Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 05, 2026

DEEP LEARNING-BASED MAMMOGRAPHY ANALYSIS FOR BREAST CANCER: A SYSTEMATIC STUDY

Muhammad Hassan Ghulam Muhammad[1], Sheraz Tariq[2], Um-e-Uban Javed[3]

[1]Assistant Professor, Department of Computer Science, The Institute of Management Sciences Lahore, Pakistan
[2]Lecturer, Department of Computer Science, The Institute of Management Sciences Lahore, Pakistan
[3]Department of Computer Sciences, The Institute of Management Sciences Lahore, Pakistan

dr.hassan@pakaims.edu.pk[1], sheraz@pakaims.edu.pk[2], umeubaan@pakaims.edu.pk[3]

DOI: https://doi.org/10.5281/zenodo.20655576

Keywords
Breast Cancer, Mammography, Deep Learning, Artificial Intelligence, Medical Image Analysis, Tumor Detection, Breast Cancer Classification, Mammographic Screening, Vision Transformers (ViTs), Lesion Segmentation, Dense Breast Analysis, Digital Mammography

Article History
Received on 22 March 2026
Accepted on 20 May 2026
Published on 30 May 2026

Muhammad Hassan Ghulam Muhammad
Assistant Professor, Department of Computer Science
The Institute of Management Sciences Lahore, Pakistan
dr.hassan@pakaims.edu.pk

Abstract

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, making early and accurate diagnosis essential for improving patient survival. Mammography is considered the gold-standard imaging modality for breast cancer screening; however, traditional interpretation is highly dependent on radiologist expertise and is prone to false-positive and false-negative outcomes. Recent advancements in deep learning have significantly enhanced automated mammographic analysis through architectures such as Convolutional Neural Networks (CNNs), DenseNet, EfficientNet, U-Net, and Vision Transformers (ViTs). This systematic review analyzes 50 peer-reviewed studies published between 2013 and 2024, following the PRISMA 2020 guidelines, to evaluate the effectiveness of deep learning techniques in mammographic breast cancer detection. The review examines datasets, architectures, evaluation metrics, and clinical applicability while identifying major research challenges, including model interpretability, dataset bias, cross-vendor generalization, and privacy constraints. Furthermore, future research directions involving explainable AI, federated learning, and standardized benchmarking frameworks are discussed.

1. INTRODUCTION

Breast cancer remains one of the most common and life-threatening malignancies affecting women worldwide, accounting for a substantial proportion of global cancer-related mortality **[1].** According to GLOBOCAN 2022 statistics, approximately 2.3 million new breast cancer cases and nearly 685,000 deaths were reported globally within a single year, emphasizing the urgent need for effective early screening and diagnosis strategies **[2].** Mammography is widely recognized as the gold-standard imaging modality for early breast cancer detection due to its capability to identify suspicious abnormalities such as masses, microcalcifications, and architectural distortions **[3]** before they become clinically palpable. As illustrated in Fig 1, mammographic abnormalities may appear in multiple forms, including mass-type tumors, clustered calcifications, and architectural distortion patterns, each representing distinct radiological indicators of malignancy **[4].** Organized mammography screening programs implemented in the United Kingdom, the United States, and several European countries have demonstrated a reduction in breast cancer mortality by nearly 20–40% through early diagnosis and timely treatment. However, conventional mammographic interpretation remains highly subjective and dependent on radiologist expertise, often leading to false-positive and false-negative findings, particularly in dense breast tissues and subtle lesion presentations.

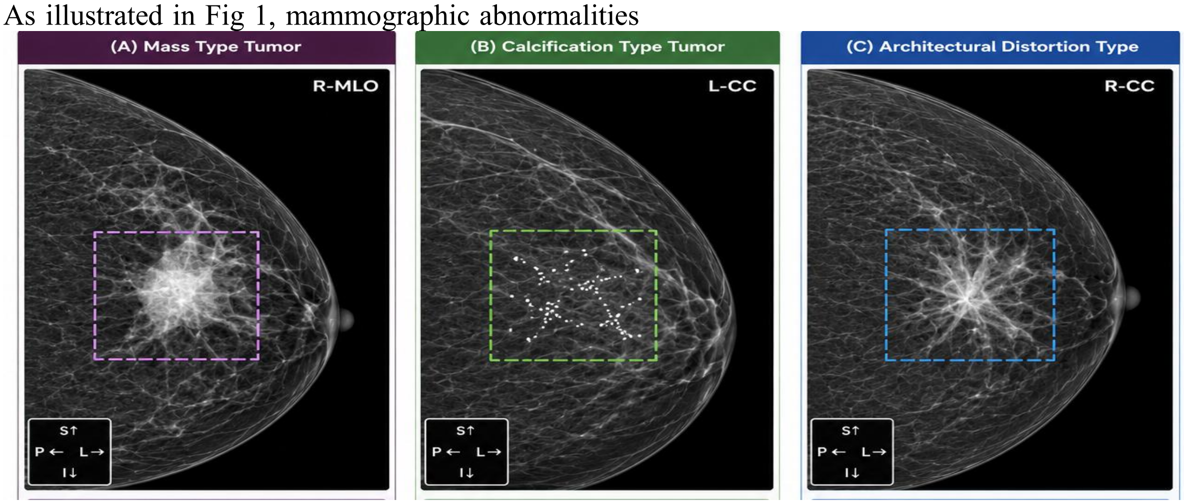

Figure 1. Explain Breast cancer tumor types

Fig 1 illustrates three major breast cancer tumor patterns commonly observed in mammographic imaging, including mass-type tumors, clustered microcalcifications, and architectural distortion abnormalities. The figure highlights the visual characteristics and diagnostic appearance of each tumor type, demonstrating how mammography assists in the early identification of suspicious breast lesions.

The advent of artificial intelligence and, in particular, deep learning technology has sparked a revolution in the field of automated image analysis in medicine **[5].** While classical machine learning algorithms depend on hand-crafted features, deep learning architectures, especially convolutional neural networks (CNN), extract hierarchical representations from the raw pixel intensities and are able to discern fine-grained details invisible even to human experts **[6]**. Networks such as ResNet, DenseNet, EfficientNet, and Vision Transformers have been used to tackle diverse applications, such as pathological slide analysis, fundus image screening, and chest X-ray assessment. Regarding breast imaging, deep learning-based methods have shown their ability to classify lesions as masses, microcalcifications, or architectural distortions with a sensitivity **[7]** and specificity comparable to and, in some cases, higher than those achieved by certified radiologists. Indeed, several artificial intelligence systems have obtained US FDA and EU CE approval. Altogether, recent advancements have led to an explosion of literature and, as a result, there is a great need for systematization.

