considering the compromise between accuracy and model efficiency, as shown in Table 2. Meanwhile, MobileNet-V3 maximizes the sensitivity value while compromising specificity. That leads to an increase in the amount of false positives. Therefore, the poor performance is attributed more to the model architecture's simplicity rather than the lack of training.

All models perform better on calcifications than on masses, with consistent improvements in AUC, F1 score, and accuracy. This difference can be explained by the simpler and more localized nature of calcifications, whereas masses have more complex and heterogeneous structures, making them harder to classify, as shown in the following Table 3.

https://doi.org/10.3390/engproc2026168013

Table 2. The three CNN models' respective performances on the CBIS-DDSM dataset are compared.

Metric	ResNet-50	EfficientNet-B0	MobileNet-V3 Large
AUC and ROC	0.7066	0.6531	0.6391
F1 (score)	0.6078	0.5777	0.5702
Accuracy	0.5582	0.4560	0.4176
Sensitivity	0.8732	0.9493	0.9855
Specificity	0.3551	0.1379	0.0514
TP	241	262	272
TN	152	59	22
FP	276	369	406
FN	35	14	4

Table 3. Performance comparison by lesion type (Mass vs. Calcification) for the three CNN models.

Model	AUC-ROC		F1-Score		Accuracy	
	Mass	Calc	Mass	Calc	Mass	Calc
ResNet-50	0.6776	0.7481	0.5821	0.6429	0.4947	0.6319
EfficientNet-B0	0.6029	0.7118	0.5603	0.6005	0.4021	0.5184
MobileNet-V3	0.5668	0.7116	0.5600	0.5828	0.3889	0.4509

All models seem to have an issue of overpredicting "malignant" for large masses, thereby leading to high sensitivity and low specificity. On the other hand, the predictions for calcification are quite balanced, proving that calcification is easy to predict owing to its more visible nature, as depicted in the subsequent Figures 7–9.

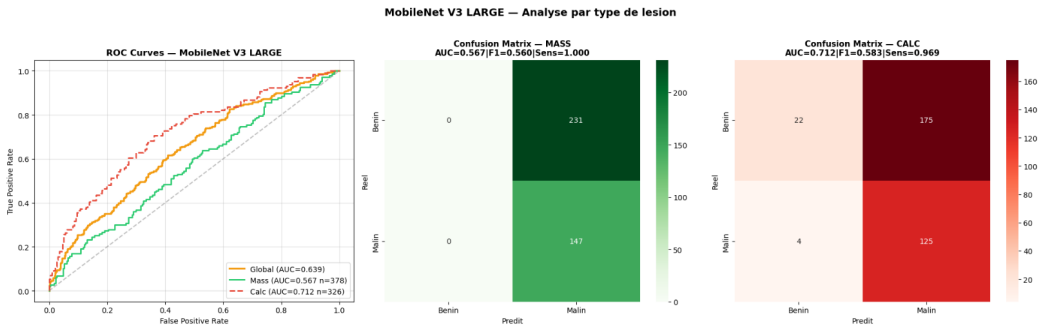


Figure 7. Confusion matrix and ROC curve for MobileNetV3-Large, illustrating high sensitivity and limited specificity. The dotted diagonal line represents a random classifier (AUC = 0.5).

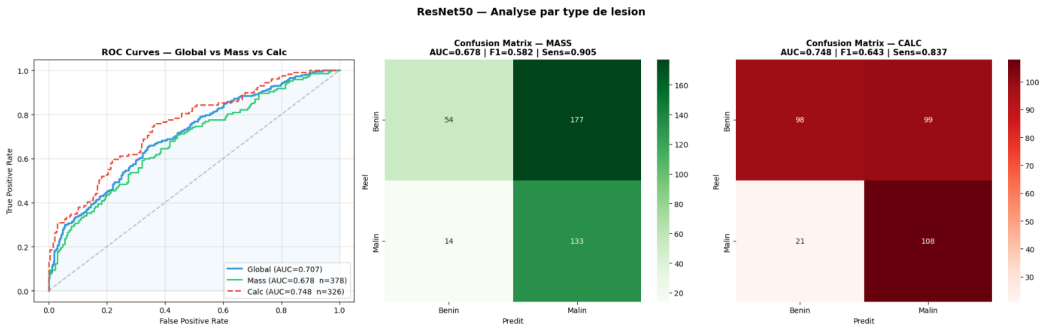


Figure 8. Confusion matrix and ROC curve for ResNet-50, illustrating strong discriminative power and balanced performance. The dotted diagonal line corresponds to a random classifier (AUC = 0.5).

