Deep Learning for Contrast Enhanced Mammography - a Systematic Review

Vera Sorin, MD[1]; Miri Sklair-Levy, MD[1]; Benjamin S. Glicksberg, PhD[2]; Eli Konen, MD[1];

Girish N. Nadkarni, MD[2,3]; Eyal Klang, MD[2,3]

[1]Department of Diagnostic Imaging, Chaim Sheba Medical Center, affiliated to the Sackler School of Medicine, Tel-Aviv University, Israel

[2]Division of Data-Driven and Digital Medicine (D3M), Icahn School of Medicine at Mount Sinai, New York, New York, USA

[3]The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Corresponding Author:

Vera Sorin, MD

Department of Diagnostic Imaging, Chaim Sheba Medical Center

Address: Emek Haela St. 1, Ramat Gan, Israel, 52621.

Tel: +972-3-5302530, Fax: +972-3-5357315, Email: verasrn@gmail.com

Abstract

Background/Aim:

Contrast-enhanced mammography (CEM) is a relatively novel imaging technique that enables both anatomical and functional breast imaging, with improved diagnostic performance compared to standard 2D mammography. The aim of this study is to systematically review the literature on deep learning (DL) applications for CEM, exploring how these models can further enhance CEM diagnostic potential.

Methods:

This systematic review was reported according to the PRISMA guidelines. We searched for studies published up to April 2024. MEDLINE, Scopus and Google Scholar were used as search databases. Two reviewers independently implemented the search strategy.
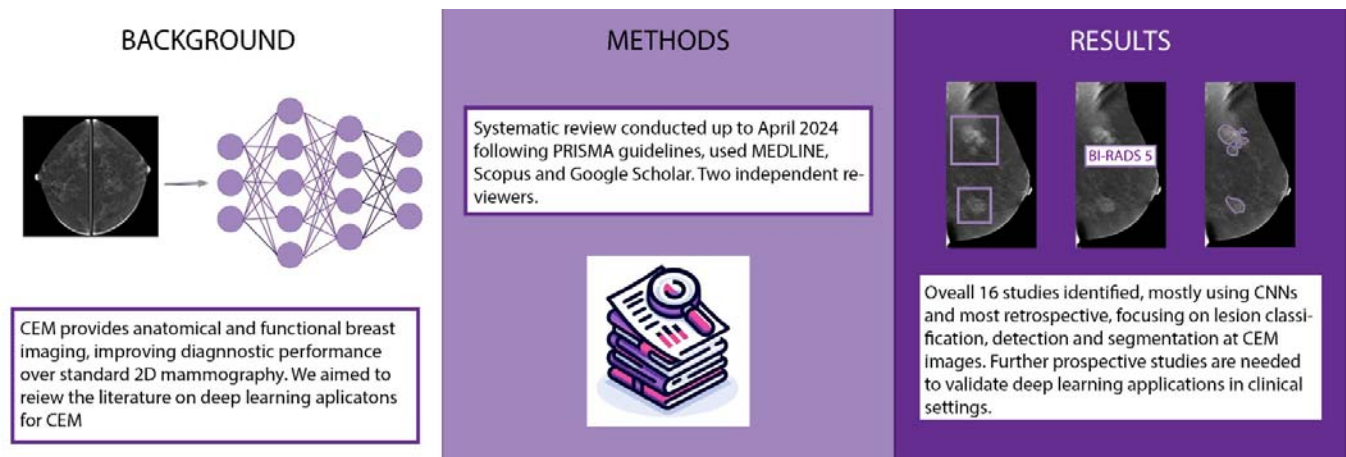
Results:

Sixteen relevant studies published between 2018 and 2024 were identified. All studies but one used convolutional neural network models. All studies evaluated DL algorithms for classification of lesions at CEM, while six studies also assessed lesion detection or segmentation. In three studies segmentation was performed manually, two studies evaluated both manual and automatic segmentation, and ten studies automatically segmented the lesions.

Conclusion:

While still at an early research stage, DL can improve CEM diagnostic precision. However, there is a relatively small number of studies evaluating different DL algorithms, and most studies are retrospective. Further prospective testing to assess performance of applications at actual clinical setting is warranted.

Graphic Abstract:

BACKGROUND

CEM provides anatomical and functional breast imaging, improving diagnnostic performance over standard 2D mammography. We aimed to reiew the literature on deep learning aplicatons for CEM

METHODS

Systematic review conducted up to April 2024 following PRISMA guidelines, used MEDLINE, Scopus and Google Scholar. Two independent reviewers.

RESULTS

BI-RADS 5

Oveall 16 studies identified, mostly using CNNs and most retrospective, focusing on lesion classification, detection and segmentation at CEM images. Further prospective studies are needed to validate deep learning applications in clinical settings.

Introduction

Despite being the cornerstone of breast imaging, 2D standard mammography faces well-known diagnostic challenges (1-3). Contrast-enhanced mammography (CEM) is an advanced imaging technique that enables both anatomical and functional breast imaging (4). The technique enhances the visibility of neo-vascularity in breast cancer. It illustrates pathologic enhancement in breast malignancies, sometimes even before the development of a distinct breast mass (5). Comparative studies have demonstrated CEM's superior diagnostic performance over conventional mammography and a comparable sensitivity to breast MRI(6).

Recent advancements in deep learning (DL) have revolutionized computer vision in radiology (7). DL algorithms are increasingly applied to enhance the segmentation, detection, and classification of pathologies across various imaging modalities, including CEM (8). However, the integration of DL in CEM is still nascent, with significant potential to improve diagnostic accuracy, and enhance radiologists' workflow.

Our study aims to systematically review the literature on DL applications for CEM, exploring how these models can further enhance its diagnostic potential.

Methods

Literature Search

We systematically searched for papers published up to April 9, 2024. MEDLINE, Scopus and Google Scholar were used as databases. The search query was: "((deep learning[Title/Abstract]) OR (CNN[Title/Abstract]) OR (AI[Title/Abstract]) OR (artificial intelligence[Title/Abstract]) OR (neural network[Title/Abstract])) AND

((CEM[Title/Abstract]) OR (CESM[Title/Abstract]) OR (contrast enhanced

mammography[Title/Abstract]) OR (contrast mammography[Title/Abstract]) OR

(DEM[Title/Abstract]))".

Eligibility Criteria

We included full publications indexed in PubMed, evaluating DL for automatic analysis of

CEM images. We excluded non-English papers, non-original manuscripts, and papers that did

not evaluate DL algorithms. **Figure 1** presents a flow diagram of the screening and inclusion

process.

Screening and Synthesis

This review was reported according to the PRISMA (Preferred Reporting Items for

Systematic Reviews and Meta-Analyses) guidelines. Two reviewers (VS, EK) independently

implemented the search strategy. The reviewers screened the titles and reviewed the abstracts

of the articles identified in the search. The reviewers then assessed the full text of the relevant

studies. Results of the studies included were summarized in a table format and included the

title, journal, year of publication, study design, size of dataset, and performance metrics.