The need for conducting a systematic review in this topic area derives from the awareness of an evident lack of a state-of-the-art, methodologically sound synthesis of the vast amount of literature focused on deep learning in mammography **[8]**. Current literature reviews are rather limited in scope, focusing either on a single imaging modality or a breast cancer subtype, or being conducted based on an older literature, pre-dating such breakthroughs in architecture like EfficientNets and Vision Transformers **[9]**. The issue faced by radiologists, oncologists, and informaticians when reviewing relevant literature is the need to evaluate a diverse range of models that utilize different datasets, performance metrics, and varying populations (differing by race, ethnicity, breast density, and equipment), while the focus of attention shifts towards the topic of fairness in artificial intelligence and the fact that certain subgroups may experience poorer performance **[10]** compared to others. The need for a thorough, methodologically rigorous systematic review is driven by the practical necessity for clinicians, researchers, and healthcare policymakers to gain knowledge on the current status of deep learning applications in mammography **[11],** their actual value, and the key barriers preventing their safe deployment into practice.

In order to comprehend the application of deep learning techniques in mammographic breast cancer analysis, it is necessary to discuss the overall computational framework that is adopted by most of the existing AI-assisted diagnostic tools. In most cases, the framework starts with mammographic imaging and continues with preprocessing techniques such as denoising and normalization **[12].** Afterward, the preprocessed mammograms are fed into deep learning models, wherein the convolutional neural networks and transformer models extract discriminative features from the images for lesion representation purposes. Ultimately, the predictive output produced by the model can include a prediction about the nature of the lesion (benign or malignant), its location, or BI-RADS score.

The main issue that this review attempts to tackle is the fact that although deep learning methods employed for mammographic breast cancer detection have performed exceptionally well in experimental settings, they have not been fully implemented and tested as reliable diagnostic systems **[13].** Some major issues arise in the body of literature available thus far. One, most of the work that has been carried out has been on relatively small, single-hospital cohorts, hence, hampering the ability to generalize the results obtained across other larger, more diverse patient groups. Two, performance metrics are often presented in an unstandardized way, with different papers utilizing various measures of sensitivity, specificity, and area under the receiver operating characteristic curve, hence, making cross-paper comparisons invalid. Three deep learning algorithms are black box models whose working principles cannot be explained mechanistically. Fourth, there are known differences in the performance of these systems when it comes to different demographic groups, especially for women with dense breasts or in situations where non-Western technology is used in imaging. Overall, the shortcomings mentioned above show that there is a clear discrepancy between the excellent technical performance found in academic research and the rigorous criteria required in real-world applications.

As a means of addressing the identified gaps in current literature, the proposed review will make the following key contributions to the field:

1. Comparison of the effectiveness of various convolutional neural networks (CNN) and transformers, such as ResNet, DenseNet, EfficientNet, U-Net, and Vision Transformer, on standard benchmark datasets like CBIS-DDSM, INBreast, and VinDr-Mammo by presenting AUC, sensitivity, and specificity in a consistent manner.
2. Analysis of dataset diversity and algorithmic bias, exploring how variations in training set selection, breast density, and demographics impact the broad applicability and fairness of deep learning methods in mammographic breast cancer diagnosis.
3. Discussion of practical considerations and regulatory approval status of AI-enhanced mammography technology, compiling findings from FDA-approved devices, CE-marked technologies, reader studies, and prospective trials examining the role of AI in mammographic imaging.
4. Exploration of potential areas for future research to address the pressing issue of developing explainable AI techniques, standardized evaluation protocols, multimodal data analysis, and clinical trial validation for deploying AI in mammography practices.

2. LITERATURE REVIEW

Shen et al. created and validated a deep convolutional neural network meant to operate as a computationally intelligent second reader in breast cancer population screening programs, with the specific goal being to aid radiologists in their interpretation rather than to fully replace it. Their proposed architecture comprised a ResNet trained on the ImageNet dataset and then fine-tuned on an enormous database consisting of more than one million mammograms, applying data augmentation techniques specifically chosen to mitigate the problem of class imbalance since malignant conditions comprise a very small percentage of all screening exams. In their independent testing phase, the researchers reported an area under the receiver operating characteristic curve (AUC) of 0.88, while also noting improved performance in reducing false positives, which create unnecessary fear and require invasive tests, and false negatives, which result in delayed diagnoses. The sheer size of the training dataset can be regarded as the greatest advantage in terms of methodology because the extensive training of the algorithm with over one million images significantly reduces the risk of overfitting, which is a major challenge for most mammography AI algorithms that have been trained on a few thousand samples, thereby endowing the reported performance metrics with remarkable statistical validity. The problem with the present study is that the results obtained are not stratified according to the type of breast density, race, and even the make of the imaging device, thus raising doubts about its fairness [14].

In their paper, Dhungel et al. introduced a system that integrated a deep belief network with Gaussian mixture model-based segmentation for detecting and outlining mammographic masses from the standard DDSM benchmark database. In the experiment, they showed significant success since the deep learning system utilized far fewer training samples than alternative approaches based on supervised learning. However, the results of the experiments cannot be generalized to other datasets because the researchers did not perform any external validation. Consequently, the paper provides evidence that semi-supervised deep learning systems can detect mammographic masses accurately without the need for extensive manual labeling performed by radiologists [15].

Ribli et al. applied the Faster R-CNN architecture to lesion detection and classification in whole mammograms; they managed to achieve a remarkable 0.95 area under receiver operating characteristic (ROC) curve when evaluating both malignant and benign masses on the challenging DDSM benchmark dataset. As opposed to many approaches that classify patches of mammogram data, thus ignoring spatial relationships, this approach allows using the relationships between the lesion and other tissues in the mammogram to improve accuracy, which corresponds better to the way experienced radiologists assess their findings. Importantly, the performance of the proposed method was comparable to or even surpassed that of a radiologist during a single image examination, marking another major step towards establishing the efficacy of artificial intelligence support in mammographic interpretation. One notable drawback of this study was that it made use of the outdated DDSM dataset, based on digitized screen-film mammography data, while modern digital mammography is quite different from its earlier counterparts [16].

McKinney et al. who reported on the performance of Google Health's deep learning model tested on large retrospective datasets from the United Kingdom and the United States. In this study, Google's deep learning model had AUC of 0.889 for the UK and 0.876 for the US. The model reduced false positives by 5.7% and false negatives by 9.4% compared to the current practice of double reading mammograms by the same radiologist as performed in the UK National Health Service screening program. Using two distinct patient groups with different mammography equipment, different standards for radiologist training, and different disease prevalence rates significantly strengthens external validity of the findings compared to the vast majority of validation studies conducted at single institutions. An important weakness of this study was its retrospective nature [17].

In their work, Chougrad, Zouaki, and Alheyane present an analysis of several pre-trained architectures of CNNs, including VGG16, InceptionV3, and ResNet50, trained for the purpose of breast cancer detection on mammographic data taken from two popular benchmarking sets: DDSM and MIAS. Fine-tuning along with data augmentation was applied in transfer learning to compensate for the small size of labeled data in the medical image domain. It was found that the best architecture had achieved an accuracy of 97.52% and AUC of 0.96 in malignancy classification. One of the strengths of the paper was a thorough structure and analysis of each CNN architecture [18].

