International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Oct-24)

http://dx.doi.org/10.12785/ijcds/1601122

Harnessing Deep Learning for Early Breast Cancer Diagnosis: A Review of Datasets, Methods, Challenges, and Future Directions

Marwa Ben Ammar[1], Faten Labbene Ayachi[2], Anselmo Cardoso de Paiva[3], Riadh Ksantini[4] and Halima Mahjoubi[5]

[1,5]*Research Laboratory of Biophysics and Medical Technologies, Higher Institute of Medical Technologies of Tunis (ISTMT), University of Tunis elManar, Tunisia*
[2]*Research laboratory Innov'COM Innovation of COMmunicant and COoperative Mobiles Laboratory , Higher School of Communication of Tunis (SUPCOM), University of Carthage, Tunisia*
[3]*Applied Computer Group NCA-UFMA, Federal University of Maranhao, Sao Luis, MA, Brazil*
[4]*Department of Computer Science, College of IT, University of Bahrain, Bahrain*

Received 16 Dec. 2023, Revised 27 Jun. 2024, Accepted 28 Jun. 2024, Published 1 Oct. 2024

Abstract: Breast cancer is the most common kind of cancer diagnosed worldwide and the leading cause of cancer-related deaths among women. Therefore it presents a significant public health risk. Therefore, early identification and diagnosis of malignant breast tumors can significantly increase patient survival rates and facilitate effective treatment. Imaging is one of the key procedures in decision-making for diagnosing breast cancer. For instance, mammography is the most efficient and highly recommended imaging technique by radiologists in identifying many types of breast abnormalities. However, with the daily growth in mammography, it is still challenging for specialists to give correct and consistent interpretations, which can lead to potential misinterpretations and unneeded biopsies. Statistics show that substantial portions, ranging from 10% to 30% of incorrect diagnoses in medical image analysis are the result of human error. Various researchers have looked into the use of mammography and Deep Learning (DL) approaches for accurate early breast cancer diagnosis. Utilizing these approaches in clinical settings can increase diagnosis accuracy, save time, lower the likelihood of mistakes and errors, increase patient satisfaction, and streamline radiologists' workloads. This paper presents a comprehensive examination of breast cancer issues, discussing the important role of the mammography examination for early cancer detection and how these image examinations can be used through Deep Learning (DL) approaches for accurate early breast cancer diagnosis. We describe the fundamental architectural deep learning components of the commonly used systems employed for breast cancer diagnosis, present the main publicly available datasets, and discuss the constraints, difficulties, and avenues for future research in the realm of breast cancer diagnosis and classification. Finally, issues and potential research objectives in this developing field are outlined. Approaching this topic, we intend to inspire and direct medical professionals, researchers, scientists, and other healthcare workers interested in creating cutting-edge applications for early breast cancer diagnosis using mammography image processing in the right direction.

Keywords: Mammography imaging, Deep Learning, Breast Cancer Diagnosis, medical images, artificial intelligence

1. Introduction

Breast cancer is a prevalent form of cancer among women on a global scale. Recent statistics from the Global Cancer Observatory (GCO), a partner organization of the World Health Organization, indicate a noteworthy change in global cancer trends in 2020. According to these statistics, female mammary carcinoma is becoming the preeminent most identified type, surpassing lung malignancy on the global scale [1]. These estimates indicate that approximately 2.3 million new cases of breast cancer were reported, accounting for 11.7% of all cancer cases. Lung cancer came in second place with 11.4% of cases, followed by colorectal cancer with 10%, prostate cancer with 7.3%, and stomach cancer with 5.6%. In addition, breast cancer accounts for 685,000 cancer-related deaths globally, and by 2070, it is projected to affect 4.4 million women [2]. In 2020, breast cancer was the first cause of mortality and the most prevalent newly diagnosed cancer in most countries, presenting over 24.5% of all cancer diagnoses and 15.5% of cancer-related fatalities in women [1]. Breast cancer is an imminent threat for all women, in general, and is also related to ageing. Being female and getting older are the two primary elements that increase the chances of developing breast cancer. There are various lifestyle factors (e.g., alcohol consumption, obesity, sedentary lifestyle) and hormonal issues (e.g., first menstruation at an early age,

E-mail address: marwa.ammar@istmt.utm.tn, faten.labbene@supcom.tn, paiva@nca.ufma.br, ksontiniriadh@yahoo.fr, halima.mahjoubi@istmt.utm.tn

late menopause, late primiparity, use of contraceptives) increasing the susceptibility to developing breast cancer. Consequently, breast cancer is a serious illness, although it typically has a fair prognosis when diagnosed early. To enhance the prognosis, elevate the patient's chance of survival by 50% [3], and reduce the potential side effects from some treatments, it is essential to diagnose this life-threatening condition as early and accurately as possible. The early-stage breast cancer diagnosis currently relies on various widely recognized imaging techniques, including mammography with X-rays [4], Ultrasonography [5], computerized tomography, also known as CT [6], and MRI, which stands for magnetic resonance imaging [7]. Mammography continues to be the modality that radiologists utilize the most frequently to diagnose this illness [8] appropriately. Radiologists should, in practice, identify aberrant lesions on mammograms during the diagnosing process to differentiate between masses, calcifications, and other frequently occurring abnormalities. Extraction of specific information about suspicious lesions (size, shape, contour, etc.) is another activity carried out by medical professionals. This information enables doctors to assess the severity of suspicious tumor regions and establish if they are benign or malignant. Finally, experts should decide how to proceed in cases of tumors with a clear indication of the level of tumor suspicion [9] and in accordance with the classification protocol outlined by the American College of Radiology (ACR), known as Breast Imaging Reporting and Data System (BI-RADS)for reporting and interpreting breast imaging findings. Unnecessary biopsies raise the expense of healthcare, exacerbate patient anxiety, and increase morbidity. However, with the daily growth in mammography, it is still challenging for radiologists and doctors to give reliable and consistent analysis, leading to diagnostic blunders and pointless biopsies. Typically, False-Positive (FP) and False-Negative (FN) error types are the two main sorts of mistakes that might happen. Since benign areas are mistaken for malignant ones, the case of false positives has undesirable outcomes. False negatives are more significant when they put the patient's life in peril, and this happens when the radiologist misses an abnormality. Additionally, studies have shown that lesions with a greater than 2% potential of being malignant will be advised for biopsy to decrease the likelihood of FN diagnosis. Only 15–30% of those who get a biopsy are ultimately found to have cancer. Sophisticated algorithms, categorized as CADe for computer-aided detection and CADx for computer-aided diagnosis systems, are developed to assist medical specialists with interpreting medical images. These systems are used to reduce the likelihood of misunderstandings and ensure early breast cancer diagnosis. Recently, scientific researchers, technology specialists, and clinicians have been continuously developing and evaluating CADe/CADx systems based on Deep Learning (DL) methods. CADe/CADx systems were developed to help doctors categorize tumors into various classifications, such as ductal cancer within situ, cancer that is invasive, lobular cancer, etc., and to help them determine whether the growth is healthy or cancerous. Additionally, these computer programs assist in preventing unneeded biopsies and preserve a lot of time for human professionals who would otherwise have to review medical images manually. The concept of CAD systems was originally developed in the 1960s to screen for breast cancer using mammograms. Currently, it is among the most important study areas for clinical image processing [10]. CADe uses computed findings to pinpoint the exact location of the lesions concerned while leaving the radiologist to make sense of these anomalies. In contrast, CADx produces quality information that assists the radiologist in making decisions regarding the observed anomalies, in particular, to further identify and classify lesions [11]. Recent considerable advancements and exceptional performance of deep learning (DL) methods have encouraged several researchers to leverage the power of DL in the diagnosis of breast cancer. The usage of DL within (CAD) systems is growing, replacing more established machine learning (ML) techniques [12]. This shift towards deep learning-based CAD offers several advantages, including the capability to discern malignant from normal breast lesions without the necessity of segmenting breast lesions, computing image features, or employing a selective approach [13]. Machine learning often requires the manual extraction of characteristics, whereas deep learning is totally automatic.

A. Research objectives and contributions

Our study aims to explore the current trends in utilizing deep learning (DL) systems for pre-symptomatic identification and precise classification of breast carcinomas through the analysis of mammography images. This research aligns with our broader goal of contributing to ongoing efforts in breast cancer research and diagnosis. Our multifaceted research objectives address key aspects of breast cancer detection and classification using deep learning techniques. Firstly, we aim to provide a comprehensive understanding of breast cancer, its fundamental concepts, and the significance of mammography in early detection and diagnosis. By synthesizing existing literature and analyzing current trends, we seek to elucidate the fundamental components of such systems employed for breast cancer diagnosis, laying the groundwork for subsequent research and development. Additionally, our study endeavours to survey the landscape of deep learning architectures and algorithms commonly utilized in breast malignancy diagnosis systems. Through evaluating their performance and identifying prevailing trends, we aim to offer insights into the potential of deep learning in enhancing diagnostic accuracy and efficiency. A primary contribution of our study lies in compiling publicly accessible datasets containing mammography images, facilitating collaboration and reproducibility in breast cancer research and advancing diagnostic methodologies. Furthermore, we examine the evaluation metrics presently employed to assess the performance of breast cancer diagnosis and detection systems, aiming to identify opportunities for refinement and enhance the rigour and comparability of future studies. Finally, our study highlights existing constraints, difficulties, and avenues for future research in breast cancer diagnosis and classification, aiming to inspire innovative solutions and guide future research directions to improve patient outcomes and reduce the global burden of breast cancer. In summary, our study offers a comprehensive review of current deep learning-based approaches to early breast cancer diagnosis, aiming to inform future research efforts and accelerate progress towards more effective detection and treatment strategies.

B. Research questions

In the attempt to investigate the main objectives as mentioned earlier, our research focused on the following key research questions (RQs), which also served as the guiding framework for this article:

- RQ1: What constitutes breast cancer?
- RQ2: What is the role of breast mammography in this context?
- RQ3: What are the fundamental architectural components of systems employed for breast cancer diagnosis?
- RQ4: What are the prevailing deep learning architectures and algorithms commonly utilized in breast malignancy diagnosis system development?
- RQ5: What are the various publicly accessible datasets containing mammography images?
- RQ6: What evaluation metrics are presently employed to assess the performance of breast cancer diagnosis and detection systems?
- RQ7: What are the existing constraints, difficulties, and avenues for future research in the realm of breast cancer diagnosis and classification?

C. Research methodology

For our research methodology, we conducted a comprehensive review of research studies published in English between 2016 and early 2024. The search process involved a thorough examination of bibliographic literature to identify relevant studies, employing keywords such as 'breast cancer', 'breast tumor', 'breast cancer diagnosis', 'breast cancer detection', 'breast cancer classification', 'breast cancer segmentation', 'mammography', 'mammogram analysis', 'computer-aided detection (CADe) systems for breast cancer', 'computer-aided diagnosis (CADx) systems for breast cancer', 'deep learning', 'deep learning in mammography', and 'convolutional neural networks (CNNs) in mammography'. Searches were conducted across various databases, including PubMed, ArXiv, IEEE Xplore Digital Library, Web of Science, Science Direct, Medline, and Google Scholar. Additionally, statistics pertaining to breast cancer mortality rates were gathered from the Global Cancer Observatory. Our search criteria aimed to include studies where mammography served as the primary modality for breast cancer detection, screening, and diagnosis. Papers were excluded if they incorporated data from screening methods other than mammography, such as Ultrasonography, Computerized Tomography (CT), and Magnetic Resonance Imaging (MRI), or if they did not meet specific criteria such as being written in English, presenting full research papers, utilizing mammographic datasets, or addressing breast cancer detection, segmentation, or classification/diagnosis. This process resulted in a total of 90 papers meeting our inclusion criteria.