Performance Characteristics

Performance characteristics for DL algorithms are evaluated based on sensitivity (also called

recall), specificity, area under receiver operating characteristic curve (AUC), precision (also

positive predictive value), accuracy and F1 score. The reported performance metrics from

each study were summarized in a Table.

The quality of the studies was evaluated based on the Quality Assessment of Diagnostic

Accuracy Studies (QUADAS-2) criteria(9).

Fundamental Concepts

Contrast-Enhanced Mammography (CEM)

CEM is an advanced breast imaging modality that combines the use of iodine-based contrast agent with conventional digital mammography (4). With CEM, dual energy for mammographic acquisition after IV administration of iodine-based contrast agent, is used. Two sets of images are obtained: images with low-energy exposures, and images at high-energy exposures, just above the kVp of iodine. The two exposures are subtracted, and two images are generated for each projection: low-energy images that are analogues to standard 2D mammography, and subtracted contrast-enhanced images displaying areas of contrast enhancement (**Figure 2**) (10).

Background parenchymal enhancement (BPE) refers to the normal enhancement of breast glandular elements and is graded on a BI-RADS scale from 0 to 3, indicating minimal, mild, moderate, or marked enhancement (**Figure 3**) (11). BPE at CEM correlates with age, menopausal status, breast density, and with BPE seen at MRI (12, 13). It is also associated with increased risk for breast cancer (13). More importantly, it may affect the diagnostic accuracy of CEM, as small abnormalities may blend with BPE when it is increased (11).

Models and Techniques

Deep Learning represents a subset of machine learning (**Figure 4**) that employs neural networks with multiple layers to analyze complex relationships in data. It excels in identifying patterns and making decisions with minimal human intervention. In the domain of medical imaging, DL algorithms process extensive imaging data, learning to analyze pathological findings with precision that often matches or exceeds that of expert radiologists. For instance, a DL model can identify malignant tumors in CEM images. This capability may

not only enhance diagnostic accuracy, but also streamline the workflow, thus impacting patient outcomes through appropriate intervention (7, 14).

Multi-Layer Perceptron (MLP) is a neural network model that consists of at least three layers of artificial neurons: an input layer, hidden layers and an output layer (15). MLP utilizes backpropagation with supervised learning for training. With supervised learning the algorithm is provided with examples of inputs with correct outputs. During backpropagation the network adjusts its internal parameters whenever it makes an error. That way, it improves its accuracy over time (15).

Convolutional Neural Networks (CNNs) are a cornerstone of deep learning in image analysis (7). These models are designed to automatically and adaptively learn spatial hierarchies of features from images. CNNs consist of multiple layers of convolutions with learnable filters, and fully connected layers, making them adept at processing 2D and 3D images. In breast imaging, CNNs can analyze CEM images to identify patterns associated with breast cancer. The architecture of CNNs enables them to focus on local features within an image, such as the shape, size, and margins of lesions, facilitating detailed analyses (7).

Attention is a relatively novel technology that was widely evaluated in DL natural language processing (NLP) (16). In fact, the novel large language models (LLMs) such as GPT are based exactly on this mechanism (17). With image analysis algorithms, attention can improve their performance by allowing the models to focus on the most relevant features within an image (18). This is similar to the way that humans focus on specific parts of an image to extract information.

Multimodal Networks integrate and process data from multiple sources or types, for example text and images (**Figure 5**) (19). This approach is highly relevant with medical data analysis, as patient care relies on diverse data sources such as clinical history, lab tests, imaging etc.

Transfer Learning is a technique used widely in DL models training. With this approach, a model developed for a specific task is reused as the starting point for a model applied for a different task. It transfers knowledge from the previously learned context to the new application, potentially enabling to achieve better performance with less data (**Figure 6**) (20).

Generative Adversarial Network (GAN) is a type of DL model that is aimed at generating new images (21). The model is composed of a generator creating fake images from input data, and a discriminator attempting to distinguish between fake and real images. The process enables both networks to improve over time through a competitive process where the generator aims to fool the discriminator (**Figure 7**) (21).

Computer Vision Tasks

In medical imaging three foundational computer vision tasks play pivotal roles: classification, detection, and segmentation **(Figure 8**). Each of these tasks leverages DL algorithms to enhance diagnostic accuracy (7, 14).

Detection refers to the identification and localization of lesions within the images. For example, using DL algorithms to mark any abnormalities at CEM images with a circle or a box region of interest (ROI). This may aid radiologists in focusing their examination on areas of interest, potentially revealing lesions that may have been overlooked.

Segmentation is the process of delineating the exact boundaries of a lesion, separating it from the surrounding tissue. With manual segmentation, a radiologist traces the contours of a lesion in a CEM image, a time-consuming task with variability between observers. Automated segmentation on the other hand, uses DL to precisely outline the lesion, facilitating a more consistent and efficient analysis. This is also valuable in assessing the size and shape of a tumor, which are critical factors in diagnosis and treatment planning. U-Net is a commonly used CNN segmentation network in the medical domain (22).

Classification involves the categorization of lesions into predefined classes, such as benign or malignant. For instance, in CEM, a DL model might analyze an image to classify a detected lesion as BI-RADS 2 (benign, requires routine screening), BI-RADS 3 (probably benign, requires short-term follow-up) or as BI-RADS 4 (potentially malignant, and requires tissue sampling).

By integrating DL algorithms into these computer vision tasks, researchers and clinicians can significantly improve the diagnostic process in breast imaging.

Results

Sixteen studies have been included in this review, published between 2018 and 2024. While all the studies but one used CNN models, two of the CNN studies augmented the algorithm with attention layers.

All studies evaluated DL algorithms for classification of lesions at CEM, while six studies also assessed lesion detection or segmentation (**Figure 8**). In three studies segmentation was performed manually, two studies evaluated both manual and automatic segmentation, and nine studies automatically segmented the lesions. The objective assessment of the risk of bias based on QUADAS-2 is reported in **Table 1**. The results of the studies are summarized in **Table 2**.

Segmentation

Danala et al. (23) were the first to publish on DL analysis of CEM images. They developed MLP for lesion segmentation at CEM, and classification between malignant and benign lesions. They were the first to use automatic segmentation of lesions based on lesion enhancement in the subtracted contrast images. Then, based on these segmentations they analyzed both the subtracted and the low-energy images for lesion classification (23).

Wu et al. (24) combined DL models with a radiomics model. Radiomics involves the extraction of a large number of quantitative features from images. The authors first used a U-Net, which is a supervised DL network, for image segmentation. Training was performed on manually segmented images. Then, using another DL network (ResNet-18) they extracted features from the segmented tumor regions of interest in both the low-energy and the subtracted contrast images. They've also incorporated radiomics features to predict non-sentinel lymph node metastases. The radiomics model without DL overperformed the DL, and the combined model (24). Beuque et al. (25) developed a full-workflow model for automatic lesion identification, segmentation, and classification. A radiomics model was also trained to classify both manual and automatically segmented lesions. The DL model had the highest sensitivity for cancer classification, while the combined DL with radiomics model had higher specificity with the highest AUC (25).

