International Journal of Multidisciplinary Research and Growth Evaluation.

BI-RADS category prediction from mammography images and mammography radiology reports using deep learning: A systematic review

Hassan Tanveer [1*], **Nimra Batool** [2], **Namoos Zahra** [3]
[1] DePaul University, USA
[2] The University of Lahore, Pakistan
[3] University of Engineering and Technology, Pakistan

* Corresponding Author: **Hassan Tanveer**

Article Info

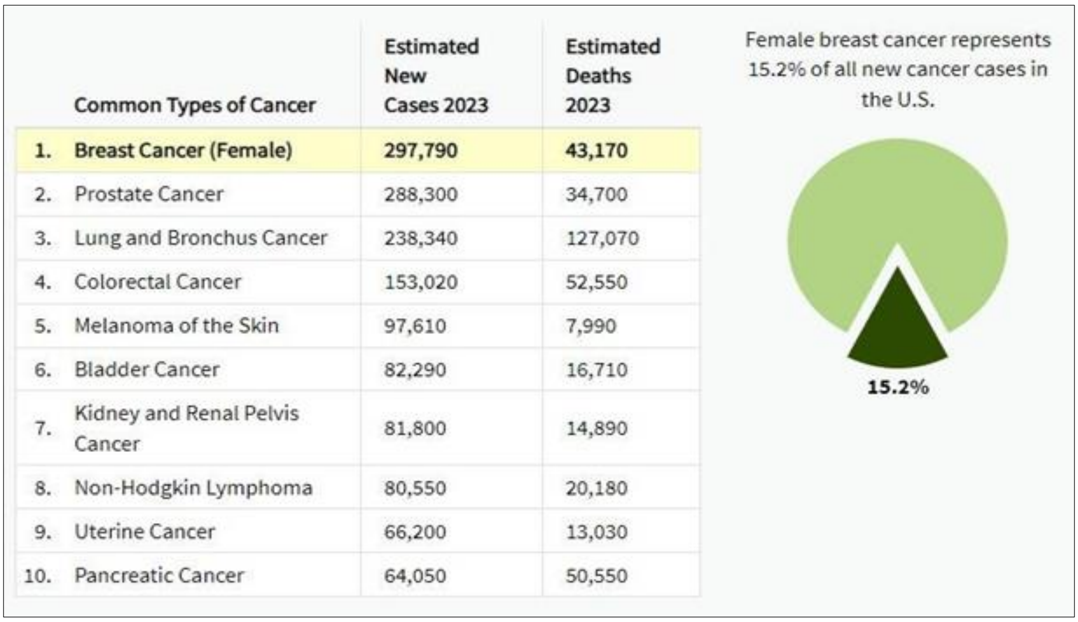
ISSN (online): 2582-7138
Volume: 06
Issue: 01
January-February 2025
Received: 15-11-2024
Accepted: 18-12-2024
Page No: 890-902

Abstract
Breast cancer is the most prevalent cancer globally and a leading cause of cancer-related deaths, with over 2.3 million new cases reported annually. It is the leading cancer in women and also significantly affects men. Early detection through routine mammography is critical, as it significantly reduces mortality. The Breast Imaging Reporting and Data System (BI-RADS) is a standard classification system used to assess mammography findings, categorizing lesions based on their likelihood of malignancy. Recent advancements in deep learning and computer-aided detection (CADe) systems have improved BI-RADS classification, aiding radiologists in identifying suspicious findings more effectively. This review explores the application of deep learning, particularly convolutional neural networks (CNNs), for BI-RADS category prediction. It discusses the strengths and limitations of existing models, highlighting the use of public datasets and the integration of mammography images and radiology reports. Additionally, it suggests a novel multi-modal approach for more accurate predictions, offering insights into the future of breast cancer detection and classification.

Keywords: Breast cancer, Mammography, BI-RADS, Deep learning, Convolutional Neural Networks (CNN)

1. Introduction

The most prevalent type of cancer worldwide, breast cancer is responsible for a significant portion of cancer- related deaths worldwide [1]. As the most common cancer diagnosed worldwide, breast cancer has surpassed lung cancer to account for 1 in 8 cancer-type diagnoses and 2.3 million new cases in both sexes i.e., male and female combined [2]. Furthermore, breast cancer ranks fifth globally in terms of deaths from cancer in females and is a major cause of cancer-related mortalities in males [2, 3]. In 2020, an estimated 685,000 women lost their lives to breast cancer, making up 16% of all female cancer deaths [4]. Following a period of insufficient public health response to this development, the World Health Organization (WHO) recently launched the Global Breast Cancer Initiative [4]. The prevalence of breast cancer varies across nations, with high-income nations typically having a higher incidence rate than low- and middle-income nations [5]. However, mortality also varies across different regions in the world [4, 5]. Numerous factors, including race, ageing, genetic changes, family history, exposure to chest radiation, and obesity, are thought to elevate the risk of developing breast cancer [6, 7]. Breast cancer mortality and financial burden can be significantly decreased with early diagnosis and treatment [8]. In most cases, routine mammography screenings along with physical exams and, if needed, additional imaging tests like ultrasound or MRI can lead to an early and successful diagnosis of breast cancer [9]. Consequently, routine mammography screening, which can identify various abnormalities in the breast even before signs appear, to determine the risk of most prevalent cancer is advised by clinical guidelines [8, 9]. The cranio-caudal (CC) and mediolateral-oblique (MLO) views of each woman's breast, or the LCC, RCC, LMLO, and RMLO views combined, are the two views that are part of routine screening mammography [10].



	Common Types of Cancer	Estimated New Cases 2023	Estimated Deaths 2023
1.	Breast Cancer (Female)	297,790	43,170
2.	Prostate Cancer	288,300	34,700
3.	Lung and Bronchus Cancer	238,340	127,070
4.	Colorectal Cancer	153,020	52,550
5.	Melanoma of the Skin	97,610	7,990
6.	Bladder Cancer	82,290	16,710
7.	Kidney and Renal Pelvis Cancer	81,800	14,890
8.	Non-Hodgkin Lymphoma	80,550	20,180
9.	Uterine Cancer	66,200	13,030
10.	Pancreatic Cancer	64,050	50,550

Fig 1: Breast Cancer statistics (as the most common cancer) in USA (https://seer.cancer.gov/statfacts/html/breast.html)

Breast Imaging Reporting and Data System (BI-RADS) was created by the American College of Radiology (ACR) to lessen discrepancies and standardize risk assessment in radiologists' reports of mammography findings [11]. The initial version that was suggested contained the final assessment category along with management recommendations, the lexicon for findings from mammography imaging, and a suggested format for a mammography report [12]. This lexicon of descriptors correlated with high predictive values linked to either benign or malignant disease through scientific analysis and literature review [12, 13]. The categorization of the imaging findings for the overall assessment was the second crucial component of the BI-RADS system as given in Table I [13]. The classification gives a lesion an approximate risk of malignancy ranging from nearly zero to more than 95% [11, 13]. The recommendations' level of ambiguity was reduced by the classification and final evaluation [14]. Six classifications for lesions were included in the most recent edition, BI-RADS 5 (2013) [15].

Table 1: Description of BI-RADS categories for mammograms and the likelihood of cancer associated with each category

BI-RADS	Definition	Likelihood of Cancer (%)
0	Incomplete; further imaging analysis is required	-
1	Normal	0
2	Benign	0
3	Probably benign	>0 to ≤ 2
4A	Minimal evidence of malignancy	>2 to ≤ 10
4B	Mildly suspicious	>10 to ≤ 50
4C	High suspicion	>50 to < 95
5	Highly suggestive of malignancy	≥ 95
6	Known biopsy-proven malignancy	100

Breast cancer screening often makes use of the BI-RADS prediction. Radiologists must put in a great deal of work to interpret screening mammograms. To increase the effectiveness of mammography interpretation, numerous computer-aided detection (CADe) systems have been developed for the effective and trustworthy BI- RADS classification. CADe primarily aids in the positioning and identification of any suspicious findings that show up in medical images, leaving the radiologist to interpret these findings. In an effort to develop more efficient computer-aided detection systems (CADe) for breast cancer, a number of studies for breast cancer detection and classification were proposed with the notable advancements in deep learning and image processing techniques.