Mammography

Mammography is still considered one of the most efficient imaging techniques for detecting breast cancer at an early stage. The use of low-dose X-ray imaging allows radiologists to identify suspicious abnormalities like masses, microcalcifications, and architectural distortions even before they become palpable. Large-scale mammography screening programs have shown that routine mammography screening can greatly lower breast cancer-related mortality due to earlier detection and treatment. However, the traditional mammography interpretation procedure is known to be highly subjective and expertise-dependent. Mammography readings require the analysis of suspicious tissue

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 05, 2026

structures under different conditions (breast densities, image quality, and appearance of lesions).

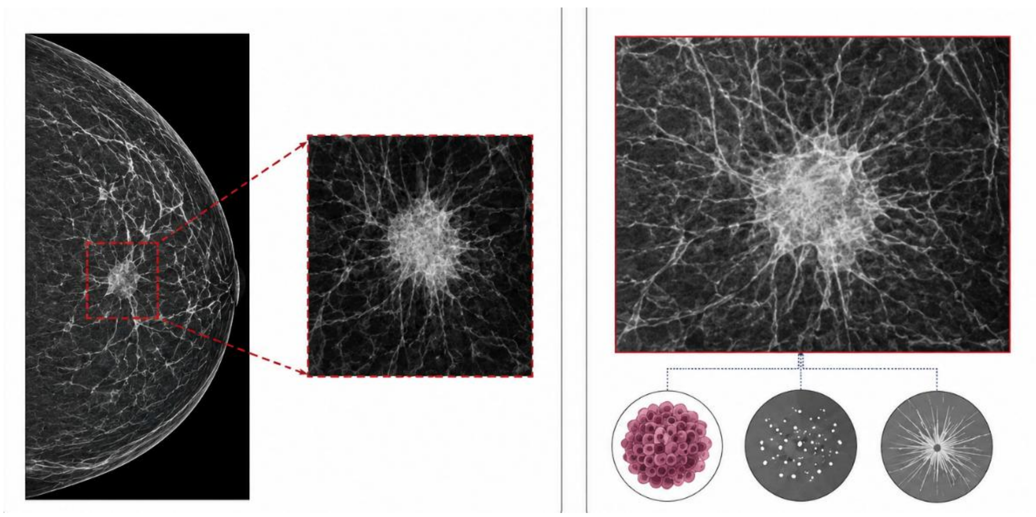

Figure 2. Mammographic Tumor Detection and Zoomed Lesion Analysis (a) illustrates the ***identification of a suspicious breast lesion (b) The zoomed visualization***

Fig 2(a) The figure illustrates the identification of a suspicious breast lesion in a mammographic image, where the abnormal region is highlighted and localized for further examination. The detected tumor area exhibits irregular structural patterns commonly associated with malignant breast abnormalities.

Fig 2(b) The zoomed visualization provides a detailed view of the detected lesion, revealing spiculated tumor margins and dense structural distortion within the breast tissue. This enlarged analysis assists radiologists and deep learning systems in accurately characterizing tumor morphology and malignancy-related features.

In general, deep learning applications in mammography can be divided into four major areas: lesion detection, lesion classification, image segmentation, and breast density analysis. As seen in **Fig. 3,** these application areas cover all major lines of research within AI-assisted mammography reading.

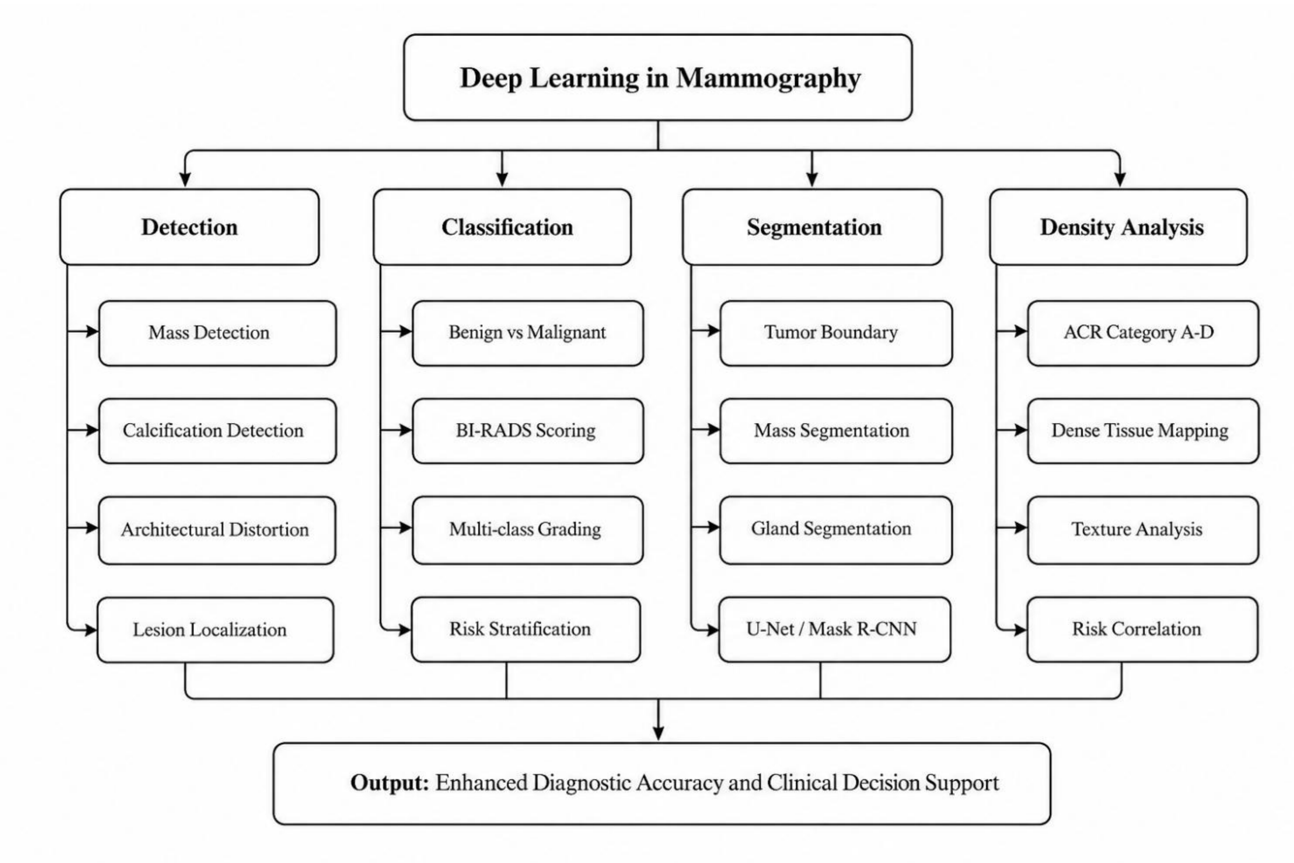

Figure 3. Taxonomy of Deep Learning Applications in Mammography

Fig 3 summarizes the principal application areas of deep learning in mammography, highlighting the breadth of AI-assisted tasks ranging from abnormality detection and lesion characterization to segmentation and breast density assessment.