Detection and classification

Gao et al. (26) developed a CNN model that classifies breast lesions based on features from both low-energy and recombined contrast-enhanced images. They have also applied a CNN model that learned the mapping between the low-energy and recombined images, to generate "virtual" contrast-enhanced images from standard digital mammography images. Using features from these "virtual" recombined images, the performance of the model in lesion classification improved, compared to incorporating only features from standard digital mammography (26). Qian et al. (27) have developed a multi-feature fusion neural network to combine information from both the low-energy and the contrast-enhanced images, for breast cancer detection.

In a recently published study, Helal et al. (28) also developed a Multiview DL model analyzing both low-energy and subtracted contrast images for breast lesions classification at

CEM studies and compared its performance to radiologists. The DL model had lower sensitivity for breast cancer detection compared to radiologists, while specificity for the model was higher (28). In 2021 Song et al. (29) evaluated a CNN that analyzed both the low-energy images and the subtracted contrast-enhanced images simultaneously to classify lesions. Later, in 2022, they've developed a model that applies GAN (**Figure 3**) used for fusion of the subtracted contrast-enhanced images and the low-energy images, and the classification model for lesion classification (30).

Perek et al. (31) applied a multimodal network for lesion classification in the subtracted CEM images, combining textual features with image analysis. They compared fine-tuning a pretrained neural network (AlexNet) and fully training a CNN. Their multimodal network resulted in a theoretical reduction in benign biopsies by up to 66% (31).

Dominique et al. (32) fine-tuned an existing CNN model that was used for chest X-rays analysis, on CEM dataset. They applied the model to predict histopathological characteristics of breast cancer from CEM images. For example, receptor status, grade and proliferative index. The model achieved the best results for estrogen receptor status prediction from the subtracted contrast-enhanced images (32).

Two studies that assessed the combination of an attention mechanism with CNN (33, 34). First, Li et al. (33) incorporated a CNN algorithm with attention for CEM images classification. Their network extracts information from all four images acquired for each breast (low energy and subtracted contrast images). In their study they report that this method significantly improved the accuracy of cancer classification and reduced false-positive cases compared to other DL models previously reported. Later, Mao et al. (34) also developed an attention-based DL model to discriminate benign from malignant breast lesions on CEM images. They've evaluated and compared between three CNNs with attention, and the same

models without attention. The attention-based models outperformed those without attention mechanism. They have also compared the best model performance to that of a radiomics model, as well as to radiologists. They show that the radiologists' performance improved with the algorithm's support (34).

Prospective data analysis validates the performance of DL models in real-world scenarios. It ensures these models are robust and applicable to future, unseen cases, beyond retrospective data. We found only two studies that evaluated their model on prospective data. Zheng et al. (35) developed a fully automated pipeline system based on CEM images for segmentation and classification of breast lesions. They assessed the algorithm as a support tool for breast radiologists, demonstrating improved sensitivity and specificity in a prospective testing cohort (35). Chen et al. (36) also developed a multi-process model for detection and classification of breast lesions at CEM. They have assessed their model in a retrospective external testing set as well as a prospective dataset. The model outperformed senior breast radiologists in lesion classification metrics (36).

Jailin et al. (37) developed a DL algorithm for detection and classification of breast lesions using data from both the low-energy and subtracted contrast images. They then evaluated their algorithm's performance at different levels of BPE (**Supplemental Figure 3**), showing increased false positive rates in cases with increased BPE (37).

Discussion

CEM is a novel imaging technique that has higher sensitivity compared to standard 2D mammography for detecting breast cancer, albeit with a decrease in specificity. We demonstrated that DL algorithms can be used for CEM lesion detection, segmentation and classification. Notably, some of the models have showed superior accuracy compared to radiologists' evaluations. Furthermore, integrating models as supportive tools for CEM

interpretation has refined specificity while keeping high sensitivity, and decreased the false-positive rate.

While the integration of DL for CEM analysis is still at a relatively early stage, there are studies that developed full-process applications that are able to detect, segment and classify breast lesions (25, 28, 35-37). Despite the promising results, currently, there is an absence of FDA-approved AI algorithms available for CEM (38).

Integrating DL algorithms into clinical workflows has the potential to enhance breast cancer screening and diagnosis. The algorithms can enable more accurate study interpretations in less time. Notably, the increased specificity reported for some algorithms highlights their utility in using these models as support tools for CEM studies interpretation. This could diminish the frequency of recalls and biopsies, while maintaining high sensitivity. Large, prospective clinical trials are necessary to assess the clinical value and real-world effectiveness of utilizing DL algorithms as a support tool for reading CEM studies. The ongoing prospective MASAI trial in Sweden is an example, assessing the application of DL in the interpretation of standard screening mammography (39).

The integration of AI into clinical workflows is anticipated to affect the role of breast radiologists. It may enable radiologists to dedicate more attention to complex cases, while spending less time on straightforward examinations. Radiologists may have higher availability for involvement in multidisciplinary teams. Initial automatic analysis of cases may allow triage and prioritization of examinations based on urgency. The transformation ultimately may streamline patient management, enhance the timeline of medical interventions, and promote a patient-focused approach in breast cancer care. Ultimately, implementation of these algorithms in everyday clinical use could improve patient outcomes and overall efficiency of healthcare delivery.

This review has several limitations. First, heterogeneity of studies and variability in measures between studies prevented a meta-analysis. Second, DL in radiology and CEM are both rapidly expanding topics. Thus, there may be relevant studies published after our review was performed. Finally, this review did not fully account for the variation in clinical settings and patient populations across the studies analyzed. This variability can affect the generalizability of the findings, as the performance of the algorithms could differ based on the study indications and patient demographics. Such differences might influence the algorithms' effectiveness and applicability in diverse contexts.

In conclusion, CEM is an emerging modality in breast imaging with a high diagnostic performance, outperforming standard 2D mammography and comparable to breast MRI. At present, the application of DL to CEM is at its beginning but showing promising performance. When used as a support tool, DL can potentially improve radiologists' accuracy and efficiency in the interpretation of CEM examinations, improving specificity while keeping high sensitivity for cancer detection. Further research is needed to validate DL models across different practices and CEM vendors, and prospectively assess their impact on real-world clinical practice.

References

1.   Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of Recall Rates with Sensitivity and Positive Predictive Values of Screening Mammography. American Journal of Roentgenology. 2001;177(3):543-9.

2.   Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. Jama. 1996;276(1):33-8.

3.   Kriege M, Brekelmans CTM, Obdeijn IM, Boetes C, Zonderland HM, Muller SH, et al. Factors Affecting Sensitivity and Specificity of Screening Mammography and MRI in Women with an Inherited Risk for Breast Cancer. Breast Cancer Research and Treatment. 2006;100(1):109-19.