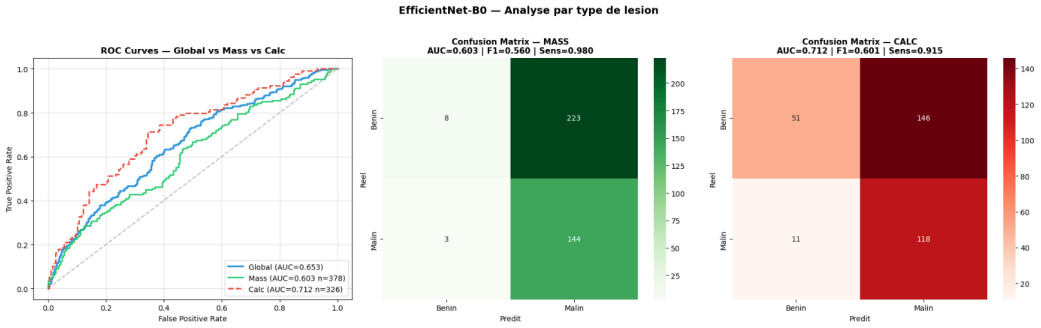


Figure 9. Confusion matrix and ROC curve for EfficientNet-B0 illustrating the trade-off between sensitivity and specificity. The dotted diagonal line corresponds to a random classifier (AUC = 0.5).

3.2. Discussion

These results indicate that there is great diversity not only in architecture but also in lesions, thus showing the unique approach of each model when used in medical image processing. The remarkable results achieved by ResNet-50 depend on its architecture and, more importantly, its capability to conduct efficient learning of hierarchically rich visual features due to residual connections. This factor is of great importance for detecting complex patterns that have already been demonstrated in numerous studies dedicated to the field of medical image processing. However, the best compromise between efficiency and performance is EfficientNet-B0.

Nevertheless, the MobileNetV3-Large model performs poorly in terms of specificity front, which leads to false positives being recorded at higher levels despite its good sensitivity level. From this, it is clear that lightweight networks, which tend to save on computation costs, fail to capture subtle discriminative features. Moreover, in terms of lesion types, there is an improved identification for calcifications as opposed to masses. These findings are in line with previous research done on CBIS-DDSM, whereby masses are difficult to classify due to high intra-class variation and heterogeneity in visual appearance. Masses are more varied in their morphology and have ill-defined borders as opposed to calcifications.

From a clinical standpoint, sensitivity being very high is a positive trait in terms of early detection because it significantly lowers the chances of missing any malignant case. However, low specificity—especially when used with light systems—may cause an uptick in the number of false positives, which translates into unnecessary further tests. Balancing these two aspects is therefore an important consideration for the effective implementation of this technology in clinical settings.

4. Conclusions and Perspective

The current research applies the CBIS-DDSM dataset in order to compare the performance of three of our models for the classification of breast cancers. It should be said that the results obtained allow us to conclude about the higher accuracy of the ResNet-50 compared to the others. In turn, the EfficientNet-B0 model allows achieving high accuracy along with cost-saving computation. Despite the extreme sensitivity of the MobileNetV3-Large model towards the tumors, it is not sufficiently specific, and therefore it may yield more false positive results than other models.

Future research directions include improving generalization through the use of larger and more diverse datasets, as well as exploring hybrid models and advanced architectures. Furthermore, the integration of multimodal data (such as clinical information) and the enhancement of explainability approaches, particularly via Grad-CAM, could increase the reliability of predictions. Finally, optimizing lightweight architectures to better bal-

ance sensitivity and specificity is a key challenge for real-time medical applications and constrained environments.

Author Contributions: Conceptualization, G.M. and S.D.; methodology, G.M.; software, G.M.; validation, G.M. and S.D.; formal analysis, G.M.; investigation, G.M.; data curation, G.M.; writing—original draft preparation, G.M.; writing—review and editing, S.D. and K.D.; visualization, G.M.; supervision, S.D. and K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted without any outside funding.