D. Paper structure

The remainder of this study is structured in the following order: In Section 2, several fundamental ideas connected to this study are briefly introduced. Section 3 briefly explores the background of cancer diagnosis and detection system architecture and highlights the current State-Of-The-Art (SOTA) for breast cancer diagnosis utilizing deep learning algorithms on mammography. Moving on to Section 4, we compile a summary of publicly accessible mammography datasets. Section 5 enumerates the prevalent evaluation metrics often employed for the experimental assessment of CAD systems in the existing literature. Section 6 then lists the limitations of CAD systems that employ deep learning methods and recommendations for future research. Lastly, Section 7 concludes the paper with a discussion and summary of our findings.

2. Basic concepts

This section presents the following concepts: normal female breast tissue, breast cancer, mammography, and deep learning.

A. Normal female breasts tissue

It's interesting to understand the types of tissue the normal breasts comprise to understand breast cancer diagnosis. In general, female breasts are glandular organs that produce milk. They are located in front of the pectoral muscles that support them. The structure of the female breast is complex and includes fat, glandular, and connective tissue. The breast lobes and breast ducts are parts of the glandular tissue. There are between 15 and 20 lobes in each breast. These lobes split into smaller lobules, each producing several tiny milk-secreting bulbs (alveoli). Milk ducts connect the lobes and lobules that gather the milk. These lead to the areola, which is the nipple in the middle of a pigmented region. Breast tumors frequently start in the lobes and ducts. Additionally, there is fatty tissue in the breast, which fills up the spaces left by the various breast structures and essentially regulates breast size. Doctors refer to all non-fatty tissue as fibro-glandular tissue. Ligaments are bands of elastic connective tissue that go from the skin to the chest wall and provide support. Blood vessels, lymph vessels, nodes, and nerves are found in each breast [14]. With age, the ratio of fat relative to glandular tissue often rises. According to studies, 33% of women aged between 75 and 79 years old and 66% of premenopausal women have breasts that are 50% or more dense [15]. Dense breast tissue, a silent storm within the breast, independently amplifies The vulnerability of developing this complex disease. This increased density poses challenges in breast cancer diagnosis due to its masking effect, which lowers the sensitivity of mammography; it also restricts the evaluation of breast cancer by medical professionals and inhibits the detection of early-stage tumors [15].

B. *Breast cancer*

Breast cancer is an extremely diverse disease that differs from woman to woman in terms of the location of the tumor's origin, its stage of development, how quickly it grows, and its propensity for metastasizing. Breast lumps or tumors are the result of aberrant cells growing out of control and causing breast cancer. To begin therapy, the specialist must be able to identify between two forms of breast cancer during the diagnosis. Both benign and malignant breast tumors fall into these two categories. Because they are less prone to spread, benign lumps are thought to be non-cancerous. Fluid-filled sacs, fibrous glandular tissue, leaf-like growths, abnormal cell overgrowth, lipid tissue death, and glandular tissue changes are a few examples of nodular formations or harmless nodules [16]. Non-invasive breast cancers (also known as in situ), invasive (also known as infiltrating), and metastatic are all examples of malignant tumors, which are cancerous growths [16], [17]. Breast tissue (e.g. lobules, ducts, intermediate tissue) can be the origin of breast cancer. Adenocarcinomas are the most prevalent type of breast cancer. These tumors develop from the epithelial layer of the breast, which is made up of the cells that line the milk-producing lobules and terminal ducts [16]. We speak of lobular carcinoma and ductal carcinoma, respectively. Other forms of malignant breast cancer exist. These cancers are called medullary, papillary, tubular, and mucinous carcinomas. They are much rarer than lobular or ductal cancers. Most often, they are tumors with a good prognosis. When the cancer cells are contained within the lobule or duct, it is called "in situ" cancer. In situ, cancer can progress and invade the surrounding tissues; breast cancer is said to be "invasive." The "in situ cancer" can exist for a long time before evolving into invasive cancer which becomes potentially metastatic, that is to say, capable of releasing cancerous cells to distant sites from the breast through lymph or blood vessels to lymph nodes or other organs in the body and developing new tumors called metastases, these being the main cause of death by breast cancer. When a breast tumor becomes in this stage, it becomes challenging to treat. Hence, the timing of the tumor's diagnosis is one of the key factors in the treatment of breast tumors. The chances of survival can significantly increase, and more effective treatment alternatives can be made available if the disease is discovered early. This underscores the importance of early diagnosis of breast tumors.

C. *Mammography*

Mammography utilizes minimal-dose X-rays, providing a non-invasive diagnostic procedure to look for any breast abnormalities. It is regarded as the most accurate method for diagnosing breast cancer in women, even before symptoms appear. Breast masses and calcifications are the two main abnormalities that can be detected by mammography. Breast lumps can be malignant or non-cancerous; malignant tumors appear in mammograms as irregularly shaped masses with spikes projecting from them. The non-cancerous masses usually have well-defined, circular, or oval borders. [18]. Both macrocalcifications and microcalcifications of the breast can occur [19]. Macrocalcifications, which look like sizable white dots randomly dispersed across the breast on mammography, are considered benign cells. In contrast, in mammography, microcalcifications manifest as minute calcium deposits resembling tiny bright dots and frequently occur in groups. Microcalcification is frequently thought of as the primary sign of early-stage malignancy in the breast or as an indication of the presence of cells at risk of developing into cancer. Every breast is imaged twice using the top-to-bottom (CC) and side-to-side oblique (MLO) projections, as shown in Fig. 1. While the top-to-bottom mammography obtains the image from above, the MLO perspective provides the image from a level that emphasizes the pectoral muscle's side view. Two primary forms of mammography are Film-based mammography and digital mammography (DMM), which are used for different tasks in breast cancer analysis, such as classifying and identifying breast lesions. The three primary subcategories of DMM are contrast-enhanced digital mammography (CEDM), breast tomosynthesis imaging (BTI), as well as comprehensive digital mammography (CDM)[20]. Present practices need a third radiologist to evaluate the mammography if an agreement cannot be reached between the initial two radiologists. This highlights the difficulties even professionals encounter when spotting possible abnormalities in a mammogram.

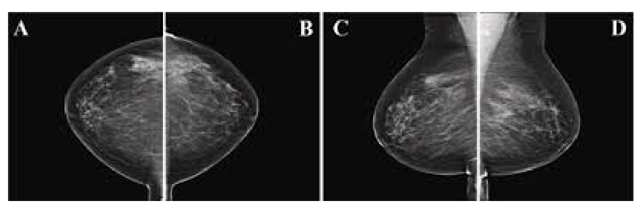

Figure 1. An illustration of the various points of view from a breast mammogram: (A) Right CC view, (B) Left CC view, (C) Right MLO view, and (D) Left MLO view are the four possible views

D. *Deep learning*

Deep learning (DL) automatically derives feature representations from input data [26]. Unlike conventional ML methods, DL has the ability to self-learn these features. In the past, manual feature extraction techniques were employed to isolate and choose features like "colors," "shapes," "edges," and "textures." However, this traditional approach to handcrafted feature extraction is labour-intensive and consumes significant processing time. On the other hand, DL algorithms allow for the automatic extraction of high-level attributes from image data. The availability of extensive datasets enables the use of these algorithms, as they

TABLE I. OPEN ACCESS TO MAMMOGRAPHY IMAGES: A VALUABLE RESOURCE FOR BREAST CANCER RESEARCH. CATEGORIES: TINY DEPOSITS OF CALCIUM IN THE BREAST TISSUE (CAD), ROUNDED DISTINCT LUMPS WITH CLEAR BOUNDARIES (MASS-C), LUMPS WITH IRREGULAR EDGES RESEMBLING SPIKES OR BRANCHES (MASS-S), LUMPS WITH UNCLEAR OR IRREGULAR BORDERS (MASS-I), ABNORMAL CHANGES IN THE BREAST TISSUE PATTERN (ARCH), UNEVENNESS IN THE BREAST TISSUE BETWEEN SIDES (ASYM), NO ABNORMAL FINDINGS DETECTED (NORM), NON-CANCEROUS (BEN), CANCEROUS (MAL), NON-CANCEROUS AND DOES NOT REQUIRE FURTHER (BENWC). DATASET NAMES: THE STUDY DRAWS UPON MAMMOGRAPHIC DATA FROM RENOWNED REPOSITORIES, INCLUDING MIAS, BCDR, DDSM, INBREAST, AND THE CURATED SELECTION OF CBIS-DDSM, COVERING A WIDE SPECTRUM OF BREAST IMAGING FINDINGS).

Dataset Title	Quantity of Images	Type	Categories	Image Presentation	View	Image quality
MIAS [21]	322	FM	CaD, Mass-C, Mass-S, Mass-I, Arch, Asym, Norm, Ben, Mal, BenWC	.PGM	MLO	1024×1024 pixels
CBIS-DDSM [22]	10239	FM	NORM, B, M .	.DICOM	MLO/CC	16 bit
DDSM [23]	10480	FM	B, C, NORM, BWC	.JPEG	MLO/CC	8-16 bit
INbreast [24]	410	DM	B, M, NORM	.DICOM	MLO/CC	14 bit
BCDR [25]	7315 3703 FFDM 3612 FM)	DM	NORM, B, M	.TIFF	MLO/CC	8-14 bit

demand substantial volumes of training data. Deep learning (DL) models acquire hierarchical attributes within the image data domain. DL models are structured with multiple layers that delve into the details of an image, encompassing Low-Level Features, Mid-Level Features, and High-Level Features [27]. The adoption of DL techniques has found application across a spectrum of medical specializations, most notably in radiology and pathology [28]. Deep learning, as an emerging technique, is surpassing traditional machine learning methods and is increasingly integrated into Computer-Aided Diagnosis (CAD) systems [29]. Deep learning techniques have recently showcased their potential in diagnosing breast cancer approximately one year earlier than traditional clinical methods [30]. Convolutional Neural Networks (CNNs) are widely utilized as one of the predominant architectures in deep learning. With enough training data, CNNs can grasp intricate and well-structured hierarchical attributes within an image. They are widely favoured for neural network-based image classification and have demonstrated impressive performance for medical image analysis and categorization [31]. As depicted in Fig. 2, a basic CNN architecture involves integrating one or more layers for convolution and pooling, subsequently complemented by one or more layers that are fully connected [32].