This review provides an overview of the important and well-known methods that have been introduced in the field of DL and CNNs for BI-RADS categories classification. Additionally, the paper shows how the models that have been developed over the last ten years have progressed. In addition to outlining the present difficulties, the paper discusses the shortcomings of the models that have been suggested in the literature for the classification of BI-RADS categories from mammography images and mammography radiology reports. The various public mammography datasets, mammography reports, and screening modalities that are currently in use are highlighted in this paper. The paper also emphasizes highlighting the advantages and disadvantages of the traditional computer aided detection systems and deep learning-based CADe systems. Finally, it suggests a novel approach to predict the BI- RADS categories from mammography images and radiology reports using a multi-modal approach.

This review answers the following questions:

RQ1: Which deep-learning algorithms are being most commonly used for BI-RADS category prediction?

RQ2: What are the common approaches for pre-processing

of mammography images and radiology reports before putting them into deep learning models?
RQ3: Which image features are most informative for predicting BI-RADS categories in mammography images?
RQ4: How are textual features extracted from mammography radiology reports and utilized in the prediction of BI-RADS category?
RQ5: Which public mammography datasets are available, and which algorithms have proven out to be best?
RQ6: How mammography images and radiology reports can be combined for the prediction of BI- RADS category?

A. Topology of Review
Section 2 describes the recent review studies conducted on breast cancer detection using deep learning approached. Section 3 provides the survey methodology. Section 4 discusses the research findings describing the architecture of CNNs and the fundamental ideas used to train them. Section 5 presents discussion for the validity of these algorithms. Section 6 discusses the benefits and covers the drawbacks of the CNN-based techniques. Section 7 concludes the review. Figure 2. presents the topology for the review.

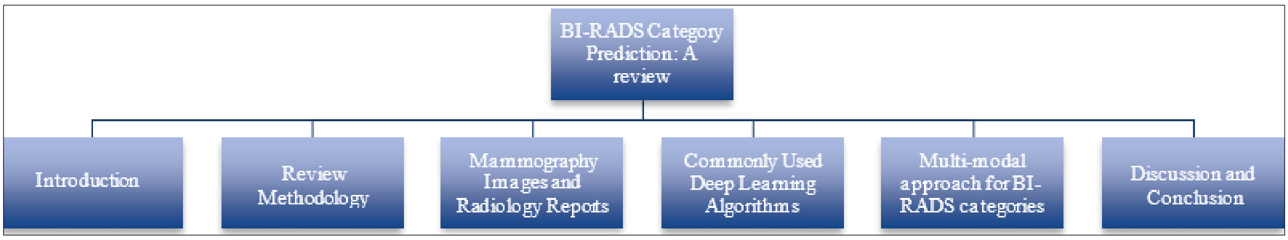


Fig 2: Organization of the Review

B. Breast Imaging Modalities
A variety of imaging modalities, including digital breast tomosynthesis (DBT), magnetic resonance imaging (MRI), mammography (MG), and ultrasound, are used to screen breasts [16].
Chest X-ray called mammography is used to look at the breast in order to find breast cancer and other diseases early on [17]. It serves as a screening and diagnostic tool. A radiologist will closely inspect a mammogram to look for areas of unusual configuration or high-density areas that deviate from normal tissue appearance [18]. These regions may be indicative of a wide range of abnormalities, such as complex cysts, fibroadenomas, benign tumors, or cancerous tumors [19]. When examining an abnormal region, radiologists consider its size, shape, contrast, and appearance around its edges or margins, as these factors can all point to the possibility of malignancy, or cancer [20]. Strong magnets and radio waves are the primary tools used in MRI, which creates comprehensive images of the inside of the breast [21, 22]. This modality is thought to be beneficial for women who are at an increased peril of breast cancer [23, 24].
Sound waves are used in ultrasound to create images of the breast's internal structure [25]. It is used for pregnant women who shouldn't be exposed to the x-ray radiations used in mammogram or for women who are at an increased risk for breast cancer but cannot go through a chest MRI [26, 27]. Ultrasonography is also frequently used to screen women with dense breast tissue [28]. Digital breast tomosynthesis, an X-ray mammography technique, in which multiple low-dose projection images obtained by moving the X-ray tube in an arc over a limited angular range are used to reconstruct tomographic images of the breast [29]. Despite the fact that the principles of tomographic imaging were developed in the 1930s, it took several decades for the clinical applications of tomosynthesis in mammography to emerge. This was because of advancements in reconstruction and post-processing algorithms, rapid computer processing, and the development of flat-panel electronic display detectors [30].

2. Literature review
In this section of the survey, few review studies have been mentioned on the usage of deep convolutional neural networks in mammography along with their pros and cons. However, no review study has been conducted on the applications of deep learning algorithms for the prediction of BI-RADS category. Abdelrahman *et al*. conducted a survey on convolutional network-based computer vision models for mammography as well as more recently developed computer assisted detection (CAD). The current literature on CNNs for four different mammography tasks i.e., mass detection and classification, calcification detection, breast asymmetry prediction, and breast density classification were presented and discussed in the survey. It included comparing and presenting the reported results for each task as well as the pros and cons of the various CNN-based approaches. It does not, however, address the topic of BI-RADS category classification with deep learning techniques [31].
Hassan *et al*. presented a survey on deep learning-based CADe systems for mass detection and classification in mammography in an organized manner [32]. In addition to providing a dataset-based comparison of the most recent techniques and the most popular evaluation metrics for the breast cancer CADe systems, the review presents the publicly available mammographic datasets as of right now [33]. The survey highlights the benefits and drawbacks of the existing literature while discussing it. The survey emphasizes the shortcomings and difficulties in the existing methods for classifying and detecting breast cancer but it was unable to discuss the BI-RADS category prediction on the basis of mass, calcification, and asymmetry [32].
Tan *et al*. conducted an overview of recent research on the use of CNNs in mammography images [33]. First, models built using one of the most significant deep learning algorithms i.e., convolutional neural networks are presented. The most recent papers are then examined for four different mammography applications: 1) Classification of breast density; 2) Asymmetry detection and classification; 3) Classification of mass; and 4) Classification of calcification. The article also covers the FDA-approved models that are used and addresses real-world applications of the algorithms that are discussed. Lastly, a list of open research challenges for CNN-based techniques to enhance breast cancer detection is provided [34].
Cè *et al*. provided an overview of the most significant AI

applications in breast malignancy imaging while examining prospective obstacles and fresh viewpoints associated with the general implementation of these innovative instruments [35]. Several studies support the idea that women, radiologists, and healthcare systems could all benefit greatly from the appropriate integration of AI into current clinical workflows [36]. The AI-based strategy has shown to be especially helpful for creating new risk prediction models that combine several data streams to plan customized screening procedures. AI models may also assist radiologists in the pre-screening and lesion detection stages, improving diagnostic precision and lowering workload and overdiagnosis-related complications. To plan a targeted treatment, radio-genomics and radiomics techniques could extrapolate the tumor's so-called imaging signature [37, 38].

Gao *et al*. provided insights into potential future development directions and specifically highlighted the most recent developments in DL techniques for mammography image analysis [39]. There are inherent deviations in the clinical explication and evaluation of the breast images because they frequently entail high wage costs and largely rely on the experience of clinicians. As a result, artificial intelligence (AI) has become a useful tool for diagnosing breast cancer. Deep Learning (DL) and Machine Learning (ML) are two aspects of artificial intelligence [40]. Both of them assist in lesion localization, lower the rate of incorrect diagnoses, and increase accuracy by modelling human behavior to learn from and process data. Using the conventional algorithms, this narrative review offered a thorough analysis of the state of mammography research today.