4.   Jochelson MS, Lobbes MBI. Contrast-enhanced Mammography: State of the Art. Radiology. 2021;299(1):36-48.

5.   Ghaderi KF, Phillips J, Perry H, Lotfi P, Mehta TS. Contrast-enhanced Mammography: Current Applications and Future Directions. RadioGraphics. 2019;39(7):1907-20.

6.   Patel BK, Lobbes MBI, Lewin J. Contrast Enhanced Spectral Mammography: A Review. Seminars in Ultrasound, CT and MRI. 2018;39(1):70-9.

7.   Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. Radiology. 2019;290(3):590-606.

8.   Kinkar KK, Fields BKK, Yamashita MW, Varghese BA. Empowering breast cancer diagnosis and radiology practice: advances in artificial intelligence for contrast-enhanced mammography. Frontiers in Radiology. 2024;3.
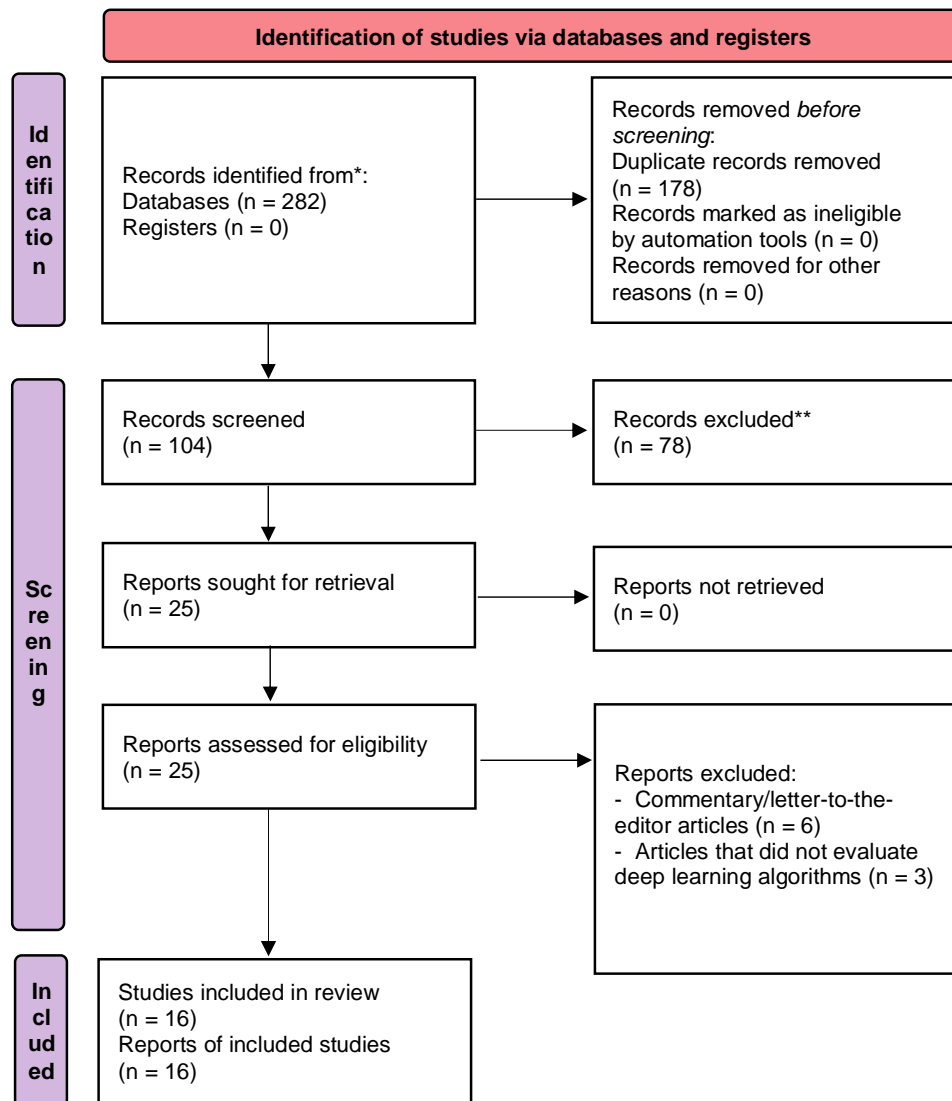
9.   Whiting PF. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Annals of Internal Medicine. 2011;155(8):529.

10.   Sorin V, Yagil Y, Yosepovich A, Shalmon A, Gotlieb M, Neiman OH, et al. Contrast-Enhanced Spectral Mammography in Women With Intermediate Breast Cancer Risk and Dense Breasts. American Journal of Roentgenology. 2018;211(5):W267-W74.

11.   Lee C, Phillips J, Sung J. ACR BI-RADS® In: ACR BI-RADS® CONTRAST ENHANCED MAMMOGRAPHY (CEM) A supplement to ACR BI-RADS® Mammography 2013 Atlas, Breast Imaging Reporting and Data System. American College of Radiology, Reston, VA, USA. 2022.

12.   Karimi Z, Phillips J, Slanetz P, Lotfi P, Dialani V, Karimova J, et al. Factors Associated With Background Parenchymal Enhancement on Contrast-Enhanced Mammography. American Journal of Roentgenology. 2021;216(2):340-8.

13.   Sorin V, Yagil Y, Shalmon A, Gotlieb M, Faermann R, Halshtok-Neiman O, et al. Background Parenchymal Enhancement at Contrast-Enhanced Spectral Mammography (CESM) as a Breast Cancer Risk Factor. Academic Radiology. 2020;27(9):1234-40.

14.   Klang E. Deep learning and medical imaging. J Thorac Dis. 2018;10(3):1325-8.

15.   Popescu M-C, Balas VE, Perescu-Popescu L, Mastorakis N. Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems. 2009;8(7):579-88.

16.   Sorin V, Barash Y, Konen E, Klang E. Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review. Journal of the American College of Radiology. 2020;17(5):639-48.

17.   Sorin V, Barash Y, Konen E, Klang E. Large language models for oncological applications. Journal of Cancer Research and Clinical Oncology. 2023;149(11):9505-8.

18.   Guo M-H, Xu T-X, Liu J-J, Liu Z-N, Jiang P-T, Mu T-J, et al. Attention mechanisms in computer vision: A survey. Computational Visual Media. 2022;8(3):331-68.

19.   Sorin V, Kapelushnik N, Hecht I, Zloto O, Glicksberg BS, Bufman H, et al. GPT-4 Multimodal Analysis on Ophthalmology Clinical Cases Including Text and Images. medRxiv. 2023:2023.11.24.23298953.

20.   Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Medical Imaging. 2022;22(1).

21.   Sorin V, Barash Y, Konen E, Klang E. Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs) – A Systematic Review. Academic Radiology. 2020;27(8):1175-85.