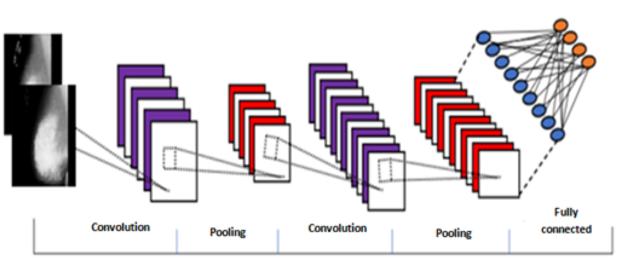

Figure 2. An illustration of a basic CNN architecture used for image diagnosis and classification

After establishing the foundational concepts of normal female breast tissue, breast cancer, mammography, and deep learning in the previous section, it becomes evident that the practical application and advancement of these concepts heavily rely on the availability of suitable datasets. In the following section, we delve into the landscape of publicly available mammography datasets, which serve as crucial resources for training and evaluating models in breast cancer detection and diagnosis.

3. Public available mammography datasets

Within this section, we try to give a succinct summary of the most frequently used publicly accessible mammography datasets for breast cancer detection and diagnosis (e.g., Kaggle, Amazon, UCI ML repository, etc.). These datasets vary in terms of their dimension, visual quality, presentation format, and the technology used to capture the images, including digital mammography (DM) or film mammography (FM), as well as the categories of abnormalities they contain. The MIAS, CBIS-DDSM, DDSM, INbreast, and BCDR datasets are integral to breast cancer research, particularly in advancing deep learning models for diagnosis. The MIAS dataset comprises 322 mammography images, each sized at 1024x1024 pixels, categorized into normal and abnormal classes. Within the abnormal class, further distinctions are made between benign and malignant cases, totalling 115 images. Each image includes details on the type of abnormality detected, such as calcifications, masses, and asymmetries. The CBIS-DDSM dataset, a subset of DDSM, encompasses 10239 mammography images from 1644 patients, ensuring a balanced representation of benign and malignant cases. DDSM, on the other hand, is a substantial database featuring 10480 digitized mammography studies from 2620 patients, spanning both normal and abnormal cases. INbreast presents 410 mammography images from 115 patients, maintaining equilibrium between benign and malignant cases and offering a mix of digital and digitized mammography images. Similarly, BCDR comprises 7315 mammography images, evenly distributed between benign and malignant cases and supplemented with clinical data like patient age and menopausal status. These datasets serve as pivotal resources for various deep learning endeavours in breast cancer diagnosis, including classification, segmentation, and detection. For instance, studies utilizing CBIS-DDSM and MIAS achieved remarkable accuracies of 96.6% and 98.88%, respectively, employing deep convolutional neural networks (DCNN) and improved marine predators algorithm (IMPA) coupled with ResNet50 models [33]. Overall, these datasets play a vital role in shaping the landscape of deep learning applications in breast cancer diagnosis, significantly advancing the field's understanding and capabilities. Table I displays a quick description of these collections.

In the previous section, we explored the landscape of publicly available mammography datasets. It is now crucial to understand how these datasets contribute to the development of systems for mammography-based breast cancer diagnosis. In the coming section, we examine the utilization of deep learning techniques in the creation and optimization of such systems, highlighting their pivotal role in enhancing diagnostic accuracy and efficiency.

4. Systems for mammography based breast cancer diagnosis using deep learning

In this section, we attempt to briefly present the typical CAD system architecture and cover some recent efforts related to DL applications in breast cancer diagnosis using mammography.

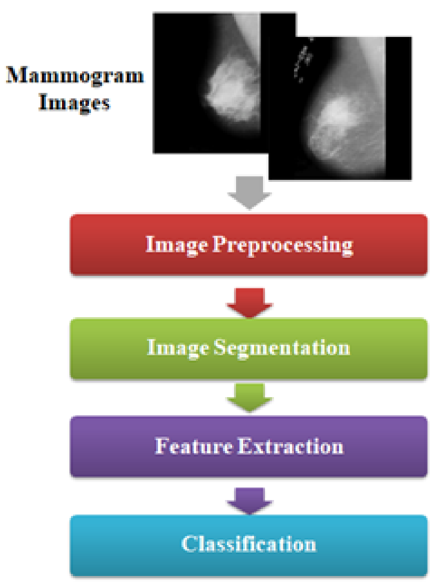

Figure 3. Illustration of the general layout of a CAD system for diagnosing breast cancer

A. CAD in breast cancer: An architectural exploration

CAD systems can differentiate between different tumor types, including mass, calcification, architectural distortion, and asymmetry, as well as classify tumors into two groups: benign or malignant. According to Fig. 3, the general framework of an automated system for diagnosing breast cancer through mammograms commonly consists of four main stages: initial image preprocessing, image segmentation, extraction of relevant features, and classification of lesions. Furthermore, these tools not only significantly reduce the time that human experts spend manually reviewing mammography images but also assist in preventing unnecessary biopsies.

Table II presents the key steps of a CAD system.

B. Deep learning models for mammography based breast cancer diagnosis

Several recent review studies have explored various deep learning methods for mammogram-based breast cancer diagnosis, classification, identification, and segmentation [44], [45], [46], and [47]. This section outlines the main deep learning approaches, encompassing CNNs and RNNs (recurrent networks), and we present transfer learning methods. The CNN comprises several layers where convolutions and max-pooling operations are applied [48]. In a recent study [49], researchers proposed the BMC system for breast mass classification into benign, malignant, and normal categories. This system combines various techniques, including clustering, recurrent network (RNN), Convolutional Neural Network (CNN), and random forest. The researchers conducted model training using the DDSM and MIAS datasets. Their algorithm reaches an accuracy of 96% using DDSM and 95% using MIAS. Another study [50] proposes an algorithm known as CNNI-BCC, which yielded impressive results, including a sensitivity of 89.47%, an accuracy rate of 90.50%, an AUC of 0.901 ± 0.0314, and a specificity of 90.71%. The utilization of this algorithm has the potential to be beneficial in classifying mammogram images into non-cancerous, cancerous, and normal classes, even without previous knowledge about the presence of a cancerous lesion. Moreover, in a different study [51], researchers introduced a Two-perspective mammogram classification model that combines a CNN with an RNN to classify breast masses in mammographic images. Their approach achieved a classification accuracy of 94.7%, recall of 94.1%, and an AUC value of 0.968. In separate research inquiries [52], [53], models based on CNNs were employed to classify mammogram abnormalities. It used the MIAS dataset. CNN-based models have demonstrated encouraging outcomes, improving the accuracy of CAD systems for breast cancer diagnosis. A deep belief network (DBN) is another important DL-based method used for breast cancer classification. It operates as an unsupervised graphical model with generative capabilities. The DBN is a stack of restricted Boltzmann machines (RBM)[54]. It is an effective tool for breast cancer diagnosis for several reasons. They can be used to reduce the input feature vector dimensionality. In

TABLE II. SUMMARY OF THE MAIN STAGES IN THE AUTOMATED DIAGNOSTIC SYSTEM FOR BREAST CANCER USING MAMMOGRAMS.

Stages	Description
Image pre-processing	Most automated image analysis systems depend on the pre-processing stage. Basically, this is done to improve image quality, reduce noise, also remove unnecessary, unwanted artifacts [34]. At this stage, image contrast enhancement methods are used based on the equalization of the histogram and noise reduction techniques (mean and median filters, among others). In addition, other operations can be carried out during the pre-processing phase, covering tasks such as image resizing, data augmentation, and normalization.
Image segmentation	One crucial step is the segmentation procedure. By separating the breast area from the background and emphasizing the suspicious area, also known as the region of interest (ROI) within the larger breast region, breast image segmentation seeks to decrease the impact of the background and facilitate the identification of anomalies within the breast area. The search space for abnormalities is reduced when the backdrop is removed [35]. There are several works approaching breast segmentation by using different methods based on thresholding, active contour, edge-based and region-based, gradient weight map, conditional network, support-pixel correlation, and statistical method.
Feature extraction	Basically, the process of extracting feature sets from mammogram images is employed to classify the considered lesions, specifically to discriminate malignant from benign breast cancer lesions. Generally, three categories of features are utilized in this process: handcrafted features, deep features, and patient-related features, encompassing factors like age and medical history. Handcrafted features encompass a variety of options for extracting information from breast mammograms [36], [37], [38], including texture, morphological aspects, and descriptors, as well as shape, intensity, and hybrid features. Additional possibilities include curvelet-based statistical features and local and global features. Other methods involve using histogram of gradients (HOG), SIFT, and wavelets. Moreover, features encompass contrast, geometrical aspects, location data, context, and patient-related information. Deep feature extraction is an entirely automated process that employs deep learning-based models to automatically extract high-level features by utilizing convolutional layers. Various architectures like deep CNN with transfer learning is proposed for feature extraction.
Classification	Following feature extraction, the final phase involves the breast lesions classification [39], [40], contributing to the categorization of mammograms and aiding medical decision-making through the utilization of the extracted features within an effective classification model. Various classification approaches apply to classifying breast cancer tissue, primarily including binary classes classification (distinguishing between cancerous and non-cancerous), multiple classes classification (encompassing categories like healthy tissue, non-cancerous lesions, in situ malignancy, and invasive malignancy), and the one-class classification (OCC) approach. Statistical ML-based Classifiers and DL-based Classifiers are the two basic classification model types used to diagnose breast cancer. Pathologists and doctors can utilize artificial intelligence-based algorithms to diagnose breast cancer to aid in their decision-making. Statistical machine-learning techniques are commonly used for the classification of breast cancer images. Convolutional networks are one of the most effective models for image analysis. There are several DL architectures based on pre-trained models such as AlexNet [41], VGG-16, ResNetXt50 [42], and Google Inception-V3 architecture [43].

[55], a novel and efficient CAD system was introduced, incorporating DBN. This system was designed to categorize mammographic masses into four evaluation sorts based on the BI-RADS classification, including not harmful (2), likely harmless (3), suspicious (4), and extremely suspicious (5). Trained on 500 DDSM images, the model reached 84.5% accuracy. Creating systems that accurately identify lesions in mammography images holds significant value for healthcare professionals. Consequently, researchers in [56] devised a system for mass detection utilizing the Faster R-CNN framework. The INbreast dataset and CBIS-DDSM (curated breast imaging subset of the DDSM) were used to evaluate the approach's performance. The study's findings showed that the true positive rate for CBIS-DDSM was 0.9345, with 2.2805 false positives per picture, while the true positive rate for INbreast was 0.9554, with 0.3829 false positives per image. The You Only Look Once (YOLO) detector has greatly enhanced classification model performance, resulting in encouraging breast lesion diagnosis outcomes [57]. The YOLO effectiveness was assessed in [58] for detecting lesions in the breast. Subsequently, they made modifications to and evaluated a traditional Multi-Layer Perceptron, 50-Layer Convolutional Model, as wellInception ResNet Version 2 (InceptionResNet-V2). The architectures were subject to evaluation using the DDSM and INbreast datasets. The detection reached the accuracy of 99.17% for DDSM and 97.27% for INbreast, along with F1-scores of 99.28% and 98.02%, respectively. For the classification in DDSM, the three models reached accuracies of 94.50%, 95.83%, and 97.50%, and for the INbreast dataset, 88.74%, 92.55%, and 95.32%. In addition, Samuel et al. [59] devised a model aimed at aiding radiologists in mass