3. Methodolgy

The steps that we adopted to conduct this review are as follows:

A. Articles Collection

To ensure a top-notch review of the literature on BI-RADS category prediction from mammography images and radiology reports using deep learning, a number of protocols were followed. A comprehensive search of the peer-reviewed literature was conducted in January 2024; short papers, reports, editorials, posters, and dissertations were not included. The guidelines for Preferred Reporting Items for Systematic Reviews and Meta- Analyses (PRISMA) were taken into account. The following terms were used to extract all of the articles: BI- RADS, breast density, laterality of breasts, mammograms, mammography images, breast radiology report, deep learning, neural networks, prediction, and classification from Web of Science, PubMed, Google Scholar, MDPI, Elsevier, and IEEE- Xplore. During the search process, 1149 peer-reviewed publications were discovered. Only BI-RADS categories were targeted, and articles published after 2014 were included in the literature selection process for this study.

B. Search Strategy

Clearly defining inclusion and exclusion criteria is important because they will be used to evaluate the literature review's overall validity during the selection process [41]. We employed the quality standards listed below, which drew inspiration from pertinent research [41]. Therefore, studies concentrating on deep learning- based prediction of BI-RADS categories qualified for inclusion. Using the specified criteria of the selection process, the papers were assessed first on the basis of their titles, then on the basis of their abstracts, and finally on the basis of their full texts. Figure 3 illustrates the overall plan. The following are the quality standards that were taken into account when deciding which research articles to include:

1. Articles released in the previous ten years were comprised.
2. The research that looked into the application of deep learning and natural language processing for BI- RADS category prediction
3. Studies that provided a thorough explanation of the feature extraction, fusion, data preprocessing, and architecture of the deep learning and machine learning models in use
4. Research that discussed the quantifiable results pertaining to specificity, accuracy, sensitivity, and AUC/ROC.
5. To guarantee credibility and quality, only peer-reviewed journals and conferences were included.

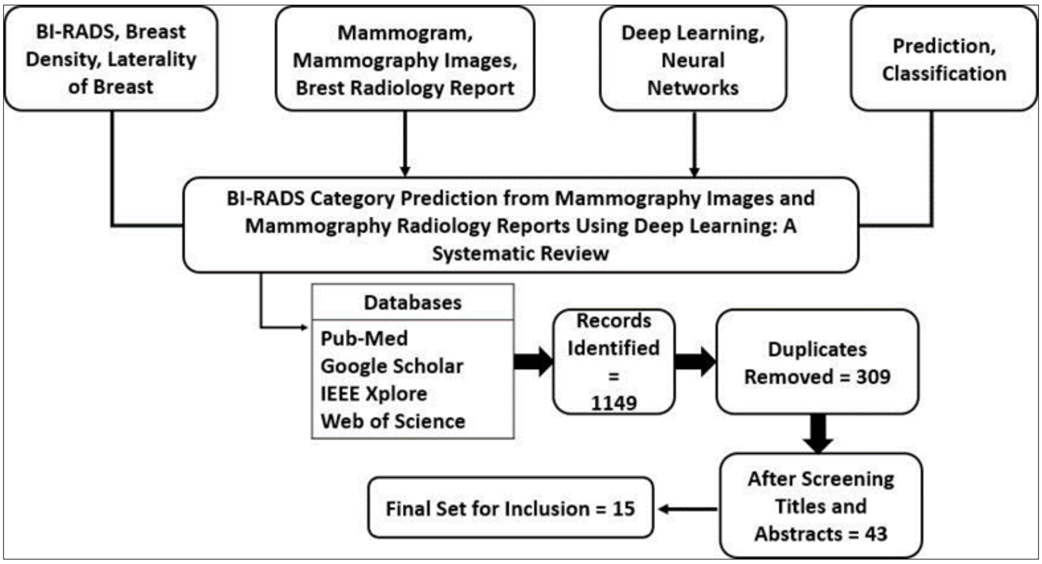


Fig 3: Strategy for articles screening and inclusion

3. Research findings

The research findings from this review are mentioned below:

A. Conventional Computer Aided Detection (CADe) Systems

The 1990s saw the introduction of CADe systems by researchers, who used them to locate aberrant or suspicious areas in medical images and notify medical professionals for follow-up [42]. These systems were based on traditional machine learning models. The main objective of CAD is to lower the false- negative rate which could be caused by fatigue or mistakes on the part of observers while raising the detection rate of diseased regions.

The CADe uses various forms of clinical medical imaging to screen and detect breast cancer. Images include magnetic resonance imaging (MRI), ultrasounds, and mammograms based on X-rays. We have selected mammography as a screening tool for in-depth review in this article. Mammography screening has made early cancer detection more effective, resulting in 500 fewer deaths from cancer for every 100,000 women screened and a 91% 5-year survival rate in USA [43]. When medical professionals discover abnormalities in a mammography, they may request additional imaging tests, like an MRI [44]. Doctors recommend a more intrusive histological image analysis for serious patients or the patients with visible masses on their radiographs [44]. In contrast, mammography is a minimally invasive procedure. Mammographic analysis, as a first-line screening technique, should be greatly enhanced with CADe techniques, especially for BI-RADS category prediction.

The conventional framework for a CADe system is: candidate area extraction (Figure 4. (a)); using image enhancement techniques which frequently rely on a comprehensive set of handwrought features, that experts spend hours extracting, candidate detection (Figure 4. (b)); experts create a statistical or morphological feature set to represent the potential regions, candidate classification (Figure 4. (c)); a statistical classifier generates an output, or prediction, of a disease state based on the engineered features. CNN-based techniques have recently replaced conventional CADe in research [45]. Support Vector Machine (SVM), Random Forest (RF), and other statistical techniques were common prior to the rise in popularity of deep Learning and CNNs [46, 47]. However, their dependence on laborious feature engineering has driven investigators towards convolutional neural network-based techniques, or hybrid techniques that combine statistical algorithms with convolutional neural network models for automatic feature extraction [48].

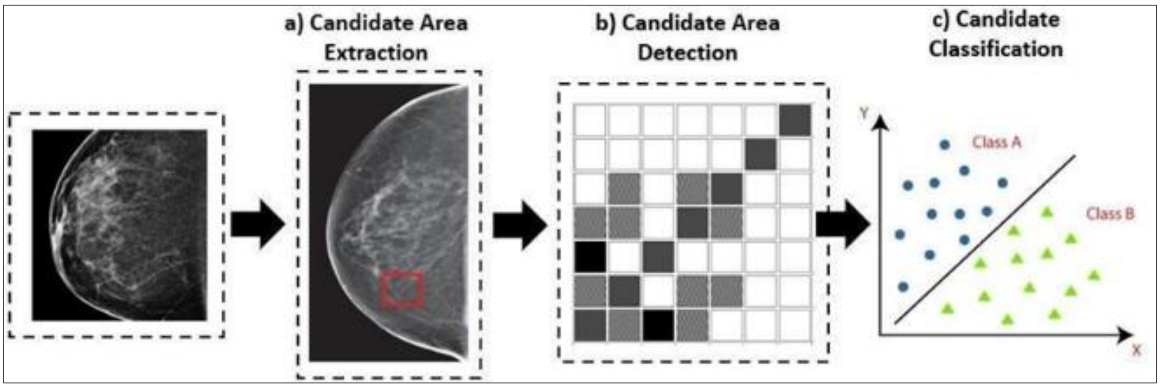


Fig 4: This figure shows the conventional CADe framework. (a) The images are annotated by researchers. (b) Relevant features are chosen using other filtering techniques and correlation matrices. (c) A statistical model draws conclusions

B. Conventional CADe Systems in Mammography Screening

Researchers attempted to address difficult mammography prediction tasks, such as classifying tissue density and identifying and classifying asymmetries, calcifications, and masses, using various datasets. Various convolutional neural networks which have been developed for mass detection and class prediction, calcification detection and class prediction, breast asymmetry detection and classification, and breast density classification have been discussed here.