22.   Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. 2018;11045:3-11.

23.   Danala G, Patel B, Aghaei F, Heidari M, Li J, Wu T, et al. Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms. Annals of Biomedical Engineering. 2018;46(9):1419-31.

24.   Wu X, Guo Y, Sa Y, Song Y, Li X, Lv Y, et al. Contrast-Enhanced Spectral Mammography-Based Prediction of Non-Sentinel Lymph Node Metastasis and Axillary Tumor Burden in Patients With Breast Cancer. Frontiers in Oncology. 2022;12.

25.   Beuque MPL, Lobbes MBI, van Wijk Y, Widaatalla Y, Primakov S, Majer M, et al. Combining Deep Learning and Handcrafted Radiomics for Classification of Suspicious Lesions on Contrast-enhanced Mammograms. Radiology. 2023;307(5).

26.   Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, et al. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Computerized Medical Imaging and Graphics. 2018;70:53-62.

27.   Qian N, Jiang W, Guo Y, Zhu J, Qiu J, Yu H, et al. Breast cancer diagnosis from contrast-enhanced mammography using multi-feature fusion neural network. European Radiology. 2023;34(2):917-27.

28.   Helal M, Khaled R, Alfarghaly O, Mokhtar O, Elkorany A, Fahmy A, et al. Validation of artificial intelligence contrast mammography in diagnosis of breast cancer: Relationship to histopathological results. European Journal of Radiology. 2024;173:111392.

29.   Song J, Zheng Y, Zakir Ullah M, Wang J, Jiang Y, Xu C, et al. Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images. International Journal of Computer Assisted Radiology and Surgery. 2021;16(6):979-88.

30.   Song J, Zheng Y, Xu C, Zou Z, Ding G, Huang W. Improving the classification ability of network utilizing fusion technique in contrast enhanced spectral mammography. Medical Physics. 2021;49(2):966-77.

31.   Perek S, Kiryati N, Zimmerman-Moreno G, Sklair-Levy M, Konen E, Mayer A. Classification of contrast-enhanced spectral mammography (CESM) images. International Journal of Computer Assisted Radiology and Surgery. 2018;14(2):249-57.

32.   Dominique C, Callonnec F, Berghian A, Defta D, Vera P, Modzelewski R, et al. Deep learning analysis of contrast-enhanced spectral mammography to determine histoprognostic factors of malignant breast tumours. European Radiology. 2022;32(7):4834-44.

33.   Li X, Cui J, Song J, Jia M, Zou Z, Ding G, et al. Contextual Features and Information Bottleneck-Based Multi-Input Network for Breast Cancer Classification from Contrast-Enhanced Spectral Mammography. Diagnostics. 2022;12(12):3133.

34.   Mao N, Zhang H, Dai Y, Li Q, Lin F, Gao J, et al. Attention-based deep learning for breast lesions classification on contrast enhanced spectral mammography: a multicentre study. British Journal of Cancer. 2022;128(5):793-804.

35. Zheng T, Lin F, Li X, Chu T, Gao J, Zhang S, et al. Deep learning-enabled fully automated pipeline system for segmentation and classification of single-mass breast lesions using contrast-enhanced mammography: a prospective, multicentre study. eClinicalMedicine. 2023;58:101913.

36. Chen Y, Hua Z, Lin F, Zheng T, Zhou H, Zhang S, et al. Detection and classification of breast lesions using multiple information on contrast-enhanced mammography by a multiprocess deep-learning system: A multicenter study. Chinese Journal of Cancer Research. 2023;35(4):408-23.

37. Jailin C, Mohamed S, Iordache R, Milioni De Carvalho P, Ahmed SY, Abdel Sattar EA, et al. AI-Based Cancer Detection Model for Contrast-Enhanced Mammography. Bioengineering. 2023;10(8):974.

38. AI Central. American College of Radiology Data Science Institute. https://aicentral.acrdsi.org. Accessed April 12, 2024.

39. Lång K, Josefsson V, Larsson A-M, Larsson S, Högberg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. The Lancet Oncology. 2023;24(8):936-44.

40. Khaled R, Helal M, Alfarghaly O, Mokhtar O, Elkorany A, El Kassas H, et al. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. Scientific Data. 2022;9(1).