screening for breast abnormalities and prioritizing patients. Their approach integrates an ensemble of EfficientNet-based classifiers with YOLOv5, a method for detecting suspicious masses, to identify abnormalities. Incorporating YOLOv5 detection is pivotal for explaining classifier predictions and enhancing sensitivity, especially in cases where the classifier fails to detect abnormalities. To further improve the screening process, the researchers also introduced an abnormality detection model. The classifier model achieves an F1-score of 0.87 and a sensitivity of 82%. By integrating suspicious mass detection, sensitivity increases to 89%, albeit with a slightly lower F1-score of 0.79. For dataset construction, the study utilized two primary sources. The first source, VinDr-mammo [60], comprises approximately 5000 studies of four-view mammography exams, providing breast-level assessments and finding annotations. The second source includes a locally prepared dataset consisting of 3123 breast scans from 1028 patients. Additionally, the Mini-DDSM [61] dataset was utilized for model evaluation, containing 679 CC and MLO scanned breast mammography views from 679 unique cancer cases, along with 2408 images from 602 unique patients with normal mammography readings. Transfer learning has become a widely adopted technique, and it addresses the challenge of insufficient data, particularly when dealing with small datasets. Additionally, it offers advantages such as reduced computational costs and shorter model training times [62]. Recent studies have increasingly embraced this approach, as an example, authors in [63] introduced a modified AlexNet architecture for mammogram classification of masses as benign/malignant. MIAS testing yielded 95.70% accuracy for the final model. In their study, researchers [64] devised an effective Deep Learning Architecture (DLA) coupled with a Support Vector Machine (SVM) for diagnosing breast cancer using mammograms. They utilized the state-of-the-art Visual Geometric Group (VGG) architecture with 16 layers to extract dominant features crucial for breast cancer classification. Subsequently, SVM was integrated into the output layer to enhance classification outcomes. The results indicate that the VGG-SVM model exhibits significant potential for classifying images from the Mammographic Image Analysis Society (MIAS) database, achieving an accuracy of 98.67%, sensitivity of 99.32%, and specificity of 98.34%. Another study [65] employs MobileNetV2, a low-computational Deep Convolutional Neural Network (Deep-CNN) model, for binary classification of mammography images into malignant or benign categories. Two methods, transfer learning and scratch learning, are explored. MobileNetV2 is employed with three transfer learning variants, including transfer learning without fine-tuning, transfer learning with fine-tuning, and fixed feature extraction, to achieve binary classification, and a seven-layer CNN architecture is developed to accomplish mammography image classification through scratch learning. The best results are achieved by combining MobileNetV2 with a Random Forest classifier using fixed feature extraction, yielding an accuracy of 99.4% in 63.87 seconds compared to the other transfer learning variants. On the other hand, a seven-layer CNN model developed using scratch learning achieved a classification accuracy of 96%, but it required a longer training period of 7980 seconds. Utilizing MobileNetV2 and Random Forest reduces trainable parameters and training time, making it suitable for low-cost embedded platforms. The researchers conducted model training using mammography images from the DDSM dataset. Additionally, a hybrid technique for the quick and precise classification of breast cancer using mammogram images was proposed and tested in a separate study [66]. This approach employed three different Deep Learning (DL) Convolution Neural Network (CNN) models, including Inception-V3, ResNet50, and AlexNet as feature extractors. The effectiveness of this method for classification was evaluated using the MIAS image database. Results indicate that the average classification accuracy was 97.81% for 70% of training data, 98% for 80% of training data, and reached its optimal value for 90% of training data. In [67], DL models, including AlexNet, ResNet18, and ResNet34, were utilized for accurate detection and characterization of microcalcifications on mammography. The dataset employed for model evaluation comprised 1,986 mammography images from 1,000 patients aged 21 to 73 years. This dataset included 611 benign lesions and 389 histologically proven breast cancers. AlexNet demonstrated superior performance, achieving a sensitivity of 98%, specificity of 89%, and AUC of 0.98 for microcalcifications detection and a sensitivity of 85%, specificity of 89%, and AUC of 0.94 for microcalcifications classification. For microcalcifications detection, ResNet18 and ResNet34 achieved sensitivities of 96% and 97%, specificities of 91% and 90%, and AUCs of 0.98 and 0.98, respectively. Regarding microcalcifications classification, ResNet18 and ResNet34 exhibited sensitivities of 75% and 84%, specificities of 85% and 84%, and AUCs of 0.88 and 0.92, respectively. Additionally, a separate study [68] presents a breast X-ray mammography image classification model utilizing Convolutional Neural Networks (CNN). The model distinguishes between benign and malignant classes in mammography images, modifying the VGG 16 network architecture and conducting experiments on datasets sourced from the Medical Imaging Department of Ganzhou People's Hospital and The Sixth Affiliated Hospital of Jinan University. Experimental findings highlight the model's exceptional classification capabilities, achieving an average accuracy rate of 96.945%. Moreover, in an alternate study [69], researchers utilized CNN-based pre-trained architectures such as modified VGGNet and SE-ResNet152 to improve their ability to differentiate between normal and suspicious mammography regions. They also employed hybrid deep neural network approaches, including CNN+LSTM and CNN+SVM, for breast lesion classification. These models were trained using mammogram images sourced from both public and private databases. Their algorithms exhibited a sensitivity of 99% and an overall AUC of 0.99, representing substantial enhancements in mammogram analysis. In another investigation [70], researchers introduced a Mammo-Light convolutional neural network (CNN) model designed for mammogram classification. This model features a reduced number of layers and parameters compared to conventional CNNs. The effectiveness of the proposed approach was evaluated using two widely used publicly accessible datasets, CBIS-DDSM and MIAS. Mammo-Light achieved test accuracies of 99.17% and 98.42%, respectively, for the CBIS-DDSM and MIAS datasets, surpassing the performance of ten state-of-the-art transfer learning (TL) models in terms of accuracy and other evaluation metrics. Notably, Mammo-Light demonstrated exceptional performance with fewer parameters and computational time due to its lightweight design, potentially contributing to advancements in early breast cancer diagnosis and facilitating prompt treatment. Researchers in a study [71] introduced a semi-automatic real-time detection approach employing deep learning to differentiate between microcalcifications and masses within a breast cancer dataset. The primary objective was the detection of microcalcifications, which may act as precursors to breast cancer. The proposed architecture utilized SAE (Hierarchically Stacked Autoencoders). The SAE model utilized a training technique based on a greedy search to extract low-level characteristics linked to microcalcifications. The approach encompassed two scenarios: (1) identifying microcalcifications and (2) simultaneously identifying microcalcifications and masses. Their method shows a discriminative accuracy in distinguishing calcifications using the SVM classifier. In an independent study [72], researchers introduced a DL approach to handle the availability of limited and imbalanced data. The approach employed an infilling technique to generate synthetic mammogram patches using generative adversarial networks (GAN). First, a multiscale GAN generator was trained to produce synthetic elements within the designated image. This generator used a refinement process to produce multiscale features and guarantee stability at higher resolutions. Importantly, the GAN was confined to infill exclusively lesions, including both masses and calcifications. A ResNet-50 classifier was employed to assess the quality of the images generated. The study compared the classification performance of data enhancement using GANs and traditional methods, revealing that synthetic augmentation enhances classification accuracy. Lately, several investigations have adopted the End-To-End (E2E) training approach, which has demonstrated promising outcomes for breast cancer detection [73]. In this context, researchers in [74] introduced a CNN model based on an E2E training strategy. The primary objective is to label mammographic images as normal or malignant. The proposed model is based on two components: contextual features and classification. It utilizes a Multi-level CNN for deep high and low-level feature extraction. The experiments achieved 96.47% of accuracy and a 0.99 AUC score using the mini-MIAS dataset. Full-field digital mammography (FFDM) is a common screening method for breast cancer (BC). Still, its effectiveness in dense breast regions is limited, reducing BC detection sensitivity due to mammographic breast density (MBD). Contrast-enhanced spectral mammography (CESM), a newer FFDM variant, offers enhanced contrast resolution, facilitating lesion assessment and detecting multicentric and multifocal lesions. CESM is increasingly used for a comprehensive investigation of suspicious cases and as a cost-effective alternative to MRI in certain clinical scenarios, providing morphological information comparable to FFDM. Comparative studies have shown CESM's superiority in background suppression, reduced background parenchymal enhancements (BPE), and higher positive-predictive values (PPV) compared to MRI. Additionally, CESM shows promise in classifying microcalcifications, determining cancer volume, and facilitating precise biopsy localization. For instance, Zheng et al. [75] developed a fully automated pipeline system (FAPS) using contrast-enhanced mammography (CEM) to segment and classify breast lesions. The model, combining RefineNet and Xception + Pyramid pooling module (PPM), was evaluated on a dataset of 1912 women with single-mass breast lesions. Results showed FAPS achieved Dice similarity coefficients (DSCs) of 0.888 ± 0.101, 0.820 ± 0.148, and 0.837 ± 0.132 for segmentation in internal, pooled external, and prospective testing sets. For classification, FAPS achieved AUCs of 0.947, 0.940, and 0.891, outperforming radiologists in classification efficiency (6 seconds vs. 3 minutes). The study demonstrates FAPS's potential for segmenting and classifying breast lesions with high efficiency and generalization. The research conducted in [76] explores the feasibility of a computationally efficient computer-aided diagnosis (CAD) system for breast lesion classification using contrast-enhanced spectral mammography (CESM). Additionally, the synthesis of contrast-enhanced (SynCESM) images is investigated to eliminate the need for intravenous contrast agents. The study collected 504 pairs of low-energy (LE) and CESM images from 160 female subjects. Lesion segmentation was performed using a semi-automatic active-contour method, followed by feature extraction. To enhance computational efficiency, the wavelet packet transform (WPT) was applied. Results showed that using LE images, a sigmoid-kernel SVM classifier achieved 90.20% accuracy, while CESM images yielded 93.26% accuracy. Interestingly, SynCESM images still provided reasonable performance with 92.14% accuracy. The combination of LE and CESM images resulted in the best performance, with 96.87% accuracy. The proposed system demonstrated clinical feasibility, lower complexity, and reduced reliance on contrast agents through synthetic data generation. Deep Learning models have emerged as indispensable tools in addressing complex challenges like cancer detection, owing to their exceptional ability to process vast datasets with precision and efficiency. Recent strides in medical research emphasize the significance of identifying molecular subtypes in breast cancer, crucial for tailoring personalized treatment strategies due to varied responses to different therapies. For instance, in [77], MOB-CBAM, a novel lightweight dual-channel attention-based deep learning model, was introduced for precise breast cancer detection and subtype prediction. By leveraging the MobileNet-V3 architecture and incorporating a Convolutional Block Attention Module (CBAM), MOB-CBAM showcased remarkable performance in discerning various breast cancer features, including masses, calcifications, and molecular subtypes such as Luminal A, Luminal B, HER-2 Positive, and Triple Negative. Through rigorous evaluations on the Chinese Mammography Dataset (CMMD) [78], MOB-CBAM demonstrated outstanding accuracy in both coarse-grained and fine-grained classifications. Notably, coarse-grained classification achieved 99% accuracy, while fine-grained tasks like mass and calcification identification boasted an impressive 98% accuracy rate. Further validation through cross-validation on MIAS and CBIS-DDSM datasets affirmed MOB-CBAM's effectiveness, with accuracies of 97% and 98%, respectively. This study underscores MOB-CBAM's potential as a reliable tool for breast cancer diagnosis and subtype prediction, offering enhanced precision through its innovative attention mechanism. Furthermore, Panambur et al. [79] explored the classification of luminal subtypes in full mammogram images. They utilized transfer learning from a breast abnormality classification task to fine-tune a ResNet-18-based model for distinguishing between luminal and non-luminal subtypes. Results obtained from the CMMD dataset showcased substantial enhancements over the baseline, achieving a mean AUC score of 0.6688 and a mean F1 score of 0.6693 on the test dataset.