1. Classification of Breast Density

The characteristics of the breast tissue as seen in a mammography image have been referred to as breast density. Because of the attenuation of the X-rays, the breast's stromal and epithelial tissues appear as shades of grey to white on a film-screen mammography. On a mammography, however, the fat inside the breast appears much darker and is more radiolucent. Breast density has been described by categorizing these various tissues' visual appearance on a mammogram in both qualitative and quantitative ways [49]. Physicians categorize breast composition (density) into four groups using the BI-RADS: a) nearly totally fatty, b) sporadic fibro glandular density areas, c) diversely dense (which could include small, obscuring masses), and d) immensely dense (which reduces mammography sensitivity) [50]. Matsuyama *et al*. used spectral data from mammograms to build an interpretable neural network-based model for breast density classification. Using a reliability diagram, they assessed whether the model's evaluation metrics produced dependable likelihood values and illustrated the foundation for the ultimate prediction. They applied modifications to ResNet50 to build the classification model, including new algorithms for quantifying prediction ambiguity, visualizing network behavior, and extracting and inputting image spectra. According to the experimental findings, the suggested model outperformed traditional CNN models that make use of image pixel information in terms of classification accuracy, reliability, and interpretability [51].

2. Breast asymmetries detection

Mammographic breast asymmetry is defined as unilateral deposits of fibrous and glandular tissue that do not meet the criteria for a mass. These deposits are categorized by medical professionals as developing, focal, global, and asymmetric [49]. The goal of asymmetry identification research is to both identify and measure the existence of asymmetry. Researchers have used distinct volume calculations to define asymmetry mathematically. Liao *et al*. examined the diagnostic performance of a deep learning system for benign and malignant asymmetric lesions in mammography using the DenseNet convolutional neural network [52]. They examined 460 women's clinical and imaging data, comparing the deep learning system's performance to that of senior and junior radiologists. The deep learning system outperformed junior radiologists in terms of specificity (0.909) and precision (0.872), with statistically significant differences in diagnostic efficacy ($p < 0.05$). Furthermore, the deep learning model's AUC-ROC (0.778) was noticeably higher than junior radiologists' (AUC: 0.579, 0.564).

3. Detection and Classification of Calcifications

Calcifications are white spots on a mammography that are flakes of calcium phosphate, calcium oxalate, or magnesium within the breast tissue [53]. The ductal system, breast acini, stroma, and vessels all develop calcifications. But it's unclear how calcifications happen and by what mechanisms [54]. The morphology, distribution, and occasional changes over time of breast calcifications identified by mammography are analyzed in the diagnostic process [49]. Radiologists classify calcifications as benign or suspicious based on this analysis; the latter category necessitates a biopsy [55]. Because microcalcifications (MCs) are present in about 55% of intangible breast malignancies, investigators have focused on segmenting and detecting them. They are responsible for 85–95% of ductal carcinoma in situ (DCIS) cases found by screening mammography, and they may also show up in metastatic cancers [56]. Banumathy *et al*. evaluated the Convolutional Neural Network's accuracy to identify the most effective early detection strategy for breast tissue malignancies, formation of masses, and breast MCs on mammograms. CNNs with class activation map (CAM) has also been used to perform breast microcalcifications detection to find a specific class in the biopsy image. This was done by a pre-trained CNN of Residual Network (ResNet50) for breast cancer detection, in order to obtain the discriminative localization. According to the test results, this method outperformed the others by nearly 225.15% when it came to using images of breast microcalcifications to precisely locate the disease (Discriminative Localization). For images, ResNet50 appears to have the highest level of accuracy of cases of malignant tumors (MT) and benign tumors (BT) at 97.11%. The pre-trained neural network accuracy of ResNet 50 is 94.17% on average [57].

4. Detection of Mass

A mass is defined as a 3D lesion that occupies space and is seen in two distinct projections. In mammography, a mass's shape, margins, and density are its descriptors [49]. For instance, the most prevalent form of invasive breast cancer is called invasive ductal carcinoma (IDC) (not otherwise specified); it usually manifests as a new focal asymmetry or as a spiculated, irregular mass [58]. Classification, segmentation, and detection of masses in the image are topics covered by researchers in the literature. Sun *et al*. proposed a novel method for detecting breast mass on a mammography image. The suggested technique used the mathematical method to preprocess the mammogram and the image matching method to identify the suspected breast mass regions. Next, it used a convolutional neural network (CNN) to classify these suspicious regions into mass and related categories in order to obtain the regions of breast mass. All of the flashing regions are converted into approximately circular areas by the mathematical morphology method, which also yields the bounding box of mass obtained by the image template matching method. The ideal solution for a breast mass bounding box regression should be found in the feasible region. As a result of particle swarm optimization algorithm, they refined the bounding box of mass. The DDSM is an open database that was used to evaluate the suggested mass detection method and the compared diagnostic methods. The experimental findings show that, in terms of detection performance, the suggested method outperforms every other detection method that was compared [59].

C. CNNs for BI-RADS Classification

Convolutional Neural Networks (CNNs) feature filters with specific patterns that preserve spatial features in the image, they are highly effective for image analysis [60]. Furthermore, because interrelated architectures flatten input images and ignore crucial spatial features, they perform better than fully connected neural network techniques as shown in Figure 5 [61].

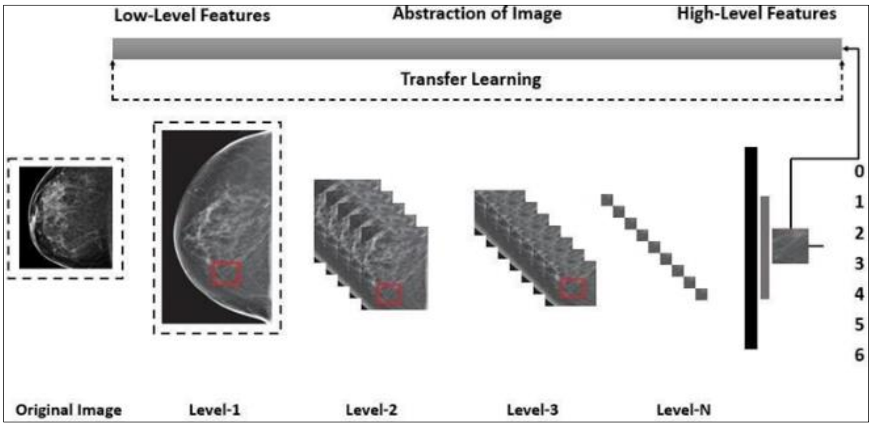


Fig 5: This CNN architecture illustrates the significant reliance of CNN-based techniques on end- to-end learning. The framework prioritizes unprocessed input data with minimal feature engineering for various tasks, including BI-RADS category prediction (0 to 6). Pattern-specific feature maps are produced by the network's deeper layers identifying abstract patterns, while the network's shallow layers identify features that resemble the input image

1. BI-RADS category prediction from Mammography Images

In breast imaging with mammography, the current processing rapidly converts into BI-RADS categories by means of deep learning algorithms and CNNs. Such classifications are very important factor to evaluate the probability of the woman to be breast cancer which in turn to help in further diagnostic and treatment process. Various deep learning algorithms have been developed for BI-RADS category prediction from mammography images. In addition, a lot of natural language processing techniques have been utilized for the similar task. But the number of studies is very less which are predicting BI-RADS categories [62-75] as most of the studies have focused on mass detection and classification, calcification detection, and density prediction [50-59]. Table 1 gives the description of all the studies which have been done for the BI-RADS category prediction directly from mammography images or radiology reports.