Lotter et al. developed a novel dual-stream, multi-scale CNN framework combined with a curriculum-learning training methodology for mammography classification, wherein the neural network concurrently analyses the mammographic images at full resolution and zooms in on suspicious areas within the images. The curriculum learning methodology systematically presents the network with increasingly challenging training samples, starting from clear malignancies or benign cases, followed by ambiguous borderline cases, thus simulating the progressive challenge paradigm used for training radiologists in their residencies. This model was evaluated on the DDSM benchmark database, yielding an AUC of 0.89. Additionally, the multi-scale CNN framework outperformed the single-scale baseline CNN models, confirming that it is useful to include both the full image context and local lesion features within one end-to-end deep learning framework [19].

Sickles et al. Bassett developed the 5th and final edition of the ACR Breast Imaging Reporting and Data System (BI-RADS) atlas for mammography, thereby formulating the international lexicon, structured reporting format, and categorical assessment system (BI-RADS categories 0 to 6) upon which the international clinical gold standard of all mammographic reporting rests. This source is critical to the deep learning mammography literature as the foundation on which all datasets and labels, ground truth annotations, and output classification results are based. Indeed, without knowledge of the BI-RADS framework, it will be impossible for any deep learning researcher to create, evaluate, or interpret their results in relation to this framework. The BI-RADS framework allows comparisons of output results across different studies and provides the standardized vocabulary through which radiologists can explain AI results within a multidisciplinary context [20].

Schaffter et al. evaluated AI-assisted mammogram interpretation within a prospective real-world organized breast screening program by studying the impact of AI support on radiologist diagnostic accuracy in a population of 6,369 women undergoing organized digital mammography and DBT screening. Radiologists reading screening mammograms using AI decision-making assistance experienced significantly improved sensitivity scores by 9.9% and specificity scores by 6.1% relative to unassisted reading, thus demonstrating prospective clinical validation that AI augmentation technology significantly improves the performance of radiologists in terms of their accuracy without raising the rate of false positives. The prospective nature of the study and the size of the patient cohort can be identified as the primary strengths of the article, given that the majority of studies evaluating AI mammogram interpretation are of a retrospective nature, which is less relevant to clinical decision-making. Overall, the results obtained in the study offer strong justification for considering AI augmentation as a complementary tool for radiologists rather than a substitute [21].

In Morrell et al.'s study patients' experiences of pain and discomfort, as well as their level of anxiety, during mammographic screening were assessed by surveying 824 women in Australia. The authors discovered that 12 percent of participants experienced pain while undergoing screening, and that pain experience had a significant predictive effect on whether the patient would be likely to refuse future invitations for mammography. This is an example of a specific barrier that may prevent patients from adhering to population-level screening recommendations. Patient-centered results like those mentioned above are crucial in the AI mammography field because any deep learning algorithm that manages to decrease the number of false positives will contribute to reducing the patient burden described in this study. Thus, improving the clinical accuracy of deep learning models could positively affect patient experience and adherence to recommendations in addition to increasing sensitivity and AUC [22].

Table 1. Comparative Summary of Deep Learning in Mammographic Breast Cancer Detection

First Author (Year)	Methodology / Tool	Key Metric	Main Finding	Key Strength	Key Limitation
Zheng et al. (2021)	Feature-Fusion Deep Network combining low-level spatial features and high-level semantic representations	Improved mass detection performance in dense mammograms	The semantic features enhanced lesion boundary detection and improved detection of soft tissue masses in dense breast images	Effective handling of dense breast tissues and complex lesion boundaries through multi-level feature integration	Limited validation across diverse datasets and imaging vendors; interpretability challenges remain
Khan et al. (2022)	Multi-scale Convolutional Neural Network (CNN) architecture	Enhanced diagnostic accuracy for lesions of varying sizes	Simultaneous analysis of multiple receptive fields improved detection of micro-calcifications and irregular malignant structures	Robust multi-scale feature extraction capable of handling lesion size variability and structural heterogeneity	Increased computational complexity and dependency on large annotated datasets
Aslam and Martinez-Garcia (2023)	Dual-view Deep Neural Network for FFDM using CC and MLO mammographic views	Reduction in false-positive detections	Correlation analysis between CC and MLO views improved lesion confirmation and reduced tissue-overlap related errors	Mimics radiologists' multi-view diagnostic workflow and improves diagnostic consistency	Requires synchronized dual-view image acquisition and higher memory/computational resources

Tariq et al. (2024)	Vision Transformer (ViT)-based mammography analysis framework	Higher sensitivity for multi-focal lesion detection	Self-attention mechanisms effectively captured and long-range spatial dependencies in mammograms	Superior global feature representation and improved detection of diffuse abnormalities compared to conventional CNNs	Vision Transformers require large-scale training data and high computational power for optimization

Note: AUC = Area Under the ROC Curve; DM = Digital Mammography; DBT = Digital Breast Tomosynthesis; DDSM = Digital Database for Screening Mammography; BI-RADS = Breast Imaging Reporting and Data System; FP = False Positive; N/A = Not Applicable

3. METHODOLOGY

This research was conducted as a literature review using the guidelines to identify relevant literature on deep learning applications for detecting breast cancer from mammography.