Figure 1. Flow Diagram of the inclusion process

Flow diagram of the search and inclusion process based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Figure 2. Contrast-enhanced mammography. An example of a contrast-enhanced mammography examination in a 45-year old woman. Mediolateral oblique low-energy (A) and subtracted contrast-enhanced (B) images of both breasts show multiple enhancing masses in her left breast. Ultrasound-guided biopsies were performed, confirming multifocal invasive ductal carcinoma (IDC).
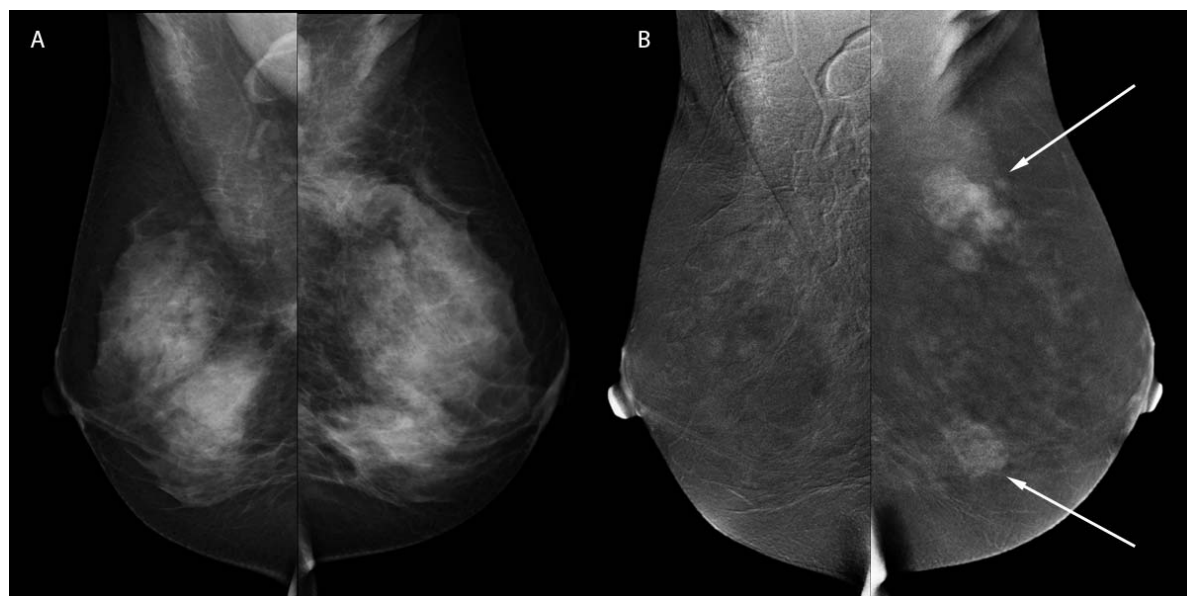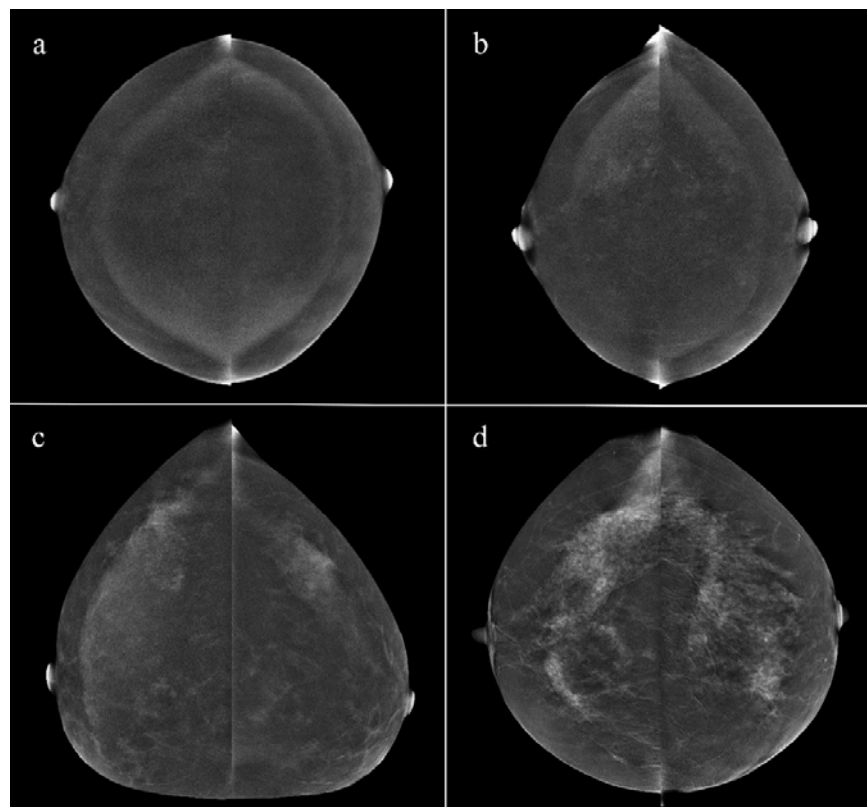
Figure 3. Background parenchymal enhancement (BPE) at contrast-enhanced mammography. Examples illustrating BPE grading at contrast-enhanced mammography: (a) Minimal BPE, < 25%; (b) Mild BPE, 25-50%; (c) Moderate BPE, 50-75%; (d) Marked BPE, > 75%.

Reprinted with permission from Sorin et al. *Background Parenchymal Enhancement at Contrast-Enhanced Spectral Mammography (CESM) as a Breast Cancer Risk Factor*. Academic Radiology, 2020.