EfficientNet-Based DNN has been utilized for multi-class prediction of BI-RADS categories. This model was trained on images blocks picked from mammogram datasets for categorization [62]. One method implemented Softmax Regression classification when MobileNetV2 was embedded for feature extraction [63]. This method was projected to cover the category 4 lesions that are BI-RADS (subcategories 4A, 4B, 4C). It was designed to predict for binary classification which differentiated the benign from the malignant cases. Similarly, a deep convolutional neural network (DCNN) was used to delineate microcalcifications and distinguish between malignant and benign masses in mammograms. This method categorized findings into three cohorts: no microcalcification, benign microcalcification, and suspicious microcalcification, thus we can identify the abnormality or disease in the initial stage of its formation. On the basis of mentioned studies [62-75], we propose a BI-RADS category prediction from mass and calcification using a multi-label approach given in Figure 6.

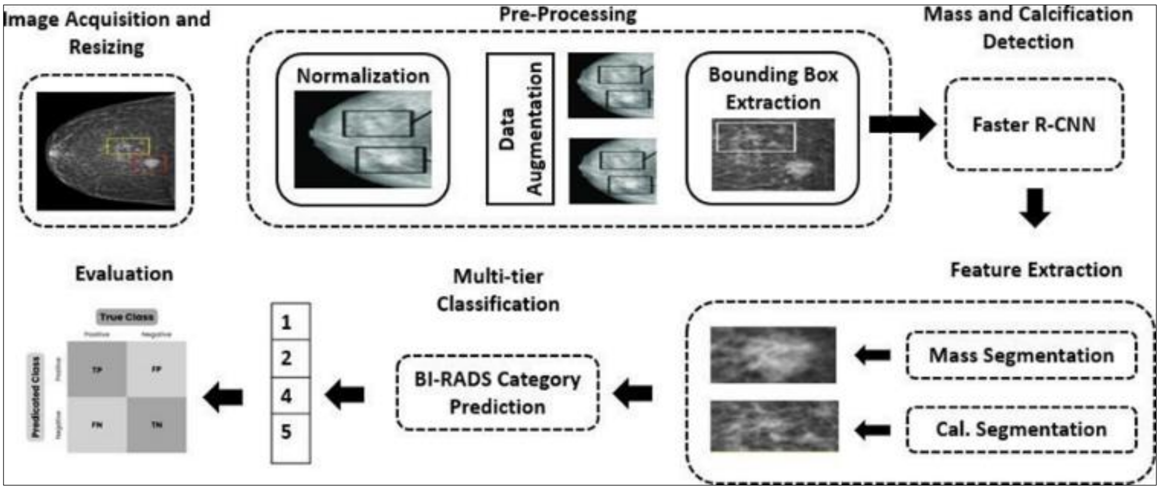


Fig 6: A multi-label classification framework for BI-RADS category prediction

2. BI-RADS categories from Mammography Radiology Reports

Exploring deep learning models (also known as deep neural networks) for various tasks has become more and more of a focus in the field of natural language processing. Deep neural network architectures free researchers from laborious feature engineering procedures by automatically learning high-level features from huge corpora. Researchers have created two well-liked deep architectures, the RNN and the CNN, to capture long-term dependencies in a word sequence. NLP technologies are needed to unlock detailed information from the vast number of clinical documents that are now electronically available due to the rapid growth of electronic health record (EHR) systems. NLP techniques have been developed by researchers to extract clinical data from clinical texts [76]. Symptoms and pathogenesis in medical records [77], symptoms in clinical notes [78], and issues, tests, and treatments in hospital discharge summaries [79] are just a few of the studies that have focused on the named entity recognition (NER) task.

Researchers have extracted BI-RADS categories on the basis of assessment and findings from breast radiology reports in various languages like English [65, 72-75] and Chinese [71] by using DL based Natural Language Processing (NLP) techniques such as PART algorithm, GATE-NLP, and Bayesian Network. Rule- based techniques were frequently used in to extract BI-RADS. For example, a rule-based system was created and assessed by Gao *et al.* [72] using SAS Base to extract four different kinds of BI- RADS findings from reports on breast radiology. With an F1-measure of 0.911, the deep learning-based techniques have outperformed the rule- based system.

Table 2: The summarized literature in this review

Author, Year	Target Variable	Input	Architecture	Pre- Processing	Dataset	Outcome	Output	Result
Tsai *et al.* - 2022	Risk of Breast Cancer	Mammograms	DNN based on EfficientNet	Block-Based Images segmented from a mammogram dataset	E-Da Hospital, Taiwan - 5733	BI-RADS category	Multi-class	Acc.: 94.22
Liu *et al.* - 2021	Risk of Breast Cancer	Clinical Data Mammograms	Softmax Regression	Feature extraction using MobileNetV2	Mammography dataset-384	BI-RADS category 4	Binary	AUC: 0.91

			Classification			(4A, 4B, 4C)		
Schönenberger *et al.* - 2020	Risk of Breast Cancer	Mammograms	dCNN	Image Segmentation	94 Patients- 268 Mammograms- 56,000	BI-RADS category (3 cohorts)	No Microcalcification, Benign microcalcification, and Suspicious Microcalcification	Acc.:99.6
BOROUMANDZADEH & PARVINNIA – 2021	Risk of Breast Cancer	Mammography Radiology Reports	MLF	word2vec and TFIDF	Namazi Hospital and Saadi Hospital Iran - 5076	BI-RADS category	Multi-class	Acc.: 0.89
Yang *et al.* - 2024	Malignancy in dense breasts	Mammograms Ultrasound Images	CNN	Image segmentation	US and MG- 992	BI-RADS category 4A	Binary	AUC: 0.94
Sabani *et al.* - 2022	Tissue Opacities in Breasts	Mammograms	4 dCNN	Image Labelling Data augmentation	PACS – 438 Patients – 1744 Mammograms	BI-RADS category (3 cohorts)	Normal tissue, Probably benign tissue opacities and suspicious opacities	Acc.: 0.89
Nguyen *et al.* - 2022	Risk of Breast Cancer	Mammograms	ResNet-34 EfficientNet-B2	Automatic Detection using YOLO v5	VinDrMammo, DDSM, and Hanoi Medical University	BI-RADS category and Density Prediction	Normal, Benign, and Malignant	F1: 0.672
					Hospital – 36,138			
Vanderheyden & Xie - 2020	Risk of Breast Cancer	Mammograms	OPHLall	Cropping and resizing	CBIS-DDSM-2238	BI-RADS category	1 to 5	MAE: 0.612
Hejduk *et al.* - 2022	Detection and Classification of Lesions	Mammograms	dCNN	Data augmentation	ABUS - 665	BI-RADS category	BI-RADS 2, 3, and 4	Acc.: 79.9
Miao *et al.* - 2018	Risk of Breast Cancer	Breast Ultrasound Reports	Bidirectional RNN with embeddings	Annotations and Character embedding	Affiliated Hospital with Nanjing Medical University - 540	BI-RADS findings	Multi-class	F1: 0.911
Gao *et al.* - 2015	Likelihood of Breast Cancer	Mammogram Reports	Natural Language Processing	Sentence Segmentation Tokenization	Group Health - 93,705	Mammogram Findings	masses, calcifications, asymmetries, and architectural distortions	Sensitivity: 0.92
Bozkurt *et al.* - 2014	Likelihood of Breast Cancer	Mammogram Reports	GATE NLP	Tokenization Stemming Named Entity Recognition	Reports - 190	Laterality of BI-RADS	Multi-class	Acc.: 81%
Banerjee *et al.* - 2019	Likelihood of Breast Cancer	Mammogram Reports	NLP-Bayesian Network	Word Embeddings Report Vector Creation	radTF Stanford University, Oncoshare database - 1, 22,109	BI-RADS category	Multi-class	F1: 0.87
Castro *et al.* - 2017	Likelihood of Cancer	Mammogram Reports	NLP PART algorithm	Clinical Text Analysis	University of Pittsburgh TIES – 24 M	Annotation of all BI-RADS categories and Classification of laterality	Multi-class	Acc.: 0.87