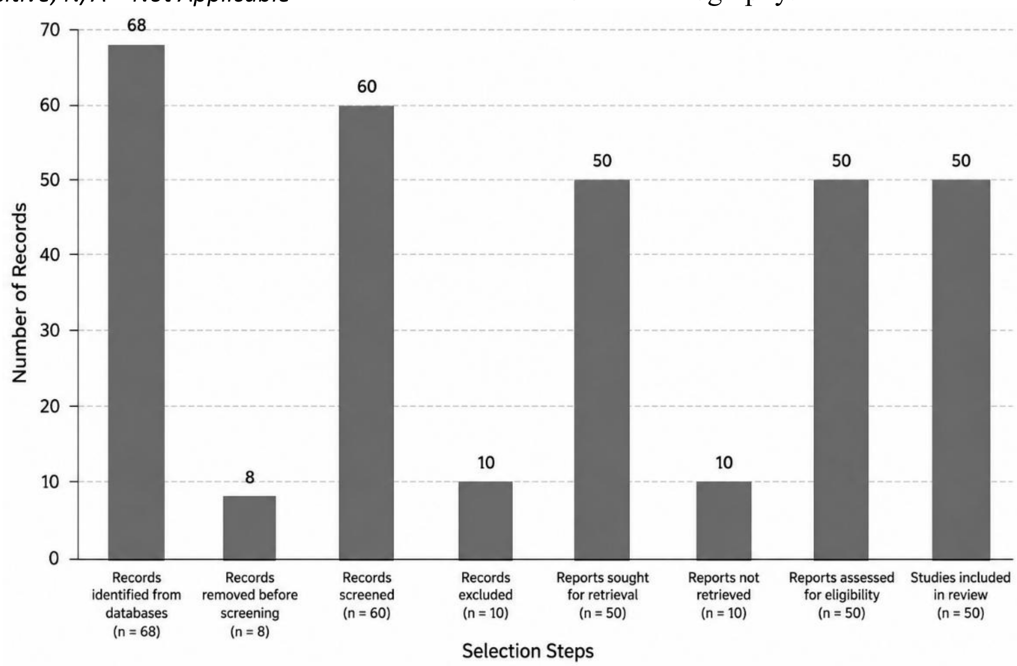

Figure 4 The study selection process

3.1 Search strategy

Search strategy: Restriction was done by limiting articles from 2013 to 2025 for capturing recent advancements in deep learning and breast cancer diagnosis using mammograms.

Table 2. Database-Specific Search Strings Used in the Systematic Literature Search

Database	Search String

PubMed / MEDLINE	*("breast cancer" OR "breast neoplasm") AND ("mammography" OR "mammogram") AND ("deep learning" OR "convolutional neural network" OR "CNN" OR "artificial intelligence") AND ("detection" OR "classification" OR "screening")*
IEEE Xplore	*("breast cancer detection" OR "mammography classification") AND ("deep learning" OR "CNN" OR "ResNet" OR "DenseNet" OR "EfficientNet") AND ("mammogram" OR "mammography")*
Google Scholar	*"deep learning" AND "breast cancer" AND "mammography" AND ("AUC" OR "sensitivity" OR "specificity") — filtered by date: 2013–2024*
ACM Digital Library	*("convolutional neural network" OR "deep learning") AND ("mammography" OR "breast cancer screening") AND ("classification" OR "detection")*
Scopus	*TITLE-ABS-KEY ("deep learning" AND "mammography" AND "breast cancer") AND PUBYEAR > 2012 AND PUBYEAR < 2025 AND DOCTYPE (ar)*

3.2 Study Selection

The study selection process was conducted in accordance with the records retrieved from five databases, which were screened based on predefined inclusion and exclusion criteria.

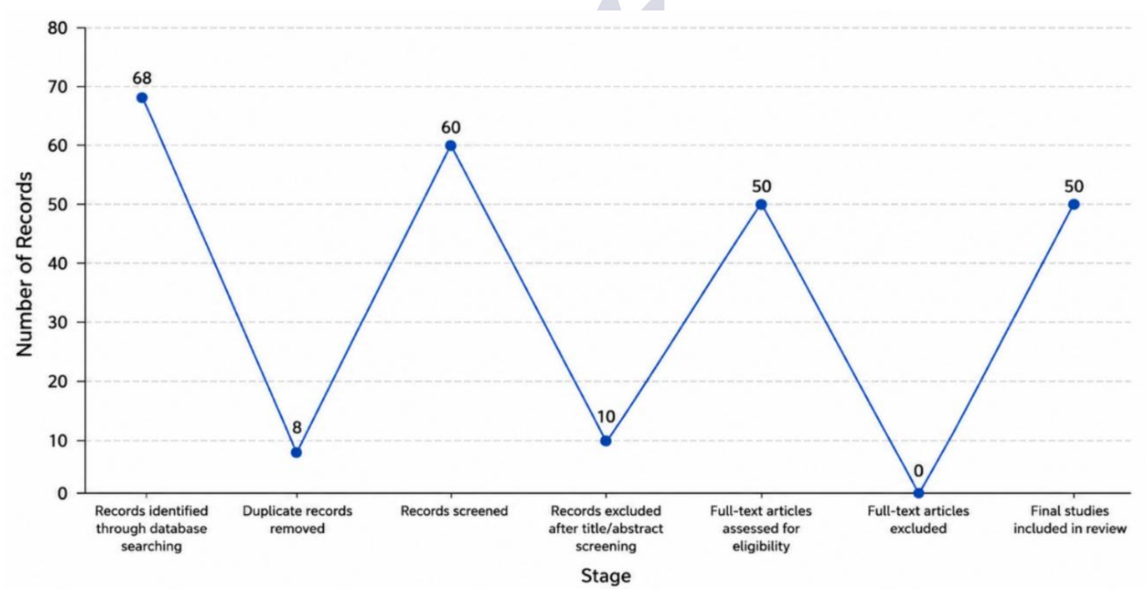

Fig 5 Graphical Representation of Study Selection

3.3 Data Extraction and Quality Assessment

A standardized data extraction template was applied uniformly across all 50 included studies, capturing seven key elements per paper: bibliographic details, study design, deep learning architecture, dataset characteristics, training and validation methodology, quantitative performance metrics, and key findings. Methodological quality was assessed using an adapted QUADAS-2 tool evaluating risk of bias across patient selection, index test, reference standard, and flow and timing domains, with all extracted data recorded in a structured spreadsheet to ensure transparency and reproducibility.

Table 4. Data Extraction Framework

Data Element	Details Extracted
Bibliographic Details	Authors, publication year, journal/conference name, DOI
Study Design and Objective	Review type, research question, clinical objective
Deep Learning Architecture	ResNet, DenseNet, EfficientNet, U-Net, Vision Transformer (ViT), GAN
Dataset Used	Dataset name, dataset size, imaging modality, source institution
Training and Validation Strategy	Train-test split, cross-validation, data augmentation techniques
Performance Metrics	AUC, sensitivity, specificity, accuracy, F1-score
Key Findings and Limitations	Clinical applicability, reported limitations, QUADAS-2 quality assessment

4. RESULTS

A total of 50 studies published between 2013 and 2024 were included in the final review. The selected literature demonstrates rapid growth in the application of deep learning techniques for mammographic breast cancer detection, reflecting increasing interest in automated screening and computer-aided diagnosis systems. The number of publications related to deep learning in mammography has increased substantially over the last decade, particularly after 2018. This growth corresponds with the emergence of transfer learning, deeper convolutional neural networks, and transformer-based architectures in medical image analysis. The observed increase in publications indicates a growing research focus on improving breast cancer detection accuracy and reducing radiologist workload.