Table III summarizes studies focusing on breast cancer diagnosis from mammogram images using DL techniques, along with their respective performance metrics.

Existing review articles on breast cancer diagnosis-based deep learning techniques using mammography have significantly advanced the field by offering thorough insights into diverse methodologies, datasets, and challenges. These reviews meticulously analyze individual studies, delving into the architectures, datasets, and performance metrics employed, thus providing a nuanced understanding of each approach's strengths and limitations. Moreover, they frequently delve into the datasets utilized across different studies, encompassing both publicly available datasets like MIAS and CBIS-DDSM, as well as proprietary ones, which are crucial for validating findings. Despite these contributions, certain limitations persist. Some studies have narrowly focused on binary classification tasks, neglecting more nuanced classifications. Additionally, the variability in DL model performance across datasets underscores the need for robust validation across diverse datasets. Furthermore, existing reviews often concentrate on specific aspects, such as microcalcification detection or mass classification, limiting the broader perspective needed for comprehensive benchmarking. Moreover, discrepancies in reporting performance metrics like accuracy, sensitivity, specificity, and AUC pose challenges in identifying the most effective approaches for breast cancer diagnosis. Another supposed problem with this analysis is that the researchers do not compare the results of the classifier with those collected by the clinician to determine whether the classifier is more reliable. Additionally, several publications lack explicit disclosure of experimental approaches, complicating reproducibility and evaluation. Addressing these issues is imperative to foster advancements in breast cancer diagnosis and enhance the efficacy of deep learning techniques in clinical settings. In this review article, we endeavour to enhance the existing body of work by offering a comprehensive examination of deep learning (DL) methods employed in mammography-based breast cancer diagnosis, classification, and segmentation. By synthesizing a diverse array of tasks and the latest emerging techniques, our aim is to furnish readers with insightful comparisons and a more exhaustive presentation of available details. Our study encompasses a broad spectrum of DL architectures, ranging from Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to Deep Belief Networks (DBNs), YOLO, and pre-trained models like VGG and MobileNet. Additionally, we delve into transfer learning and end-to-end training approaches, ensuring a thorough exploration of the deep learning landscape in breast cancer diagnosis. Furthermore, our review expands its purview to include various tasks such as classification (normal/cancerous/non-cancerous), segmentation (lesion boundaries), detection (mass/calcification), and even subtype prediction, providing a holistic view of the applications of DL in this domain. Moreover, we emphasize the diverse performance metrics utilized in the studies, ranging from accuracy and sensitivity to specificity, Area Under the Curve (AUC), and F1-score. This comprehensive approach allows for a nuanced understanding of the effectiveness of different models and methodologies.

With an understanding of the development and utilization of deep learning systems for mammography-based breast cancer diagnosis, it is essential to establish robust evaluation metrics to assess their effectiveness and reliability. In the following section, we delve into the various metrics employed to evaluate the performance of these systems, providing insights into their accuracy, sensitivity, specificity, and other key measures crucial for assessing diagnostic efficacy.

5. Evaluation metrics

This section presents the evaluation measures employed to assess the performance of methods for diagnosing and detecting breast cancer.

A summarized overview of the calculation formulas and explanations for the most frequently employed evaluation metrics in the literature can be found in Table IV.

To calculate various evaluation metrics, several key terms are employed:

- True Negative (TN): Cases where both the actual and predicted outcomes are negative.
- True Positive (TP): Cases where both the actual and predicted outcomes are positive.
- False Negative (FN): Cases where the actual outcome is positive, but the prediction is negative (missed

TABLE III. COMPREHENSIVE ANALYSIS OF STATE-OF-THE-ART RESEARCH IN BREAST CANCER DIAGNOSIS ON MAMMOGRAPHY UTILIZING DEEP LEARNING. ACCURACY (ACC); SENSITIVITY (Sn); SPECIFICITY (Sp); AREA UNDER THE CURVE (AUC); DETECTION (DET); SEGMENTATION (SEG); CLASSIFICATION (CLA).

Paper	Year	Application	Model	Dataset	Evaluation Metric
[49]	2021	Classification	k-means + LSTM + RNN CNN + Random Forest	DDSM MIAS	DDSM: Acc=96% MIAS: Acc=95%
[50]	2019	Classification	CNNI-BCC	MIAS	Sn = 89.47% Acc = 90.5% AUC = 0.90 Sp = 90.7%
[51]	2021	Classification	CNN-RNN	DDSM	Acc = 94.7% Recall = 94.1% AUC = 0.968
[52]	2021	Classification	GNN + CNN	MIAS	Acc = 96.1%
[53]	2021	Classification	CNN with Knowledge transfer	MIAS	Acc = 98.87% F1-score = 99.3%
[55]	2020	Classification	DBN	DDSM	Acc = 84.5%
[56]	2020	Mass Segmentation	Faster R–CNN	CBIS-DDSM INbreast	CBIS: TP = 0.93 INbreast: TP = 0.95
[80]	2022	Lesions Segmentation	CNN	DDSM	Dice = 65%
[58]	2020	Lesions Segmentation Classification	YOLO CNN ResNet-50 InceptionResNet-V2	DDSM INbreast	SEG DDSM: F1-score = 99.28% SEG INbreast: F1-score= 98.02% CLA DDSM: Acc = 97.5% CLA INbreast: Acc = 95.32%
[59]	2024	Detection	YOLOv5	VinDr-mammo Private dataset Mini-DDSM	F1-score = 87% Sn = 82 %
[63]	2020	Classification	AlexNet (Augmentation)	MIAS	Acc=95.70%
[64]	2022	Classification	VGG16 + SVM	MIAS	Acc = 98.67% Sn = 99.32% Sp = 98.34%
[65]	2023	Classification	MobileNetV2 + Random Forest CNN	DDSM	MobileNet Acc=99.4% CNN Acc=96%
[66]	2023	Classification	Inception-V3 + ResNet50 + AlexNet	MIAS	Acc = 97.81% (70% training) Acc = 98% (80% training) Acc ~optimal value (90% training)
[67]	2023	Microcalcifications Detection and classification	AlexNet ResNet18 ResNet34	Private	DET AlexNet Sn = 98% Sp = 89% AUC = 0.98 DET ResNet18 Sn = 96% Sp = 91% AUC =0.98 DET ResNet34 Sn = 97% Sp = 90% AUC =0.98 CLA AlexNet Sn = 85% Sp = 89%, AUC = 0.94 CLA ResNet18 Sn = 75% Sp =85% AUC =0.88 CLA ResNet34 Sn = 84% Sp =84% AUC =0.92
[68]	2024	Classification	VGG16	Private	Acc = 96.945%
[69]	2024	Classification	CNN VGGNet SE-ResNet152	Public and private	Sn = 99% AUC = 0.99
[70]	2024	Classification	Mammo-Light CNN model	CBIS-DDSM MIAS	Acc = 99.17% (CBIS-DDSM) Acc = 98.42% (MIAS)
[71]	2016	Microcalcification Detection and classification	Stacked autoencoder	Private	Acc =87%
[72]	2018	Microcalcification Detection and classification	GAN + ResNet50	DDSM	AUC = 0.896
[74]	2020	Classification	CNN	mini-MIAS	Acc = 96.47% AUC = 0.99
[75]	2023	Segmentation and classification	RefineNet Xception with PPM	Private	SEG DICE = 0.888 CLA AUC = 0.947
[76]	2023	Classification LE + CESM Images	SVM	Private	LE Acc= 90.20% CESM Acc=93.26% SynCESM images: Acc=92.14% , LE + CESM images Acc =96.87%
[77]	2024	Detection and subtype prediction	MobileNet-V3 with CBAM	CMMD MIAS CBIS-DDSM	DET Acc=99% SubType Acc=98% MIAS Acc=97% CBIS-DDSM Acc= 98%
[79]	2023	Classification	ResNet-18 (transfer learning)	CMMD	Mean AUC = 0.6688 Mean F1score=0.6693

positives).

- False Positive (FP): Cases where the actual outcome is negative, but the prediction is positive (false alarms).

Having examined the evaluation metrics used to assess the performance of mammography-based breast cancer diagnostic systems, it's imperative to acknowledge the challenges encountered in their implementation and consider the future directions of research in this domain. The subsequent section delves into the various obstacles faced and outlines potential avenues for advancement, emphasizing the need for innovative solutions to address emerging challenges and propel the field forward.

6. Challenges and future directions

This section discusses some of the challenges and research directions in DL-based systems diagnosing breast malignancies. Effectiveness in utilizing DL systems for diagnosing and detecting malignancies in the breast might be greatly impacted by the limited data problem in medical imaging analysis. A number of models have been put up to use X-ray mammography pictures to automate the diagnostic procedure for breast cancer. Many researchers have trained their deep learning architectures using publicly available breast imaging datasets. On the other hand, it is commonly recognized that DL architectures need a large quantity of training data. Regretfully, in order to train these models successfully, many of the current existing publicly accessible datasets, including MIAS and INbreast, might need to be improved. Training on datasets this tiny, usually only a few hundred samples in size, may cause problems like overfitting. In the existing literature, two commonly adopted approaches are employed to tackle the issue of limited data and enhance the robustness and accuracy of such a proposed DL model. The primary and widely used method involves expanding the training dataset size through data augmentation, which generates multiple slightly altered versions of the original images. This data augmentation technique encompasses various methods, such as rotating images within specific angle ranges, adjusting image sizes within specified factors, shifting and flipping images in different orientations, cropping images, and producing images with transformed shapes and intensities using various techniques. When all augmented image versions are pre-generated and integrated with the original dataset before the training, Offline data augmentation. The model then utilizes this dataset in randomized Mini-Batches during training. Conversely, Online augmentation is designed to execute operations (e.g., affine transformation) as part of the DL model pipeline. Users may configure the input parameters for each form of augmentation in this arrangement, including the likelihood and range. This way, every picture in a Mini-Batch is randomly altered according to the given probabilities, using the initial training set as input. The selection between Offline and Online approaches for augmentation is based on the dataset size. Offline augmentation is the preferred choice for smaller datasets, while Online augmentation is better suited for larger datasets, particularly if the augmentation process can be implemented on a GPU. It's worth noting that Offline augmentation demands more memory, while the Online approach consumes more computational time. Extensive research has demonstrated that data augmentation effectively mitigates the risk of overfitting when dealing with small training sets, as evidenced by studies like [82] and [83]. Another effective strategy involves utilizing transfer learning, which has demonstrated significant success in analyzing mammography images, as exemplified in [84]. Initially undergo training on extensive image datasets from a diverse range of domains, essentially encompassing any general imaging dataset. Subsequently, these models undergo refinement using a dataset specific to breast images, which typically pertains to the targeted domain. ImageNet frequently employs a general imaging dataset for this purpose [85], serving as a foundational resource. Numerous deep models based on transfer learning have undergone pre-training on this dataset, including VGG-16, ResNet, Inception-V3, and others. Moreover, a significant limitation observed in mammography datasets for breast cancer diagnosis pertains to the substantial imbalance between negative and positive classes. Specifically, breast mammography image datasets, as evidenced in [86], exhibit a pronounced class imbalance, with approximately 97% of examples belonging to the negative class and only around 3% representing the positive class. An ideal classification scenario would entail a balanced rate that achieves equivalent accuracy in predicting both the majority and minority classes within the dataset, ideally reaching 100% accuracy for both. However, practical classification outcomes reveal a substantial imbalance, with precision rates of 100% for the majority class and ranging from 0% to 10% for the minority class. To put this into perspective, a 10% precision rate for the minority class implies that 2% of patients with cancer may be erroneously classified as noncancerous. In the medical domain, such an error is considerably more costly than classifying a cancerous patient as noncancerous. Imbalanced datasets, particularly in terms of class distribution, are a recurring challenge encountered in addressing real-world classification scenarios like breast cancer diagnosis. Imbalanced datasets are characterized by a skewed class distribution, where one or more groups have a significantly larger number of examples than others. In medical diagnosis datasets, it's common to have an imbalance, where there are many more instances of benign (normal/healthy) cases recorded than malignant (abnormal/cancerous) cases. When a dataset exhibits such an unequal distribution, it tends to be biased toward the majority class, which may not be of primary interest. Consequently, when deep learning algorithms are trained on imbalanced datasets, they also tend to be biased by the majority class. This poses significant challenges in learning from severely imbalanced datasets, a topic called imbalanced learning. For instance, in a simulation study conducted to address this issue [87], researchers examined how well a CNN could classify breast masses into malignant or benign categories. They used a