5. Discussion and open research areas

Mammography is a commonly employed technique for the prompt detection of breast cancer. It provides X-ray images with high resolution, making it possible to see the various layers of breast tissue. Hyperplasia, benign and malignant tumors, and abnormalities in the structure of the breast tissue can all be detected with this method [80]. Additionally, mammography is very important for early prediction, diagnosis, and management because it produces clear images that make previous and prior comparisons easier. This narrative review analyses and discusses the importance of BI-RADS category prediction using deep learning approaches both from mammography images and mammography radiology reports. The article begins by defining current deep learning algorithms for BI-RADS category prediction and outlining the basic theoretical concepts. The research on the use of CADe systems in mammography such as feature selection, segmentation of lesion, and classification of benign and malignant lesions for BI-RADS prediction is then covered. The paper then explores the use of deep learning in mammography, including mass detection, calcification detection and classification, and briefly summarizes the benefits and drawbacks of such systems. There are still numerous obstacles to overcome in spite of the advancements in research on computerized breast cancer screening systems:

1. There is a dearth of standardized sample data for BI-RADS prediction into all categories based on mass and calcification detection. The majority of current research ignores the examination of mass and calcification at the same time in favor of mass and microcalcification detection separately. Thus, it is imperative that researchers give top priority to developing automated tools for detecting mass and calcification for BI-RADS category prediction, as well as models for detecting multiple lesions.
2. Unstable lesion depiction and a higher chance of misdiagnosis are caused by the irregularities in the forms and borders of suspicious lesions as well as the hazy lining between the lesion and the surrounding tissues. Even though the accuracy DL-assisted breast screening has greatly increased, developing a reliable mass and calcification identification system and putting it into practice on a big scale are still difficult research topics. Consequently, more advancements in the methods for the automated identification and division of breast lesions in mammograms are required.
3. Even with the growing body of research on deep learning, there are still barriers to its widespread use. First off, a lot of raw data is needed for training a new deep learning system. Second, the replicability of research findings across different datasets is hampered by the absence of a common standard in datasets as a result of technological, equipment, and operator limitations. Furthermore, deep learning technology is expensive due to its intricate internal workings [81]. Moreover, output results from DL models are usually presented without clear explanations.
4. Completely annotated data for both mass and calcification are scarce, and the current database is tiny. For training, both images and reports need a sizable amount of labelled and tagged data respectively. However, because medical data is highly specialized and fragmented, it can be difficult to obtain such data. Thus, in order to improve data availability, it is essential that future research enlarge the dataset and investigate the application of deep learning for BI-RADS category prediction
5. Lastly, there is a need to develop more robust models which can analyze certain features from mammography images and mammography radiology reports and combine deep learning and natural language processing techniques for BI-RADS category prediction. And, the potential area of research is to develop such algorithms which can classify the breast lesions into all the seven categories instead of binary or no, benign, and malignant classification.

6. Challenges and limitations

As one of the first oncologists use to assess risk status, classifying breast density for BI-RADS category prediction is a difficult but important task. Depending on the type of mammography machine used, up to 22% of women have different BI-RADS prediction [82]. Oncologists, above all, should have faith in the way mammogram labels are applied. Although clinics now have highly magnified imaging techniques, it is difficult to standardize CAD performance across different machines.

Despite the fact that mammography is our primary focus, our investigation of the literature reveals that the majority of reported CNN-based asymmetry detection for BI-RADS category prediction is on thermography, a clinical imaging modality that is not commonly used in clinics. We encourage more research focused on the identification of asymmetry in mammography for BI-RADS category prediction from 0 to 6. From a technical standpoint, we point out that for ground truth measurements, researchers must report what they interpret as an asymmetry for each category when they are predicting BI-RADS categories. Asymmetry is discussed in general terms in many of the papers that use neural networks for asymmetry detection, and there is no explicit mathematical formalism of what represents asymmetry as ground truth for each category of BI-RADS. We observe that the majority of publicly available datasets include both the CC and MLO views. We do acknowledge, though, that most publicly available datasets do not have annotations for asymmetry and density or even breast volume to predict BI-RADS categories. We invite papers that either add these essential empirical annotations to the datasets or introduce new datasets with annotations for asymmetry along with mass and calcification for BI- RADS prediction.

Datasets with the BI-RADS labels are the only ones used in the literature on CNNs used for classification. Most published literature is restricted to mammography repositories that are privately collected. The findings show that attention-based mechanisms in conjunction with CNNs have become more popular over the last three years, which has improved model performance. Growing ROC-AUC trends demonstrate how transfer learning enhances attention-based models. Analyzing the quantitative outcomes from diverse approaches, we observe that the majority of recent publications now report the ROC-AUC curves. Because it enables readers to compare TP and FP rates, the ROC-AUC is a more illuminating statistic. To improve model performance, researchers have recently used complex convolutional neural networks block combinations like squeeze-excitation and attention networks.

Enhancements have been made by researchers to improve BI-RADS category prediction; 0.99 is the highest described sensitivity. However, there is still opportunity to further enhance these models' specificity. It makes sense why the generative adversarial network-based data augmentation model performs worse: The investigators test the synthetic image model based on DDSM, indicating that further research should enhance the GANs' capacity for data augmentation. The main focus in BI-RADS category prediction is the sensitivity, which is also referred to as the true-positive rate. The sensitivity rate for BI-RADS category prediction is usually 0.90 employing multiclass classification into normal, benign and malignant. To increase reported sensitivity rates, further research into the recent mass segmentation applications of autoencoders is warranted. However, we emphasize the need for research on learning transfers across domains.

A. Gold Standard in Current Approaches

The "radiologist" is the gold standard that is commonly used in AI research to determine whether the algorithm outperformed or was on par with the radiologist. Using deep learning for mammography screening, Becker *et al.* demonstrated that while radiologists had higher specificity and lower sensitivity than neural networks, there was no discernible difference in the AUCs between the trained algorithm and radiology expert [83]. Similar to this, a good accuracy was achieved comparable to a subspecialty radiologist by using texture parameters from multiparametric MR images with a random forest model [84].

Significant deep learning research has been documented in the literature, but notably absent is a radiologist- augmented approach for BI-RADS category prediction. With the aim of replacing radiologists, the algorithms that have been evaluated whether to predict BI-RADS category from mammography images or mammography radiology reports, typically operate without regard to the opinions of radiologists. While this could alleviate radiologists' workloads and enhance patient flow, the question of whether we can improve ourselves is still unanswered. Working together, the radiologist and deep learning algorithms could extend the radiologist- augmented workflow for determining the BI-RADS category in mammography screening and produce outcomes that are more in line with the gold standard of diagnostic breast radiology.

Models that included both the BI-RADS category and the BI-RADS descriptions were more accurate than those that only included the BI-RADs descriptions of mass and calcification margin, density, and shape along with the patients' age. Despite the small amount of data and extremely basic parameters used in these models, the results have successfully shown that it is possible to implement a CNN workflow that augments radiologists, enabling better patient screening and BI-RADS category classification. The development of computer-aided detection algorithms has made extensive use of the texture parameters of breast lesions in literature. Lesion segmentation was not done and instead a very basic texture model was used.