Institutional Review Board Statement: Since this paper only uses a publicly available dataset, it does not directly include any experiments.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Cancer Imaging Archive database makes the CBIS-DDSM dataset available free of charge. This information may be used for research purposes. This study did not result in the creation of any new datasets.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

This text uses the following acronyms:

CNN	Convolutional Neural Network
AUC	Area Under the Concentration-Versus-Time Curve
ROI	Region of Interest
CBIS-DDSM	Selected Digital Database's Breast Imaging Subset for Mammography Screening

References

1. Narayanankutty, A. Phytochemicals as PI3K/AKT/mtor inhibitors and their role in breast cancer treatment. *Recent Pat. Anti-Cancer Drug Discov.* **2020**, *15*, 188–199. [CrossRef] [PubMed]
2. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): The Latest Review to read first. *Radiol. Phys. Technol.* **2020**, *13*, 6–19. [CrossRef] [PubMed]
3. Gupta, J.; Pathak, S.; Kumar, G. Deep learning (CNN) and Transfer learning: A review. *J. Phys. Conf. Ser.* **2022**, *2273*, 012029. [CrossRef]
4. Arun, M.; Deenadhayalan, M.; Deepak, R.; Balaraman, S.; Deepak, D. Lung cancer detection using resnet-50 CNN architecture. In Proceedings of the 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, India, 24–27 July 2024; pp. 1–6. [CrossRef]
5. Kansal, K.; Chandra, T.B.; Singh, A. ResNet-50 vs. EfficientNet-B0: Multi-centric classification of various lung abnormalities using deep learning. *Procedia Comput. Sci.* **2024**, *235*, 70–80. [CrossRef]
6. Kavyashree, P.S.; El-Sharkawy, M. Compressed MobileNet V3: A Light Weight Variant for Resource-Constrained Platforms. In Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 27–30 January 2021; pp. 104–107. [CrossRef]
7. Liao, L.; Aagaard, E.M. An open codebase for enhancing transparency in Deep Learning-based breast cancer diagnosis utilizing CBIS-DDSM Data. *Sci. Rep.* **2024**, *14*, 27318. [CrossRef] [PubMed]
8. Yengec Tasdemir, S.B.; Tasdemir, K.; Aydin, Z. A review of mammographic region of interest classification. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1357. [CrossRef]
9. Sánchez-Femat, E.; Galván-Tejada, C.E.; Galván-Tejada, J.I.; Gamboa-Rosales, H.; Luna-García, H.; Flores-Chaires, L.A.; Saldívar-Pérez, J.; Reveles-Martínez, R.; Celaya-Padilla, J.M. CBIS-DDSM-R: A curated radiomic feature dataset for Breast Cancer Classification. *Data* **2025**, *10*, 179. [CrossRef]
10. Sandhu, J.K.; Sharma, C.; Kaur, A.; Singh Gogna, P.V.; Sharma, V. Improving breast cancer detection with deep learning techniques: A study using CBIS-DDMS dataset. In Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 13–15 November 2024; pp. 1703–1707. [CrossRef]
11. Ali, L.; Alnajjar, F.; Al Jassmi, H.; Gochoo, M.; Khan, W.; Serhani, M.A. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors* **2021**, *21*, 1688. [CrossRef] [PubMed]

12. Hoang, V.-T.; Jo, K.-H. Practical analysis on architecture of EfficientNet. In Proceedings of the 2021 14th International Conference on Human System Interaction (HSI), Gdańsk, Poland, 8–10 July 2021; pp. 1–4. [CrossRef]
13. Guo, C.; Zhou, Q.; Jiao, J.; Li, Q.; Zhu, L. A modified MobileNetV3 model using an attention mechanism for eight-class classification of breast cancer pathological images. *Appl. Sci.* **2024**, *14*, 7564. [CrossRef]
14. Zhao, J.; Huang, Q.; Liang, D.; Chen, Y.; Wang, G. Deep Learning for Medical Image Reconstruction: Focus on MRI, CT and PET. In *Deep Learning for Medical Image Analysis*, 2nd ed.; Academic Press: San Diego, CA, USA, 2024; pp. 247–278. [CrossRef]
15. Lee, R.S.; Gimenez, F.; Hoogi, A.; Miyake, K.K.; Gorovoy, M.; Rubin, D.L. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* **2017**, *4*, 170177. [CrossRef] [PubMed]
16. Falconi, L.G.; Perez, M.; Aguila, W.G.; Conci, A. Transfer learning and fine tuning in breast mammogram abnormalities classification on cbis-DDSM database. *Adv. Sci. Technol. Eng. Syst. J.* **2020**, *5*, 154–165. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.