TABLE IV. EVALUATION METRICS COMMONLY USED FOR BREAST CANCER DIAGNOSIS [81].

Metrics	Description	Formula
Accuracy (Acc)	It is computed by taking the proportion of correct predictions and dividing it by the overall predictions generated. Essentially, provides insight into the proportion of the model's predictions that were accurate.	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$ (1)
Precision (Pr)	It evaluates the correctness of the positive predictions. It is computed by dividing the number of true positive results by the total number of actual positive cases, which includes both correctly identified cases and those erroneously labelled as positive by the classifier.	$Pr = \frac{TP}{TP+FP}$ (2)
Sensitivity (Sn) Recall (R) TPR	It quantifies the fraction of real positive instances that the classifier should have accurately identified as true positives. Maintaining high values for both Sn and Pr is essential in medical image diagnosis to reduce the chances of misdiagnosing patients with malignancies.	$Sn = \frac{TP}{TP+FN}$ (3)
Specificity (Sp)	It is calculated by considering the ratio of accurately identified instances from the negative class to the overall count of negative instances.	$Sp = \frac{TN}{TN+FP}$ (4)
F1-score	It is typically employed when dealing with imbalanced datasets, especially those with significant class imbalances. It assesses the model's accuracy for each class and is calculated based on precision and recall.	$F1 - score = \frac{2 \times R \times Pr}{R+Pr}$ (5)
ROC-AUC (FPR)	Receiver Operating Characteristics (ROC) curve holds significance as a vital performance measure for CAD systems, depicts the relation between True-Positive Rate (TPR) and False-Positive Rate (FPR) across various decision points. The Area Under the ROC Curve (AUC) indicates the system's capability to differentiate between positive and negative classes.	$FPR = \frac{FP}{FP+TN}$ (6)

potentially corrupted training set, with corruption levels ranging from 0% to 50% of samples. The findings showed that although classification performance might reach 100% on the training set, as the degree of training label corruption rose, it became less effective when applied to unseen test samples. In the literature, two frequently employed methods are discussed to address the aforementioned issue, namely oversampling and undersampling. Some studies suggest that in the case of oversampling, there is a potential risk of overfitting [88], which could affect model generalization. Conversely, another study [89] has indicated that undersampling may be more effective than oversampling, but it does come with the drawback of discarding valuable samples from the dataset. In addition to these well-established techniques, recent years have seen the emergence of the One-Class Classification (OCC) technique, particularly in identifying abnormal samples in comparison to known class instances. This approach offers a promising solution to address challenges associated with severely imbalanced datasets [90], which are particularly prevalent in large-scale data scenarios. While conventional classification techniques, whether binary or multi-class, aim to assign a data object to one of the several existing classes, there is an approach that aims to determine whether a data instance belongs to a specific class or not, named One-Class Classification (OCC). It trains the model exclusively on samples from a single class, referred to as the target class, and treats all other samples as outliers. This approach proves valuable in situations where samples from other classes are either scarce or entirely unavailable. Such scarcity of samples can arise from various factors, including the challenges associated with data collection, high computational requirements, rare events, and more. Consequently, it is suggested that future research endeavours should consider the utilization of Deep Learning-based One-Class Classification models to increase the accuracy of a cancer diagnosis on breast images. Moreover, most research articles focus largely on the accuracy measure when evaluating the performance of their model, frequently ignoring other important aspects. This approach proves inadequate because the accuracy metric fails to differentiate between errors in the positive and negative classes specifically. It is recommended that forthcoming studies incorporate, at the very least, AUC and F1 Scores as part of their evaluation criteria to gauge the effectiveness of each model comprehensively. A notable discovery is that there is currently a limited number of DL models that integrate clinical Data (such as patient age, menopausal status, medical history, etc.) with image data. Future researchers may find it worthwhile to conduct additional investigations and develop more hybrid algorithms utilizing DL-based approaches that merge clinical data with image data.

As we reflect on the challenges and potential avenues for advancement in mammography-based breast cancer diagnosis, it becomes evident that addressing these challenges is paramount for the continued progress of the field. In the final section, we consolidate our findings, providing a comprehensive discussion of the implications of our research and offering concluding remarks on the current state of the art and avenues for future exploration.

7. Discussion and conclusion

Breast cancer, a prevalent malignancy among women globally, necessitates accurate diagnosis for tailored treatments that enhance outcomes and survival rates. Collaborative efforts within the medical community aim to innovate early diagnosis technologies for breast cancer. However, the lack of interdisciplinary information and understanding among medical professionals, researchers, scientists, and healthcare workers poses challenges in developing novel diagnostic tools. This review aims to impart crucial insights into breast cancer, emphasizing the importance of multidisciplinary collaboration to drive the creation of more efficient diagnostic tools. While existing reviews predominantly focus on DL-based methods for breast cancer diagnosis, lacking comprehensive disease explanations, this review distinguishes itself by consolidating information on normal female breast tissue anatomy, core breast cancer concepts, and established diagnostic modalities like mammography. Our study aims to advocate for adopting deep learning approaches among field specialists, including medical professionals, researchers, scientists, and healthcare workers, for accurate breast cancer diagnosis and evaluation. By providing an overview of various deep learning algorithms employed for breast cancer detection, segmentation, and classification using mammography, this article contributes to bridging the gap between traditional diagnostic methods and cutting-edge technologies. In recent years, the field of deep learning for diagnosis has experienced significant advancements, with numerous works contributing to the state-of-the-art. This paper presents a comprehensive review of the latest developments, offering novel insights and contributions that distinguish it from existing literature. Our review comprehensively explores a range of recent deep learning architectures utilized in breast cancer diagnosis. This includes supervised models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Recurrent Convolutional Neural Networks (RCNN), and You Only Look Once (YOLO), as well as specific models like CNNI-BCC End-to-End CNN. Additionally, we delve into unsupervised models like stacked autoencoders and Generative Adversarial Networks (GAN). We also examine the relevance of pre-trained models such as ResNet 50, Inception V2, and AlexNet in the context of breast cancer diagnosis. Furthermore, machine learning classification techniques employed in breast cancer diagnosis are discussed. The review delves into various stages of image processing, including image pre-processing, segmentation, feature extraction, and classification. These are crucial for analysing mammography images for breast cancer diagnosis and detection. Our analysis, based on the reviewed studies, highlights a significant trend where most existing studies focus on various modalities-based deep learning approaches for breast cancer diagnosis, such as Ultrasonography, Computerized Tomography (CT), and Magnetic Resonance Imaging (MRI). In contrast, there is a notable gap in the comprehensive and systematic analysis of mammography-based deep learning techniques, with limited coverage in the literature. This review uniquely centres solely on mammography for breast cancer diagnosis, recognizing it as the most commonly utilized imaging modality for assessing breast abnormalities and emphasizing its effectiveness in early breast cancer detection. While numerous researchers have made substantial contributions to the field of breast cancer diagnosis, this study aims to provide valuable insights to radiologists and the medical community, facilitating the development of new and efficient diagnostic techniques for early-stage breast cancer detection. Moreover, the paper provides a summary of the most frequently used publicly accessible mammography datasets to facilitate future research. It explores the most widely used metrics for evaluating computer-aided breast cancer detection and diagnosis systems. The findings of the current review underscore the Convolutional Neural Network (CNN) as the most accurate and prevalent model for breast cancer detection, with accuracy metrics emerging as the predominant method for performance evaluation. CNNs have gained widespread adoption due to their capability to extract intricate features from mammographic images effectively. For instance, research by Malebary et al. [48] introduced the BMC system, RNN, CNN, and random forest techniques for breast mass classification, achieving notable accuracies of 96% and 95% on DDSM and MIAS datasets, respectively. Similarly, the CNNI-BCC algorithm proposed by Ting et al. [49] exhibited high sensitivity, accuracy, and AUC values, showcasing its potential for classifying mammogram images into non-cancerous, cancerous, and normal categories. Furthermore, the integration of CNNs with RNNs in the Two-perspective mammogram classification model [50] yielded promising results, achieving a classification accuracy of 94.7%. These studies underscore the efficacy of DL models in accurately identifying breast abnormalities from mammography images. DL techniques have also been found to be applicable to specific tasks such as mass detection and lesion classification. For instance, researchers in [55] developed a Faster R-CNN-based system for mass detection, demonstrating high true positive rates and low false positive rates. Similarly, the YOLO detector [56] exhibited promising results in breast lesion diagnosis, achieving high accuracy and F1-scores.