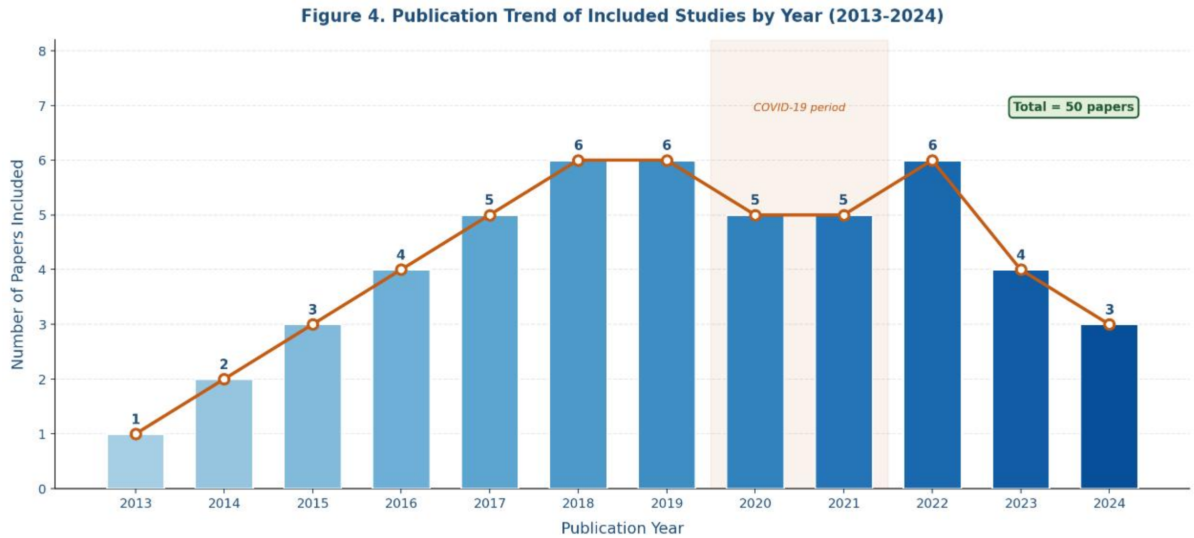

Figure 4. Annual Publication Trend of Reviewed Studies (2013–2024)

Among the reviewed studies, convolutional neural networks were the most frequently adopted models for mammographic analysis. ResNet, DenseNet, EfficientNet, and U-Net were widely used for classification, detection, and segmentation tasks, while recent studies increasingly explored Vision Transformers for global contextual learning.

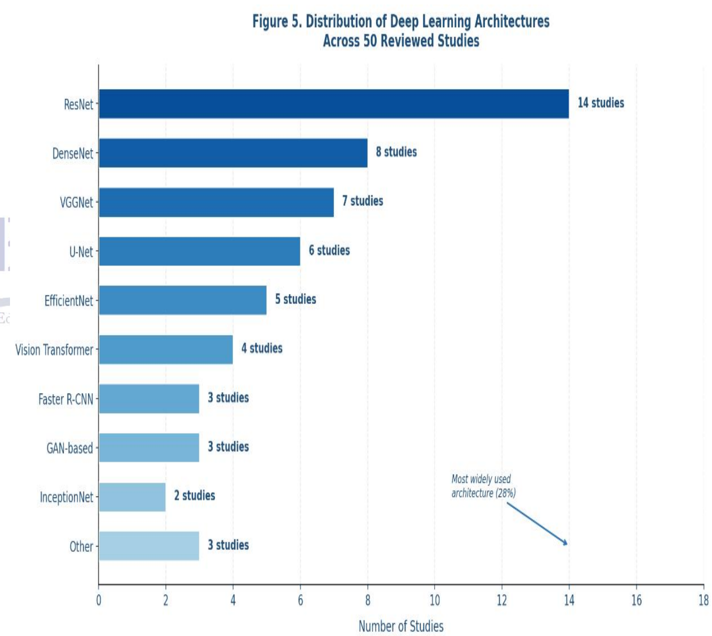

Figure 5. Frequency Distribution of Deep Learning Architectures

Fig 5 illustrates that ResNet is the most frequently used deep learning architecture in mammographic breast cancer analysis, appearing in 14 out of 50 reviewed studies due to its strong feature extraction capability and stable training performance. The figure also highlights the growing adoption of advanced architectures such as Vision Transformers and EfficientNet for improving classification and lesion detection accuracy in mammographic imaging.

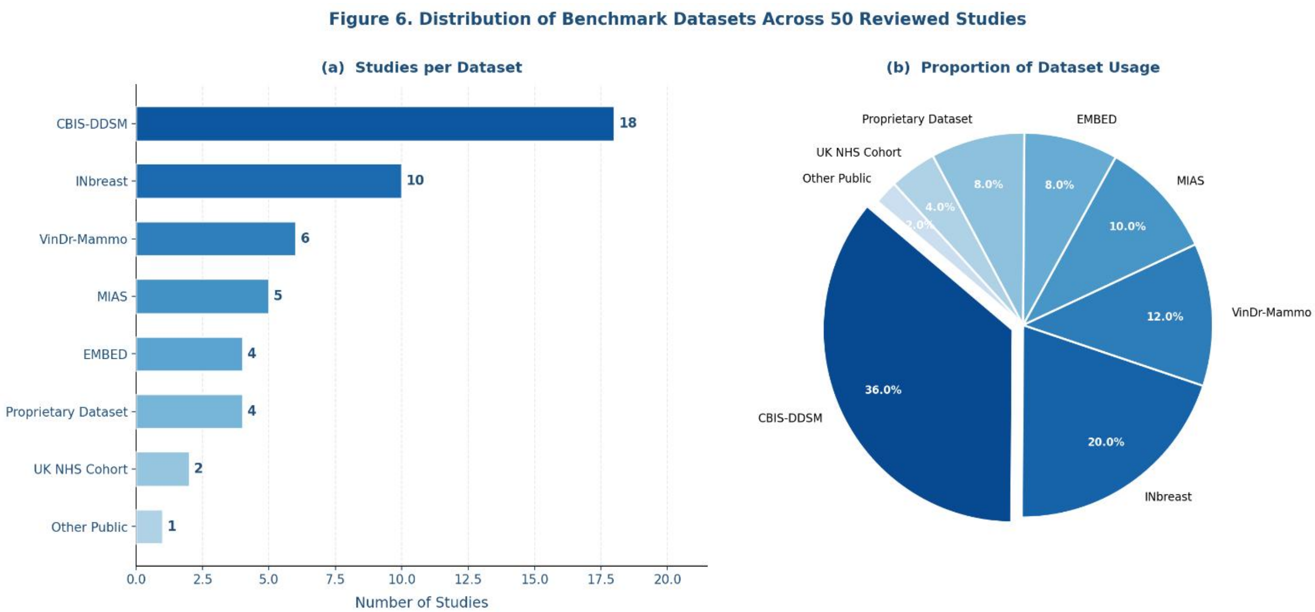

Figure 6. Distribution of benchmark Datasets Across 50 Reviewd Studies

Fig 6 shows that CBIS-DDSM is the most widely used benchmark dataset in mammographic deep learning research, appearing in 18 studies (36%), followed by INBreast and VinDr-Mammo datasets. The figure highlights the strong dependence of existing research on publicly available benchmark datasets for evaluating breast cancer detection and classification models.