7. Conclusion

We highlight important unresolved research issues across the literature. The application of knowledge from larger and annotated datasets to sparser ones has been aided by transfer learning. Further investigation is necessary for the BI-RADS category prediction based on mass, calcification, and asymmetry, though. This method necessitates more annotated datasets, so future research should concentrate on producing large corpora. Our review presents a need for public data by tabulating the currently available public datasets of breast images. It is also recommended that practitioners must test their methods on multiple datasets in order to verify the cross-domain robustness of a model. The current corpus in this paper is biased towards tasks related to BI-RADS category prediction based on mass detection. Given that there are comparatively fewer related publications for BI-RADS prediction on the basis of density, asymmetry, and mass or calcification, researchers ought to expand their use of CNN-based techniques in these areas. Researchers should look into predicting BI-RADS from fine- grained mass types as there isn't much literature on the localization and classification of breast masses that goes beyond differentiating between benign and malignant masses. Moreover, research that focuses on this gap could greatly enhance CAD systems. Models perform well in fatty breast tissue but poorly in dense tissue.

8. Acknowledgment

The authors acknowledge the support from Buch International Hospital, Multan, Pakistan. https://www.buchhospital.com/

9. Refrences

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-249.
2. Xu S, Liu Y, Zhang T, Zheng J, Lin W, Cai J, *et al*. The global, regional, and national burden and trends of breast cancer from 1990 to 2019: results from the global burden of disease study 2019. Front Oncol. 2021;11:689562.
3. Giordano SH. Breast cancer in men. N Engl J Med. 2018;378(24):2311-2320.
4. Anderson BO, Ilbawi AM, Fidarova E, Weiderpass E, Stevens L, Abdel-Wahab M, Mikkelsen B. The Global Breast Cancer Initiative: a strategic collaboration to strengthen health care for non-communicable diseases. Lancet Oncol. 2021;22(5):578-581.
5. Lin L, Yan L, Liu Y, Yuan F, Li H, Ni J. Incidence and death in 29 cancer groups in 2017 and trend analysis from 1990 to 2017 from the Global Burden of Disease Study. J Hematol Oncol. 2019;12:1-21.
6. Ataollahi MR, Sharifi J, Paknahad MR, Paknahad A. Breast cancer and associated factors: a review. J Med Life. 2015;8(Spec Iss 4):6.
7. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. Cancers (Basel). 2021;13(17):4287.
8. Misra S, Solomon NL, Moffat FL, Koniaris LG. Screening criteria for breast cancer. Adv Surg. 2010;44(1):87-100.
9. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med. 2016;375(15):1438-1447.
10. Mercado CL. BI-RADS update. Radiol Clin North Am. 2014;52(3):481-487.
11. Orel SG, Kay N, Reynolds C, Sullivan DC. BI-RADS

categorization as a predictor of malignancy. Radiology. 1999;211(3):845-850.
12. Strigel RM, Burnside ES, Elezaby M, Fowler AM, Kelcz F, Salkowski LR, DeMartini WB. Utility of BI-RADS assessment category 4 subdivisions for screening breast MRI. AJR Am J Roentgenol. 2017;208(6):1392-1399.
13. Malmartel A, Tron A, Caulliez S. Accuracy of clinical breast examination's abnormalities for breast cancer screening: cross-sectional study. Eur J Obstet Gynecol Reprod Biol. 2019;237:1-6.
14. Ciritsis A, Rossi C, Eberhard M, Marcon M, Becker AS, Boss A. Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making. Eur Radiol. 2019;29:5458-5468.
15. Gastounioti A, McCarthy AM, Pantalone L, Synnestvedt M, Kontos D, Conant EF. Effect of mammographic screening modality on breast density assessment: digital mammography versus digital breast tomosynthesis. Radiology. 2018;291(2):320-327.
16. Kissane J, Neutze JA, Singh H. Breast imaging. In: Radiology Fundamentals: Introduction to Imaging & Technology. 2020;139-154.
17. Bick U, Diekmann F. Digital mammography: what do we and what don't we know?. Eur Radiol. 2007;17:1931-1942.
18. McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol. 1996;15(6):423.
19. Horsley RK, Kling JM, Vegunta S, Lorans R, Patel BK. Baseline mammography: what is it and why is it important? A cross-sectional survey of women undergoing screening mammography. J Am Coll Radiol. 2019;16(2):164-169.
20. Adcock KA. Initiative to improve mammogram interpretation. Perm J. 2004;8(2):12.
21. Hashemi RH, Bradley WG, Lisanti CJ. MRI: the basics. 2nd ed. Lippincott Williams & Wilkins; 2012.
22. Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. Lancet. 2011;378(9805):1804-1811.
23. Morris EA. Breast cancer imaging with MRI. Radiol Clin North Am. 2002;40(3):443-466.
24. Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. J Magn Reson Imaging. 2019;50(2):377-390.
25. Kopans DB. Breast-cancer screening with ultrasonography. Lancet. 1999;354(9196):2096-2097.
26. Feig S. Cost-effectiveness of mammography, MRI, and ultrasonography for breast cancer screening. Radiol Clin North Am. 2010;48(5):879-891.
27. Gartlehner G, Thaler KJ, Chapman A, Kaminski A, Berzaczy D, Van Noord MG, Helbich TH. Adjunct ultrasonography for breast cancer screening in women at average risk: a systematic review. JBI Evid Implement. 2013;11(2):87-93.
28. Yaffe MJ, Jong RA. Adjunctive ultrasonography in breast cancer screening. Lancet. 2016;387(10016):313-314.
29. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol. 2016;71(2):141-150.
30. Skaane P, Gullien R, Bjørndal H, Eben EB, Ekseth U, Haakenaasen U, Krager M. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. Acta Radiol. 2012;53(5):524-529.
31. Abdelrahman L, Al Ghamdi M, Collado-Mesa F, Abdel-Mottaleb M. Convolutional neural networks for breast cancer detection in mammography: A survey. Comput Biol Med. 2021;131:104248.
32. Hassan NM, Hamad S, Mahar K. Mammogram breast cancer CAD systems for mass detection and classification: a review. Multimed Tools Appl. 2022;81:20043-20075.
33. Raza S, Goldkamp AL, Chikarmane SA, Birdwell RL. US of breast masses categorized as BI-RADS 3, 4, and 5: pictorial review of factors influencing clinical management. Radiographics. 2010;30(5):1199-1213.
34. Tan YJ, Sim KS, Ting FF. Breast cancer detection using convolutional neural networks for mammogram imaging system. In: 2017 International Conference on Robotics, Automation and Sciences (ICORAS); 2017 Nov. p. 1-5. IEEE.
35. Cè M, Caloro E, Pellegrino ME, Basile M, Sorce A, Fazzini D, *et al*. Artificial intelligence in breast cancer imaging: Risk stratification, lesion detection and classification, treatment planning and prognosis— A narrative review. Exp Target Anti-tumor Ther. 2022;3(6):795.
36. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. In: Semin Cancer Biol. 2021 Jul;72:214-225. Academic Press.
37. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction. In: Semin Cancer Biol. 2021 Jul;72:238-250. Academic Press.
38. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. Breast Cancer Res Treat. 2018;169:217-229.
39. Gao YE, Lin J, Zhou Y, Lin R. The application of traditional machine learning and deep learning techniques in mammography: a review. Front Oncol. 2023;13:1213045.
40. Yu X, Zhou Q, Wang S, Zhang YD. A systematic survey of deep learning in breast cancer. Int J Intell Syst. 2022;37(1):152-216.
41. Chen C. Science mapping: a systematic review of the literature. J Data Inf Sci. 2017;2(2):1-40.
42. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? J Am Coll Radiol. 2010;7(10):802-805.
43. De Jesus C, Moseley TW, Diaz V, Vishwanath V, Jean S, Elhatw A, *et al*. The benefits of screening mammography. Curr Breast Cancer Rep. 2023;15(2):103-107.
44. Ali S, Tanveer H. A focus on brain health through artificial intelligence and machine learning. Int J Multidiscip Res Growth Eval.
45. Ali MT, Ali U, Ali S, Tanveer H. Transforming cardiac care: AI and machine learning innovations. Int J Multidiscip Res Growth Eval.
46. Ahmad A, Hussain HK, Tanveer H, Kiruthiga T, Gupta K. The intelligent heart rate monitoring model for survivability prediction of cardiac arrest patients using deep cardiac learning model. In: 2023 International

Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS).
47. Tanveer H, Zahra N, Batool N. Time Series Classification and Anomaly Detection with Deep Models. Int J Multidiscip Res Growth Eval.
48. Tanveer H, Adam MA, Khan MA, Ali MA, Shakoor A. Analyzing the performance and efficiency of machine learning algorithms, such as deep learning, decision trees, or support vector machines, on various datasets and applications. Asian Bull Big Data Manag. 2023;3(2):126-136.
49. Tanveer H, Zahra N, Batool N. Efficient data delivery through decentralized networks. Int J Multidiscip Res Growth Eval.
50. Zahra N, Tanveer H, Batool N. Efficient data delivery through decentralized networks. Int J Multidiscip Res Growth Eval.
51. Zahra N, Tanveer H, Batool N. Technological advancement in covid-19 rehabilitation: Therapists' views. Int J Multidiscip Res Growth Eval.
52. Zahra N, Tanveer H, Batool N. Technological advancement in covid-19 rehabilitation: Therapists' views. Int J Multidiscip Res Growth Eval.
53. Henrot P, Leroux A, Barlier C, Genin P. Breast microcalcifications: the lesions in anatomical pathology. Diagn Interv Imag. 2014;95(2):141-152.
54. Wilkinson L, Thomas V, Sharma N. Microcalcification on mammography: approaches to interpretation and biopsy. Br J Radiol. 2017;90(1069):20160594.
55. Bent CK, Bassett LW, Orsi CJD, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. Am J Roentgenol. 2010;194(5):1378-1383.
56. Stomper PC, Geradts J, Edge SB, Levine EG. Mammographic predictors of the presence and size of invasive carcinomas associated with malignant microcalcification lesions without a mass. Am J Roentgenol. 2003;181(6):1679-1684.
57. Banumathy D, Khalaf OI, Romero CAT, Raja PV, Sharma DK. Breast calcifications and histopathological analysis on tumor detection by CNN. Comput Syst Sci Eng. 2023;47:2219-2230.
58. Baines CJ, Woolcott CG, Kahn L, Lee K. Impact of mammography screening on breast cancer mortality: the Canadian National Breast Screening Study. J Natl Cancer Inst. 2014;106(11):dju231.
59. Sardanelli F, Pupi A, Fileni A, Ciatto S. Italian national screening program for breast cancer. Radiol Clin North Am. 2010;48(5):835-844.
60. Bese NS, Ulger Y, Ozdemir N. The role of mammography in breast cancer detection. J Cancer Res Ther. 2013;9(3):401-407.
61. Pinaya WHL, Vieira S, Garcia-Dias R, Mechelli A. Convolutional neural networks. In: Machine learning. Academic Press; 2020. p. 173-191.
62. Tsai KJ, Chou MC, Li HM, Liu ST, Hsu JH, Yeh WC, *et al*. A high-performance deep neural network model for BI-RADS classification of screening mammography. Sensors. 2022;22(3):1160.
63. Liu H, Chen Y, Zhang Y, Wang L, Luo R, Wu H, *et al*. A deep learning model integrating mammography and clinical factors facilitates the malignancy prediction of BI-RADS 4 microcalcifications in breast cancer screening. Eur Radiol. 2021;31:5902-5912.
64. Schönenberger C, Hejduk P, Ciritsis A, Marcon M, Rossi C, Boss A. Classification of mammographic breast microcalcifications using a deep convolutional neural network: A BI-RADS–based approach. Investig Radiol. 2021;56(4):224-231.
65. Boroumandzadeh M, Parvinnia E. Automated classification of BI-RADS in textual mammography reports. Turk J Electr Eng Comput Sci. 2021;29(2):632-647.
66. Yang Y, Zhong Y, Li J, Feng J, Gong C, Yu Y, *et al*. Deep learning combining mammography and ultrasound images to predict the malignancy of BI-RADS US 4A lesions in women with dense breasts: a diagnostic study. Int J Surg. 2024:10-1097.
67. Sabani A, Landsmann A, Hejduk P, Schmidt C, Marcon M, Borkowski K, *et al*. BI-RADS-based classification of mammographic soft tissue opacities using a deep convolutional neural network. Diagnostics. 2022;12(7):1564.
68. Nguyen HT, Tran SB, Nguyen DB, Pham HH, Nguyen HQ. A novel multi-view deep learning approach for BI-RADS and density assessment of mammograms. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 2144-2148.
69. Vanderheyden R, Xie Y. Mammography image BI-RADS classification using ohplall. In: 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE; 2020. p. 120-127.
70. Hejduk P, Marcon M, Unkelbach J, Ciritsis A, Rossi C, Borkowski K, Boss A. Fully automatic classification of automated breast ultrasound (ABUS) imaging according to BI-RADS using a deep convolutional neural network. Eur Radiol. 2022;32(7):4868-4878.
71. Miao S, Xu T, Wu Y, Xie H, Wang J, Jing S, *et al*. Extraction of BI-RADS findings from breast ultrasound reports in Chinese using deep learning approaches. Int J Med Inform. 2018;119:17-21.
72. Gao H, Bowles EJA, Carrell D, Buist DS. Using natural language processing to extract mammographic findings. J Biomed Inform. 2015;54:77-84.
73. Bozkurt S, Rubin D. Automated detection of ambiguity in BI-RADS assessment categories in mammography reports. In: Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilising Big Data. IOS Press; 2014. p. 35-39.
74. Banerjee I, Bozkurt S, Alkim E, Sagreiya H, Kurian AW, Rubin DL. Automatic inference of BI-RADS final assessment categories from narrative mammography report findings. J Biomed Inform. 2019;92:103137.
75. Castro SM, Tseytlin E, Medvedeva O, Mitchell K, Visweswaran S, Bekhuis T, Jacobson RS. Automated annotation and classification of BI-RADS assessment from radiology reports. J Biomed Inform. 2017;69:177-187.
76. Casey A, Davidson E, Poon M, Dong H, Duma D, Grivas A, *et al*. A systematic review of natural language processing applied to radiology reports. BMC Med Inform Decis Mak. 2021;21(1):179.
77. Oztan O, Garner JP, Constantino JN, Parker KJ. Neonatal CSF vasopressin concentration predicts later medical record diagnoses of autism spectrum disorder. Proc Natl Acad Sci U S A. 2020;117(19):10609-10613.

78. Luo X, Gandhi P, Storey S, Huang K. A deep language model for symptom extraction from clinical text and its application to extract covid-19 symptoms from social media. IEEE J Biomed Health Inform. 2021;26(4):1737-1748.
79. Fernandes M, Sun H, Jain A, Alabsi HS, Brenner LN, Ye E, *et al*. Classification of the disposition of patients hospitalized with COVID-19: reading discharge summaries using natural language processing. JMIR Med Inform. 2021;9(2):e25457.
80. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med. 1994;331(22):1493-1499.
81. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Dean J. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24.
82. Melnikow J, *et al*. Supplemental screening for breast cancer in women with dense breasts: a systematic review for the US Preventive Services Task Force. Ann Intern Med. 2016;164(4):268-278.
83. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol. 2017;52:434-440.
84. Ueno Y, Forghani B, Forghani R, Dohan A, Zeng XZ, Chamming's F, *et al*. Endometrial carcinoma: MR imaging-based texture model for preoperative risk stratification—a preliminary analysis. Radiology. 2017;284:748-757.