Transfer learning has emerged as a valuable approach to mitigate data scarcity issues in DL models. Studies such as [62] and [63] leveraged transfer learning with modified AlexNet and MobileNetV2 architectures, respectively, for mammogram classification, achieving accuracies exceeding 95%. These findings underscore the effectiveness of transfer learning in enhancing model performance with limited data. Another significant finding from the review is the increasing preference for Contrast-enhanced mammography (CEM) over full-field digital mammography (FFDM) in classifying mammography images. This shift is driven by the limitations of mammography, particularly in dense breasts where gland shielding and overlapping affect diagnostic performance. CEM, integrating intravenous iodine-contrast enhancement with digital mammography, emerges as a promising technology endorsed by the American College of Radiology for breast cancer diagnosis. Studies indicate that CEM's sensitivity is comparable to that of magnetic resonance imaging (MRI), with a notably higher positive predictive value than MRI. However, challenges persist, including variations in technique influencing radiologists' assessments and interobserver variability in interpretation. Moreover, despite its promising potential, the relatively recent development of CEM means that diagnostic experience is still evolving. In addition, a key finding of the review is the growing emphasis on radiomics in classifying mammography images. Radiomics, which leverages advanced image analysis techniques to extract quantitative data from medical images, holds significant promise in precision medicine, particularly for breast cancer diagnosis and prognosis. As an integral component of cancer management, medical imaging provides rich information on regions of interest without invasive procedures. Radiomics facilitates the utilization of this data to develop predictive models for both diagnosis and prognosis, tailored to individual patients for optimal outcomes. This multidisciplinary approach is poised to revolutionize breast cancer care, enhancing the accuracy and precision of mammography imaging. Recent advancements in DL algorithms, coupled with vast imaging datasets, have yielded promising models for radiomics applications in breast cancer. While patient acceptance of DL in clinical practice is growing, it is viewed as a supportive tool for radiologists rather than a replacement. Radiomics has demonstrated utility in various aspects of breast imaging, including distinguishing between malignant and benign lesions, assessing tumor subtype and grade, predicting therapy response and recurrence risk, and potentially replacing physical breast biopsies in the future. The research highlights the essential need to leverage various open-access mammography datasets to validate segmentation, detection, and classification outcomes, ultimately contributing to effective diagnosis. Also, based on our analysis, it is clear that there is an urgent requirement for a substantial volume of annotated, balanced clinical data and the development of a unified, fully automated framework capable of accurately diagnosing breast cancer with minimal human intervention. Additionally, establishing a standardized repository, accompanied by ground truth annotations for the images, is essential to address these needs effectively. Therefore, establishing substantial public databases emerges as a crucial step for future research endeavours. This study also succinctly explores the challenges and future directions in deploying deep learning (DL) systems to diagnose breast malignancies through mammography images. Future research endeavours may prioritize the following areas:

- Advancing DL models to provide transparent insights into their classification or segmentation decisions, fostering trust among healthcare professionals in AI-assisted diagnosis.
- Tackling the challenge of generalizability by refining models to perform effectively across datasets with diverse ethnicities, breast densities, and imaging protocols, employing techniques like domain adaptation and data augmentation.
- Developing models capable of predicting breast cancer subtypes to tailor personalized treatment strategies, integrating clinical data with mammogram images for a comprehensive approach.
- Designing user-friendly interfaces and integrating AI models seamlessly into clinical workflows to streamline and expedite breast cancer diagnosis processes.
- Creating models specifically tailored to detect cancers that manifest between routine mammogram screenings, enabling earlier detection and potentially improving patient outcomes.
- Implementing mechanisms for continuous learning and model updates to ensure ongoing accuracy and adaptability to evolving medical knowledge as new data becomes available.
- Exploring federated learning approaches to facilitate model training on distributed datasets from different hospitals, maintaining patient privacy while advancing AI capabilities in breast cancer diagnosis.

The strength of the paper lies in its comprehensive and detailed explanation of the application of deep learning techniques in the early-stage study of breast cancer, offering valuable insights that can contribute to advancements in women's health. One primary limitation of this study is its exclusive focus on journals discussing breast cancer detection, diagnosis, and segmentation in mammography using DL techniques. To ensure the relevance of the included research papers, irrelevant publications were identified and excluded based on predefined search criteria. While this approach ensured the suitability of the selected papers for the investigation, it is acknowledged that the inclusion of additional sources, such as supplementary textbooks and conference articles, could have enriched the review.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] F. Soerjomataram, I.Bray, "Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070," *Nat Rev Clin Oncol*, vol. 18, no. 10, pp. 663–672, 2021.

[3] "Breast cancer," *World Health Organization*, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[4] M. Moghbel, C.-Y. Ooi, N. Ismail, Y. Hau, and N. Memari, "A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography," *Artificial Intelligence Review*, vol. 53, 03 2020. [Online]. Available: https://link.springer.com/article/10.1007/s10462-019-09721-8

[5] E. Kozegar, M. Soryani, H. Behnam, M. Salamati, and T. Tan, "Computer aided detection in automated 3-d breast ultrasound images: a survey," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 1919–1941, 2020.

[6] I. Domingues, G. Pereira, P. Martins, H. Duarte, J. a. Santos, and P. H. Abreu, "Using deep learning techniques in medical imaging: a systematic review of applications on ct and pet," *Artif. Intell. Rev.*, vol. 53, no. 6, p. 4093–4160, aug 2020.

[7] G. Murtaza, L. Shuib, A. Wahid, A. Wahab, G. Mujtaba, H. Nweke, M. Al-Garadi, F. Zulfiqar, G. Raza, and N. Azmi, "Deep learning-based breast cancer classification through medical imaging modalities: State of the art and research challenges," *Artif Intell Rev 53*, p. 1655–1720, 05 2020.

[8] H. G. Welch, P. C. Prorok, A. J. O'Malley, and B. S. Kramer, "Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness," *N Engl J Med*, vol. 375, no. 15, pp. 1438–1447, 2016.

[9] B. Percha, H. Nassif, J. Lipson, E. Burnside, and D. Rubin, "Automatic classification of mammography reports by bi-rads breast tissue composition class," *J Am Med Inform Assoc*, vol. 19, no. 5, pp. 913–6, 2021.

[10] K. Loizidou, R. Elia, and C. Pitris, "Computer-aided breast cancer detection and classification in mammography: A comprehensive review," *Computers in Biology and Medicine*, vol. 153, p. 106554, 2023.

[11] J. H. Yoon and E.-K. Kim, "Deep learning-based artificial intelligence for mammography," *Korean J Radiol*, vol. 22, no. 8, pp. 1225–1239, 2021.

[12] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu Rev Biomed Eng*, vol. 19, pp. 221–248, 2017.

[13] Y. Qiu, S. Yan, R. R. Gundreddy, Y. Wang, S. Cheng, H. Liu, and B. Zheng, "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," *J Xray Sci Technol*, vol. 25, no. 5, pp. 751–763, 2017.

[14] V. Bhateja, M. Misra, and S. Urooj, "Breast cancer and mammographic anomalies," pp. 9–20, 2020.

[15] W. H. Yuan, H. C. Hsu, Y. Y. Chen, and C. H. Wu, "Supplemental breast cancer-screening ultrasonography in women with dense breasts: a systematic review and meta-analysis," *Br J Cancer*, vol. 123, no. 4, pp. 673–688, 2020.

[16] Y. Feng, M. Spezia, S. Huang, C. Yuan, Z. Zeng, L. Zhang, X. Ji, W. Liu, B. Huang, W. Luo, B. Liu, Y. Lei, S. Du, A. Vuppalapati, H. H. Luu, R. C. Haydon, T.-C. He, and G. Ren, "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis," *Genes Dis*, vol. 5, no. 2, pp. 77–106, 2018.

[17] R. J. Oskouei, N. M. Kor, and S. A. Maleki, "Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges," *Am J Cancer Res*, vol. 7, no. 3, pp. 610–627, 2017.

[18] "Masses in mammography: What are the underlying anatomopathological lesions?" *Diagnostic and Interventional Imaging*, vol. 95, no. 2, pp. 124–133, 2014, radio-histological correlations in breast imaging.

[19] S. Kim, T. X. M. Tran, H. Song, and B. Park, "Microcalcifications, mammographic breast density, and risk of breast cancer: a cohort study," *Breast Cancer Research*, vol. 24, no. 1, p. 96, 2022. [Online]. Available: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-022-01594-0#citeas

[20] C. E. Comstock, C. Gatsonis, G. M. Newstead, B. S. Snyder, I. F. Gareen, J. T. Bergin, H. Rahbar, J. S. Sung, C. Jacobs, J. A. Harvey, M. H. Nicholson, R. C. Ward, J. Holt, A. Prather, K. D. Miller, M. D. Schnall, and C. K. Kuhl, "Comparison of abbreviated breast mri vs digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening," *Jama*, vol. 323, no. 8, pp. 746–756, 2020.

[21] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "Mammographic image analysis society (mias) database v1.21," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:62937784

[22] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (tcia): Maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, 07 2013.

[23] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, no. 1, p. 170177, 2017.

[24] I. Moreira, I. Amaral, I. Domingues, A. J. O. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database." *Academic radiology*, vol. 19 2, pp. 236–48, 2012.

[25] M. A. Guevara Lopez, N. Posada, D. Moura, R. Pollán, J. Franco-Valiente, C. Ortega, M. Del Solar, G. Díaz-Herrero, I. Ramos, J. Loureiro, T. Fernandes, and B. Araújo, "Bcdr: A breast cancer digital repository," pp. 1065–1066, 01 2012. [Online]. Available: https://www.researchgate.net/publication/258243150_BCDR_A_BREAST_CANCER_DIGITAL_REPOSITORY

[26] S. M. Shah, R. A. Khan, S. Arif, and U. Sajid, "Artificial intelligence for breast cancer analysis: Trends directions," *Comput Biol Med*, vol. 142, p. 105221, 2022.

[27] "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, 2016.

[28] D. Huang, H. Bai, L. Wang, Y. Hou, L. Li, Y. Xia, Z. Yan, W. Chen, L. Chang, and W. Li, "The application and development of deep learning in radiotherapy: A systematic review," *Technol Cancer Res Treat*, vol. 20, p. 15330338211016386, 2021.

[29] G. S. H. I. Shen, D. Wu, "Deep learning in medical image analysis," *Annu Rev Biomed Eng*, vol. 19, pp. 221–248, 2017.

[30] M. Nasser and U. K. Yusof, "Deep learning based methods for breast cancer diagnosis: A systematic review and future direction," *Diagnostics*, vol. 13, no. 1, 2023.

[31] "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021.

[32] J. Gao, Q. Jiang, B. Zhou, and D. Chen, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview," *Mathematical Biosciences and Engineering*, vol. 16, pp. 6536–6561, 07 2019.

[33] E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm," *Neural Comput Appl*, vol. 34, no. 20, pp. 18 015–18 033, 2022. [Online]. Available: https://doi.org/10.1007/s00521-022-07445-5

[34] N. V. Slavine, S. Seiler, T. J. Blackburn, and R. E. Lenkinski, "Image enhancement method for digital mammography"," 2018.

[35] D. Ponraj, M. Jenifer, P. Poongodi, and S. Manoharan, "A survey on the preprocessing techniques of mammogram for the detection of breast cancer," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, 01 2011.

[36] R. Llobet, R. Paredes, and J. C. Pérez-Cortés, "Comparison of feature extraction methods for breast cancer detection," pp. 495–502, 2005.

[37] A. T. Tosin, A. T. Morufat, S. A. Aladejobi, O. M. Oyedepo, O. E. Olusayo, and O. S. Olatude, "Feature extraction techniques for mass detection in digital mammogram (review)," *Journal of Scientific Research and Reports*, vol. 17, no. 1, p. 1–11, Dec. 2017.

[38] Y. Nahid, Abdullah-Al Kong, "Involvement of machine learning for breast cancer image classification: A survey," *Computational and Mathematical Methods in Medicine*, vol. 2017, p. 3781951, 2017.

[39] N. Seliya, A. Abdollah Zadeh, and T. M. Khoshgoftaar, "A literature review on one-class classification and its potential applications in big data," *Journal of Big Data*, vol. 8, pp. 1–31, 2021.