Figure 4. Venn diagram of artificial intelligence frameworks. This diagram illustrates that artificial intelligence (AI) is the broadest category, encompassing machine learning and deep learning as its subsets.
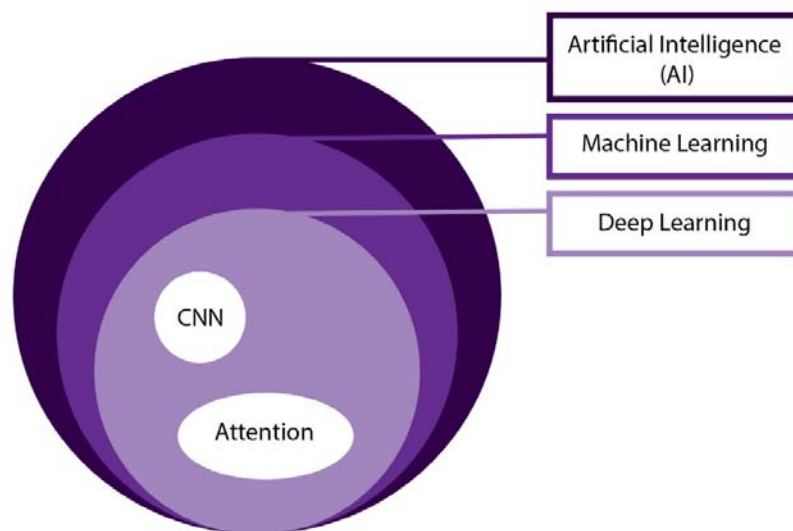
Figure 5. An illustration of a multimodal deep learning model. This diagram demonstrates the integration of a clinical note (textual data) and contrast-enhanced mammography images (visual data) within a neural network architecture. These are processed through distinct pathways before being merged in a unified framework, facilitating a comprehensive analysis in the context of patient diagnosis.
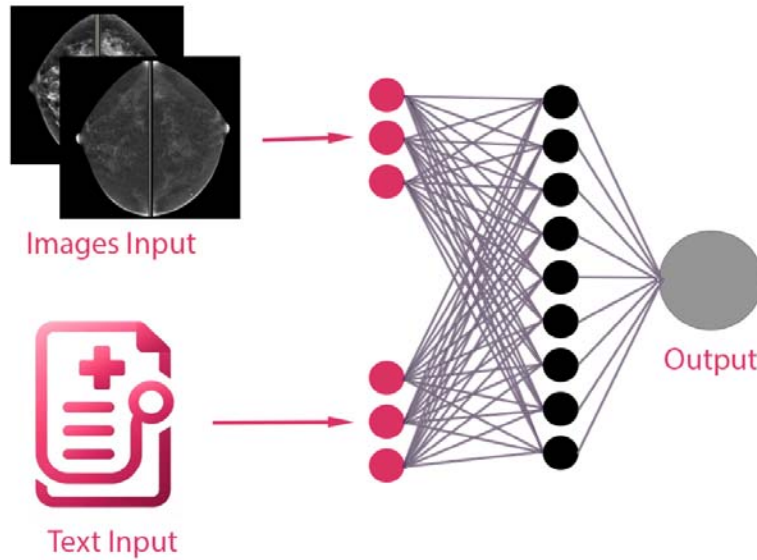
Figure 6. Illustration of transfer learning. This diagram displays the application of transfer learning between two neural network models. Initially, a chest X-ray image is processed by the first model, which is used for lung cancer classification. An arrow then signifies the transfer of learned features (in purple) to a second model, which processes a contrast-enhanced mammography (CEM) image. This second model is adapted to the characteristics of the CEM data, utilizing modified or additional components (in pink) for cancer classification at CEM.
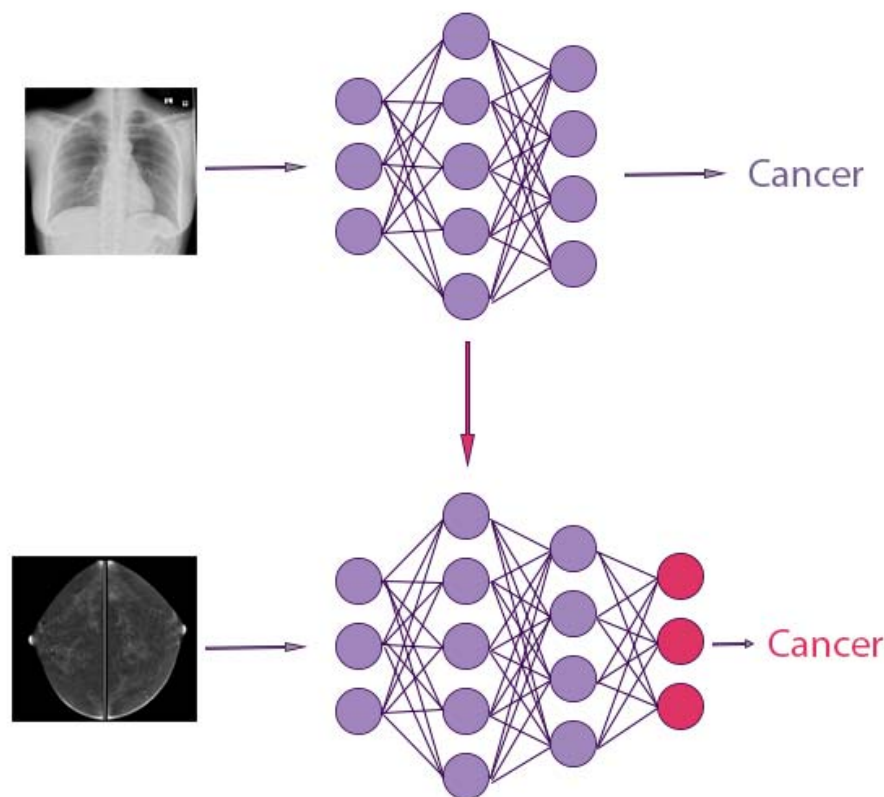
Figure 7. Schematic model of a generative adversarial network (GAN). The diagram illustrates how the generator and discriminator components are trained simultaneously. Starting with random noise as input, the generator produces synthetic images. These images, along with actual images, are supplied to the discriminator. The discriminator assigns a probability indicating whether the image is real. Based on feedback from the discriminator the generator then adjusts its parameters to enhance the realism of the synthetic images.
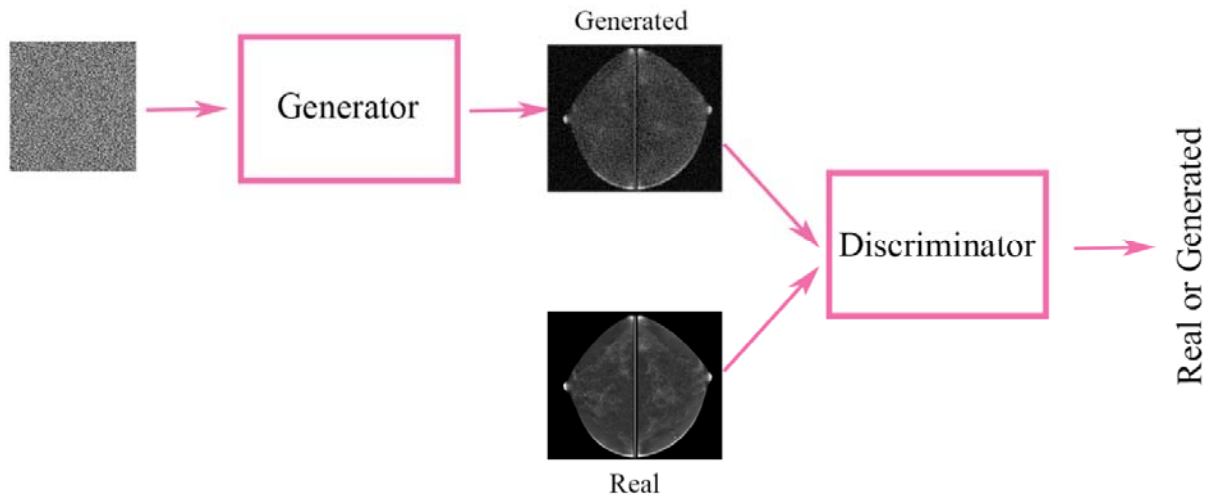
Figure 8. Image analysis tasks in medical imaging. This figure illustrates three tasks applied to contrast-enhanced mammography (CEM) images: lesion detection, classification, and segmentation.
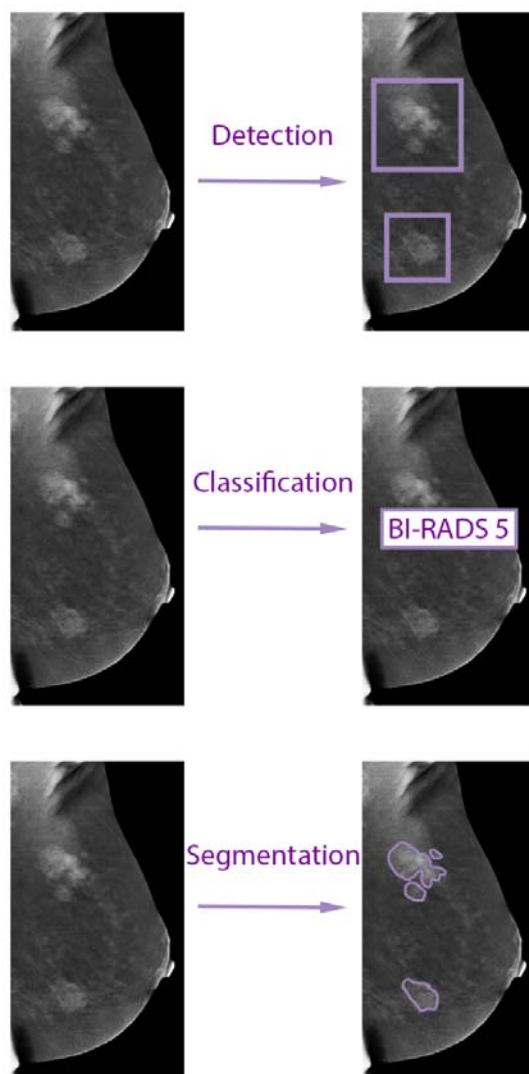
Table 1: Quality Assessment of Diagnostic Accuracy Studies-2 (QUADS-2)

🔴 = high risk of bias; 🟢 = low risk of bias. ❓ = unclear risk. N/A – not applicable