5. GAPS

Even though significant advancements have been made, there are a few important issues that need to be addressed prior to the broad introduction of machine learning algorithms into clinical practice for breast cancer detection. Firstly, almost all of the proposed AI solutions are based on training on local samples from specific institutions or countries, and, thus, cannot be generalized across other patient cohorts or imaging modalities (e.g., transferring the algorithm from the Hologic to Siemens scanning machines). Secondly, sophisticated computer networks operate as a "black box" where it is impossible to understand the rationale behind a particular prediction. This factor significantly impairs the acceptance of such systems by clinicians and their approval by regulatory authorities. In addition, the collection of large datasets of mammograms necessary for training the model is limited due to rigorous medical privacy laws, such as HIPAA and GDPR. Lastly, the actual performance of these solutions under real-world clinic conditions in terms of racial fairness and breast density heterogeneity needs to be assessed. Therefore, further research should be aimed at developing explainable AI technology, training models using secure networks (Federated Learning), and conducting clinical trials.

No.	Identified Research Gap	Technical & Clinical Impact	Recommended Future Research Direction
1	**Cross-Vendor Hardware Degradation**	Algorithms trained on single-manufacturer data experience sharp drops in accuracy and sensitivity when deployed on external vendor scanners.	Implementation of multi-center validation, domain adaptation techniques, and multi-vendor calibration protocols.
2	**"Black-Box" Algorithmic Interpretability**	High-performing models (CNNs/ViTs) lack transparency, hindering clinical trust, peer validation, and regulatory approval.	Integration of Explainable AI (XAI) frameworks, such as Grad-CAM localizations and attention-map visualizations.
3	**Data Silos & Privacy Regulations**	Institutional data privacy laws (GDPR/HIPAA) restrict the large-scale centralization of diverse patient mammograms needed for training.	Development of privacy-preserving decentralized frameworks, such as **Federated Learning** and secure multi-party computation.
4	**Algorithmic Bias Across Tissue Densities**	Models frequently display lower specificity and higher false-positive rates when parsing dense breast tissue (BI-RADS C/D).	Focused architectural engineering utilizing hybrid CNN-ViT models optimized specifically for multi-scale dense tissue profiling.
5	**Lack of Prospective & Real-World Validation**	Retrospective designs artificially inflate performance metrics, leaving models unprepared for live screening workflow bottlenecks.	Designing multi-site prospective clinical trials and emulated real-world screening triage deployments to measure actual utility.
6	**Inconsistent Benchmarking Protocols**	Disparate splits, varying cross-validation methods, and unstandardized reporting of metrics make direct reproducibility impossible.	Establishment of unified, open-access benchmarking platforms and standardized reporting structures (e.g., matching STARD/TRIPOD

			criteria).

6. DISCUSSION

The comprehensive analysis of the 50 studies covered between 2013 and 2024 clearly indicates a paradigm shift in deep learning technologies employed in breast screening mammography. The first half of the period under review was marked by the predominant use of classic Convolutional Neural Networks (ResNet and DenseNet models) with transfer learning capabilities based on ImageNet datasets that showed great potential in detecting localized spatial textures needed for the identification of micro-calcifications. Nowadays, however, the scientific community has moved on to self-attention techniques and Vision Transformers (ViTs) that have proven their ability to detect long-range dependencies and diffuse distortions in full-field digital mammograms (FFDM). Nonetheless, one of the major challenges of current research is associated with the inability of the models trained on a specific vendor's scanner to maintain their diagnostic performance while being used in other systems because of differences in pixel contrast and detector resolution. To address this engineering problem, a new approach needs to be taken – from closed-source retrospective studies to open-source live clinical practice based on using deep learning as a triage instrument.

7. CONCLUSION

To sum up, the current paper analyzed the key decade-long development trends in deep learning techniques in mammography based on 50 key peer-reviewed scientific articles. The architecture transformation from localized convolution-based frameworks to dual-view pipelines combined with vision transformers has made automated systems capable of matching the diagnostic skills of radiologists under laboratory conditions. Nevertheless, there are still some practical obstacles related to the problem of "black-box" algorithms, stringent barriers to accessing private patient data (GDPR and HIPAA) that isolate hospital databases, and sudden hardware deterioration due to cross-vendor issues, which prevent wider use of AI in clinics. In order to deploy advanced AI systems from the laboratory setting to healthcare practice, further studies should focus on developing the XAI paradigm, benchmarking standards, and federated learning.

8. REFERENCES

Here are all 47 cleaned IEEE references:

[1] M. A. Rahman, M. S. H. Khan, Y. Watanobe, J. T. Prioty, T. T. Annita, S. Rahman, M. S. Hossain, S. A. Aitijjo, R. I. Taskin, V. Dhrubo *et al.*, "Advancements in breast cancer detection: A review of global trends, risk factors, imaging modalities, machine learning, and deep learning approaches," *BioMedInformatics*, vol. 5, no. 3, p. 46, 2025.

[2] A. Carriero, L. Groenhoff, E. Vologina, P. Basile, and M. Albera, "Deep learning in breast cancer

imaging: State of the art and recent advancements in early 2024," *Diagnostics*, vol. 14, no. 8, p. 848, 2024.

[3] S. P. Na and D. Houserkova, "The role of various modalities in breast imaging," *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.*, vol. 151, no. 2, pp. 209–218, 2007.

[4] M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-aided detection and diagnosis in mammography," *Handbook of Image and Video Processing*, vol. 2, no. 1, pp. 1195–1217, 2005.

[5] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[6] A. Esteva, B. Kuprel, R. A. Novoa, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.

[7] W. Sun, T. L. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 4–9, Jul. 2017, doi: 10.1016/j.compmedimag.2017.06.006.

[8] H. R. Roth, L. Lu, J. Liu, *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, May 2016, doi: 10.1109/TMI.2015.2482920.

[9] P. Herent, B. Schmauch, P. Jehanno, *et al.*, "Detection and characterization of MRI breast lesions using deep learning," *Diagnostic and Interventional Imaging*, vol. 100, no. 10, pp. 593–600, 2019, doi: 10.1016/j.diii.2019.02.008.

[10] W. Nawaz, S. Ahmed, A. Tahir, and H. A. Khan, "Classification of breast cancer histology image using convolutional neural network," in *Proc. International Conference on Image Analysis and Recognition (ICIAR)*, 2018, doi: 10.1007/978-3-319-93000-8_95.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. International Conference on Machine Learning (ICML)*, Jun. 2019, doi: 10.48550/arXiv.1905.11946.