[40] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[41] G. Altan, "A deep learning architecture for identification of breast cancer on mammography by learning various representations of cancerous mass," *Studies in Computational Intelligence*, pp. 169–187, 09 2020.

[42] J. Gallego-Posada, D. Montoya-Zapata, and O. Quintero-Montoya, "Detection and diagnosis of breast tumors using deep convolutional neural networks," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:39348125

[43] M. Yemini, D. Y. Zigel, and D. Lederman, "Detecting masses in mammograms using convolutional neural networks and transfer learning," *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*, pp. 1–4, 2018.

[44] "Recent advancements in machine learning and deep learning-based breast cancer detection using mammograms," *Physica Medica*, vol. 114, p. 103138, 2023.

[45] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying deep learning methods for mammography analysis and breast cancer detection," *Applied Sciences*, vol. 13, no. 7, 2023.

[46] A. K. Shanmugam and S. Kala, "Review on deep learning-based cad systems for breast cancer diagnosis," *Technology in cancer research treatment*, vol. 22, p. 15330338231177977, 06 2023.

[47] L. Wang, "Mammography with deep learning for breast cancer detection," *Frontiers in Oncology*, vol. 14, 01 2024.

[48] J. Arevalo, F. González, R. Pollán, J. Oliveira, and M. A. Guevara Lopez, "Convolutional neural networks for mammography mass lesion classification," 08 2015.

[49] S. Malebary and A. Hashmi, "Automated breast mass classification system using deep learning and ensemble learning in digital mammogram," *IEEE Access*, vol. PP, pp. 1–1, 04 2021.

[50] "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, pp. 103–115, 2019.

[51] H. Li, J. Niu, D. Li, and C. Zhang, "Classification of breast mass in two-view mammograms via deep learning," *IET Image Processing*, vol. 15, 12 2020.

[52] Y. Zhang, S. Satapathy, D. Guttery, J. Gorriz, and S. Wang, "Improved breast cancer classification through combining graph convolutional network and convolutional neural network," *Information Processing and Management*, vol. 58, p. 102439, 01 2021.

[53] A. Saber, M. Sakr, O. Abo-Seida, A. Keshk, and H. Chen, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique," *IEEE Access*, vol. PP, pp. 1–1, 05 2021.

[54] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning - a new frontier in artificial intelligence research [research frontier]." *IEEE Comp. Int. Mag.*, vol. 5, no. 4, pp. 13–18, 01 2010.

[55] S. Boumaraf, X. Liu, C. Ferkous, and X. Ma, "A new computer-aided diagnosis system with modified genetic feature selection for bi-rads classification of breast masses in mammograms," *BioMed Research International*, vol. 2020, pp. 1–17, 05 2020.

[56] J. Peng, C. Bao, C. Hu, X. Wang, W. Jian, and W. Liu, "Automated mammographic mass detection using deformable convolution and multiscale features," *Medical biological engineering computing*, vol. 58, 04 2020.

[57] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 06 2016.

[58] M. A. Al-antari Aisslab and T.-S. Kim, "Evaluation of deep learning

detection and classification towards computer-aided diagnosis of breast lesions in digital x-ray mammograms," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105584, 06 2020.

[59] S. Kebede, F. Gelana, T. Debelee, M. Aleme, W. Bedane, B. Mezgebu, and Z. Merga, "Dual view deep learning for enhanced breast cancer screening using mammography," *Scientific Reports*, vol. 14, 02 2024.

[60] H. T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao, and V. Vu, "Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography," *medRxiv*, 2022. [Online]. Available: https://www.medrxiv.org/content/early/2022/03/10/2022.03.07.22272009

[61] D. C. Dissanayake, F. Afzal, E. Westerberg, and A. Cheddad, "Mini-ddsm: Mammography-based automatic age estimation," 10 2020.

[62] K. R. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 9–, 2016.

[63] E. L. Omonigho, M. David, A. Adejo, and S. Aliyu, "Breast cancer:tumor detection in mammogram images using modified alexnet deep convolution neural network," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, 2020, pp. 1–6.

[64] G. Jayandhi, J. Jasmine, and S. M. Joans, "Mammogram learning system for breast cancer diagnosis using deep learning svm." *Computer Systems Science & Engineering*, vol. 40, no. 2, 2022. [Online]. Available: http://dx.doi.org/10.32604/csse.2022.016376

[65] A. Bokade and A. Shah, "Breast cancer diagnosis in mammography images using deep convolutional neural network-based transfer and scratch learning approach," *Indian Journal Of Science And Technology*, vol. 16, pp. 1385–1394, 05 2023.

[66] A. Elkorany and Z. Elsharkawy, "Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance," *Scientific Reports*, vol. 13, 02 2023.

[67] F. Pesapane, C. Trentin, F. Ferrari, G. Signorelli, P. Tantrige, M. Montesano, C. Cicala, R. Virgoli, S. D'Acquisto, L. Nicosia, D. Origgi, and E. Cassano, "Deep learning performance for detection and classification of microcalcifications on mammography," *European radiology experimental*, vol. 7, p. 69, 11 2023.

[68] Z. Liu, J. Peng, X. Guo, S. Chen, and L. Liu, "Breast cancer classification method based on improved vgg16 using mammography images," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 2, p. 100885, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1687850724000694

[69] T. Mahmood, T. Saba, A. Rehman, and F. S. Alamri, "Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography," *Expert Systems with Applications*, vol. 249, p. 123747, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424006134

[70] M. A. K. Raiaan, N. M. Fahad, M. S. H. Mukta, and S. Shatabda, "Mammo-light: A lightweight convolutional neural network for diagnosing breast cancer from mammography images," *Biomedical Signal Processing and Control*, vol. 94, p. 106279, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424003379

[71] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Scientific Reports*, vol. 6, p. 27327, 06 2016.

[72] E. Wu, K. Wu, D. Cox, and W. Lotter, "Conditional infilling gans for data augmentation in mammogram classification," pp. 98–106, 09 2018.

[73] T. Glasmachers, "Limits of end-to-end learning," in *Proceedings of the Ninth Asian Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M.-L. Zhang and Y.-K. Noh, Eds., vol. 77. Yonsei University, Seoul, Republic of Korea: PMLR, 15–17 Nov 2017, pp. 17–32.

[74] A. Sundaresan, A. Jude, S. Immanuel, and J. Peter, "Classification of mammogram images using multiscale all convolutional neural network (ma-cnn)," *Journal of Medical Systems*, vol. 44, p. 30, 12 2019.

[75] T. Zheng, F. Lin, X. Li, T. Chu, J. Gao, S. Zhang, Z. Li, Y. Gu, S. Wang, F. Zhao, H. Ma, H. Xie, C. Xu, H. Zhang, and N. Mao, "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*, vol. 58, p. 101913, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589537023000901

[76] M. N. Amin, R. Kamal, A. Farouk, M. Gomaa, M. A. Rushdi, and A. M. Mahmoud, "An efficient hybrid computer-aided breast cancer diagnosis system with wavelet packet transform and synthetically-generated contrast-enhanced spectral mammography images," *Biomedical Signal Processing and Control*, vol. 85, p. 104808, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423002410

[77] I. Nissar, S. Alam, S. Masood, and M. Kashif, "Mob-cbam: A dual-channel attention-based deep learning generalizable model for breast cancer molecular subtypes prediction using mammograms," *Computer Methods and Programs in Biomedicine*, vol. 248, p. 108121, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260724001172

[78] H. Cai, J. Wang, T. Dan, J. Li, Z. Fan, W. Yi, C. Cui, X. Jiang, and L. Li, "An online mammography database with biopsy confirmed types," *Scientific Data*, vol. 10, p. 123, 03 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9992520/

[79] A. Panambur, P. Madhu, and A. Maier, "Classification of luminal subtypes in full mammogram images using transfer learning," 01 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2301.09282

[80] K. Atrey, B. Singh, A. Roy, and N. Bodhey, "Real-time automated segmentation of breast lesions using cnn-based deep learning paradigm: Investigation on mammogram and ultrasound," *International Journal of Imaging Systems and Technology*, vol. 32, pp. n/a–n/a, 12 2021.

[81] S. Hicks, I. Strumke, V. Thambawita, M. Hammou, M. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, 04 2022.

[82] L. Taylor and G. Nitschke, "Improving deep learning with generic data augmentation," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2018, pp. 1542–1547.

[83] J. Wang, L. Perez *et al.*, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit*, vol. 11, no. 2017, pp. 1–8, 2017.

[84] R. Samala, H.-P. Chan, L. Hadjiiski, M. Helvie, C. Richter, and K. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Transactions on Medical Imaging*, vol. PP, pp. 1–1, 09 2018.

[85] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "Imagenet: a large-scale hierarchical image database," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 06 2009.

[86] H. He and E. Garcia, "Learning from imbalanced data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, pp. 1263 – 1284, 10 2009.

[87] C. D. Richter, R. K. Samala, H.-P. Chan, L. Hadjiiski, and K. Cha, "Generalization error analysis: deep convolutional neural network in mammography," vol. 10575, p. 1057520, 2018.

[88] M. Galar, A. Fernández, E. Barrenechea, H. Sola, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, pp. 463 – 484, 07 2012.

[89] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, "A novel ensemble method for classifying imbalanced data," *Pattern Recogn.*, vol. 48, no. 5, p. 1623–1637, may 2015.

[90] A. A. Z. . T. M. K. Naeem Seliya, "A literature review on one-class classification and its potential applications in big data," *Journal of Big Data*, vol. 8, no. 1, p. 122, 2021.

Marwa Ben Ammar is a PhD student at the Higher Institute of Medical Technologies of Tunis, Tunisia. She obtained her Research Master's degree in Biophysics, Radiophysics and Medical Imaging from the Higher Institute of Medical Technology of Tunis. Her research interest includes medical devices, programming languages, Business Process Management (BPM) and business process modeling tools, medical image processing and analysis, artificial intelligence (AI) techniques.

Faten LABBENE AYACHI is a lecturer at the Higher School of Communications (SUP'COM) of Tunis, Tunisia. She obtained her PhD in Sciences from the University of Nice Sophia-Antipolis (UNSA) in December 1992. Her research interest includes algorithms and advanced data structures, procedural programming, logic programming and object-oriented programming, relational database design and advanced databases, Oracle and SQL server administration, information systems security, software engineering and UML..

Anselmo Cardoso de Paiva is a Full Professor at the Computer Science Department at the University of Maranhão, Brazil. He received a Doctor degree from PUC-Rio, Brazil in 2001. His research interests include medical image processing, natural language processing, computer vision and information systems.

Riadh Ksantini Riadh Ksantini is Associate Professor at the Dpt of Computer Science, College of IT, University of Bahrain. Skilled in Computer Science, Computer Vision, C++, Artificial Intelligence (AI), and Pattern Recognition. Strong professional with a Doctor of Philosophy (Ph.D.) focused in Machine/Deep learning and computer vision from Sherbrooke University.

Halima Mahjoubi is a professor of biophysics. She obtained a habilitation in radiological physics from the University of Tunis in 2001, and has been a doctor in radiological and medical physics at the Université Paul Sabatier since 1992. She is also the Former Director of the Higher Institute of Medical Technologies of Tunis. She is currently Vice-President of Tunis El Manar University..