	RISK OF BIAS				APPLICABILITY CONCERNS		
First Author	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Helal et al.	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Chen et al.	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Jailin et al.	🔴	🟢	🟢	🟢	❓	🟢	🟢
Qian et al.	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Beuque et al.	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Zheng et al.	🟢	🟢	🟢	🟢	🔴	🟢	🟢
Li et al.	🔴	🟢	🟢	🟢	❓	🟢	🟢
Mao et al.	🟢	🟢	🟢	🟢	🔴	🟢	🟢
Wu et al.	🟢	🟢	🟢	🟢	🔴	🟢	🟢
Dominique et al.	🟢	🟢	🟢	🟢	🔴	🟢	🟢
Khaled et al.	🔴	🔴	🟢	🟢	❓	❓	🟢
Song et al.	🔴	🟢	🟢	🟢	❓	🟢	🟢
Song et al.	🔴	🔴	🟢	🟢	❓	❓	🟢
Perek et al.	🔴	🔴	🟢	🟢	❓	❓	🟢
Gao et al.	🔴	🔴	🟢	🟢	❓	❓	🟢
Danala et al.	🔴	🔴	🟢	🟢	❓	❓	🟢

	Study (Ref)	Year	Study Design	Neural Network	Application	Training Dataset (n = patients)	Testing Dataset (n=patients)	Segmentation	Performance Metrics
1	Helal et al. (28)	2024	Retrospective	CNN	Lesion detection and classification	326 (training)	37	Automatic	AUC-ROC 0.936; (95% CI: 0.898, 0.973). Sensitivity 75%, specificity 96.3%, total accuracy of 90.1%, PPV 87.1 %, and NPV 92%
2	Chen et al. (36)	2023	Retrospective and prospective	CNN	Lesions detection and classification	1355 (training) and 352 (validation)	95 (retrospective) and 101 (prospective)	Automatic	**Lesion detection**: AFROC 0.953 (external) and 0.963 (prospective). **Lesion classification**: AUC 0.909 (external, 95% CI: 0.822–0.996), and 0.912 (prospective, 95% CI: 0.840–0.985).
3	Jailin et al. (37)	2023	Retrospective	YOLOv5 (CNN)	Lesions detection and classification	787 (training) and 150 (validation)	150	Automatic	**Lesion detection:** AUFROC 0.891 and 0.712 for low and high BPE, respectively. **Lesion classification**: AUC 0.986 (95% CI: 0.976–0.995)

									and 0.919 (95% CI: 0.866–0.962) for low and high BPE, respectively.
4	Qian et al. (27)	2023	Retrospective	CNN	Classification	1718 (training) and 255 (validation)	523	No segmentation was performed	AUC of 0.92 (95% CI: 0.89, 0.95)
5	Beuque et al. (25)	2023	Retrospective	CNN	Lesion detection, segmentation, and classification	850 (training) and 212 (validation)	279	Manual and automatic	**Lesion detection:** Accuracy 88%, sensitivity 99%. **Segmentation:** Dice coefficient 0.80 **Classification: Manual segmentation:** AUC 0.88 (95% CI: 0.84-0.93). Sensitivity 89% (95%CI 85-93%), specificity 73% (95% CI 64-81%). **Automatic segmentation:** AUC 0.87 (95% CI: 0.82-0.92). Sensitivity 100% 6(95%CI 100-100%), specificity 45% (95% CI 34-57%).
6	Zheng et al. (35)	2023	Retrospective and	CNN	Segmentation and	1538	172 (internal testing), 37	Automatic	**Segmentation**: Dice coefficient

			prospective		classification		(external testing), 105 (prospective)		0.820 ± 0.148 (external retrospective) and 0.837 ± 0.132 (prospective). **Classification:** AUC: 0.940 (external retrospective, 95% CI: 0.894–0.987) and 0.891 (prospective, 95% CI: 0.816–0.945).
7	Li et al. (33)	2022	Retrospective	CNN + Attention	Classification	98 (training) and 12 (validation)	12	Automatic	Accuracy 88.06%, Precision 88.03%, Sensitivity 88.1%, specificity 88.01%, F1-score 0.88
8	Mao et al. (34)	2023	Retrospective	CNN + Attention	Classification	1093	146	Manual and automatic	AUC of 0.970, an AUPRC of 0.988, F1 score 0.918, accuracy 0.891, sensitivity 0.848, specificity 1.000, PPV 1.000, and NPV 0.722
9	Wu et al. (24)	2022	Retrospective	CNN	Lesion segmentation and lymph nodes classification	For segmentation 177, for classification 120	For segmentation 20, for classification 31	Automatic	**Segmentation**: Dice coefficient 0.84 ± 0.10 **Classification:** Deep learning model: AUC 0.53 (95% CI 0.31-

								0.75), deep learning model with radiomics features score: AUC 0.76 (95% CI 0.59-0.93).
10	Dominique et al. (32)	2022	Retrospective	CNN	Tumor characteristics prediction (classification)	1574 images for training and 390 images for validation (no details on the number of patients in each group)	496 images (no details on the number of patients in the testing group)	Manual	Tumor grade AUC 0.611, sensitivity 60.34%, specificity 60.26%, accuracy 60.28%; ER AUC 0.858, sensitivity 80.23%, specificity 80.3%, accuracy 80.24%; PR AUC 0.615, sensitivity 80.23%, specificity 80.3%, accuracy 80.24%; HER2 AUC 0.62, sensitivity 60%, specificity 54.82%, accuracy 55.44%; Ki67 AUC 0.593, sensitivity 55.97%, specificity 56.14%, accuracy 56.05%; TN AUC 0.876, sensitivity 72.22%, specificity 73.26%, accuracy 73.19%
11	Khaled et al. (40)	2022	Retrospective	CNN	Segmentation	Overall 2006 images from 326 patients		Automatic	IOU 0.65, F1 0.71

12	Song et al. (30)	2022	Retrospective	CNN	Classification (per image, benign vs. malignant)	Overall images from 95 patients were included. Each breast was included as a single case and 66 benign cases were excluded. In the preprocessing of images, images were flipped, and noise were added, resulting ultimately in 17,360 images that were divided randomly to 80% training, 15% validation and 5% testing.	Automatic	Accuracy 96.6%, sensitivity 96.4%, specificity 96.4%, precision 96.8%, F1 score 0.966, AUC 0.966
13	Song et al. (29)	2021	Retrospective	CNN	Classification	95 patients	Automatic	Accuracy 94.78%, precision 95.02@, recall 95.91%, specificity 94.5%, F1 score 0.955, AUC 0.947
14	Perek et al. (31)	2019	Retrospective	CNN	Classification	54 patients with 129 images used for validation.	Manual	For 100% sensitivity specificity was up to 66%, depending on the model
15	Gao et al. (26)	2018	Retrospective	CNN	Classification	49 patients (training and testing unclear).	Manual	Accuracy 89%, sensitivity 93%, specificity 86%, and AUC 0.91
16	Danala et al. (23)	2018	Retrospective	MLP	Classification	111 patients	Automatic	Accuracy 68.5%, PPV 82.1%, NPV 47.7%, AUC 0.737±0.048

Table 2. Studies Evaluating Deep Learning Applications for Contrast Enhanced Mammography