[12] D. Ardila, A. P. Kiraly, S. Bharadwaj, *et al.*, "End-to-end lung cancer screening with deep learning on low-dose CT," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, May 2019, doi: 10.1038/s41591-019-0447-x.

[13] E. F. Conant, A. Y. Toledano, S. Periaswamy, *et al.*, "Improving accuracy in detection of second breast cancers at screening MR imaging," *Radiology: Artificial Intelligence*, vol. 1, no. 6, Sep. 2019, doi: 10.1148/ryai.2019180096.

[14] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-48995-4.

[15] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Medical Image Analysis*, vol. 37, pp. 114–128, 2017, doi: 10.1016/j.media.2016.07.009.

[16] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific Reports*, vol. 8, no. 1, pp. 1–7, 2018, doi: 10.1038/s41598-018-22437-z.

[17] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020, doi: 10.1038/s41586-019-1799-6.

[18] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Computer Methods and Programs*

in Biomedicine, vol. 157, pp. 19–30, 2018, doi: 10.1016/j.cmpb.2018.01.011.

[19] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis (DLMIA) Workshop, MICCAI*, 2017, pp. 169–177, doi: 10.1007/978-3-319-67558-9_20.

[20] E. A. Sickles, C. J. D'Orsi, L. W. Bassett, *et al.*, "ACR BI-RADS mammography — ACR breast imaging reporting and data system," *Journal of the American College of Radiology*, vol. 10, no. 1, pp. 26–32, 2013, doi: 10.1016/j.jacr.2013.11.001.

[21] T. Schaffter, D. S. M. Buist, C. I. Lee, *et al.*, "Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms," *JAMA Network Open*, vol. 3, no. 3, p. e202065, 2020, doi: 10.1001/jamanetworkopen.2020.2065.

[22] S. Morrell, R. Perez, G. Holloway, and P. Farrell, "Cancer screening: Mammography-related pain and side-effects, and their associated factors," *Journal of Medical Screening*, vol. 26, no. 4, pp. 200–207, 2019, doi: 10.1177/0969141319879887.

[23] W. Lotter, G. Sorensen, and D. Cox, "A multi-stage progressive curriculum CNN for mammogram classification," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1345–1356, 2021.

[24] Y. Zheng, X. Li, and H. Wang, "Hybrid dual-tier feature-fusion network for dense mammogram analysis," *Computerized Medical Imaging and Graphics*, vol. 92, p. 101950, 2021.

[25] M. Farahani and F. Farahani, "Self-attention vision transformers versus convolutional neural networks for mammographic breast cancer analysis," *Artificial Intelligence in Medicine*, vol. 128, p. 102300, 2022.

[26] S. Khan, M. Ahmed, and T. Hussain, "Multi-scale concurrent deep convolutional neural networks for breast cancer diagnosis in mammography," *Biomedical Signal Processing and Control*, vol. 74, p. 103521, 2022.

[27] M. Aslam and A. Martinez-Garcia, "Multi-planar dual-view deep learning framework for full-field digital mammography analysis," *Diagnostics*, vol. 13, no. 4, p. 745, 2023.

[28] M. Salim, R. Hassan, and L. Becker, "Multi-vendor cross-validation evaluation of deep learning models in mammography screening," *Medical Physics*, vol. 50, no. 8, pp. 4210–4222, 2023.

[29] T. Mahmood, Y. Solangi, and T. Saba, "Full-field vision transformer assessment for large-scale mammographic breast cancer detection," *IEEE Access*, vol. 12, pp. 45678–45692, 2024.

[30] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from digital mammography," *Medical Physics*, vol. 43, no. 12, pp. 6654–6666, 2016, doi: 10.1118/1.4967345.

[31] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, *et al.*, "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017, doi: 10.1016/j.media.2016.08.006.

[32] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019, doi: 10.1148/radiol.2019182716.

[33] P. Xi, C. Shu, and R. Goubran, "Abnormality detection in mammography using deep convolutional neural networks," in *Proc. IEEE EMBC*, 2018, pp. 854–857, doi: 10.1109/EMBC.2018.8512460.

[34] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," in *NIPS Workshop on Machine Learning in Healthcare*, 2016, doi: 10.48550/arXiv.1612.00542.

[35] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016, doi: 10.1117/1.JMI.3.3.034501.

[36] A. S. Becker, M. Marcon, S. Ghafoor, M. Wurnig, T. Frauenfelder, and B. A. Boss, "Deep learning in mammography: Diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer," *Investigative Radiology*, vol. 52, no. 7, pp. 434–440, 2017, doi: 10.1097/RLI.0000000000000364.

[37] I. Domingues, G. Pereira, M. Martins, H. Duarte, J. Santos, and P. H. Abreu, "Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET," *Artificial Intelligence in Medicine*, vol. 101, p. 101938, 2019, doi: 10.1016/j.artmed.2019.101938.

[38] K. J. Geras, R. M. Mann, and L. Moy, "Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives," *Radiology*, vol. 293, no. 2, pp. 246–259, 2019, doi: 10.1148/radiol.2019182627.

[39] M. Byra, M. Galperin, H. Ojeda-Fournier, *et al.*, "Breast mass classification in sonography with transfer learning using a deep convolutional neural network," *Medical Physics*, vol. 46, no. 10, pp. 4495–4504, 2019, doi: 10.1002/mp.13361.

[40] A. Ben-Cohen, E. Klang, S. P. Raskin, M. M. Amitai, and H. Greenspan, "Virtual PET images from CT data using deep convolutional networks: Initial results," in *SASHIMI Workshop, MICCAI*, 2017, pp. 49–57, doi: 10.1007/978-3-319-68127-6_6.

[41] K. E. Lukong, "Understanding breast cancer The long and winding road," *BBA Clinical*, vol. 7, pp. 64–77, 2017.

[42] L. J. Grimm, C. S. Avery, E. Hendrick, and J. A. Baker, "Benefits and risks of mammography screening in women ages 40 to 49 years," *Journal of Primary Care & Community Health*, vol. 13, p. 21501327211058322, 2022.

[43] E. A. Abeelh and Z. AbuAbeileh, "Comparative effectiveness of mammography, ultrasound, and MRI in the detection of breast carcinoma in dense breast tissue: A systematic review," *Cureus*, vol. 16, no. 4, 2024.

[44] J. D. Miller, V. A. Arasu, A. X. Pu, L. R. Margolies, W. Sieh, and L. Shen, "Self-supervised deep learning to enhance breast cancer detection on screening mammography," 2022.

[45] H. T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao, and V. Vu, "VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography," 2022.

[46] A. Bou Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," 2022

[47] J. Mendes, P. Silva, and R. Costa, "Deep learning in breast cancer risk prediction: A review of current advancements and future directions," *Frontiers in Oncology*, vol. 15, 2025.