International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

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], 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 Maranhão, Sao Luis, MA, BR*
[4] *Department of Computer Science, College of IT, University of Bahrain, Bahrain*

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

Received ## Mon.20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

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. In instance, mammography is the most efficient and highly recommended imaging technique by radiologists in the identification of many types of breast abnormalities However, with the daily growth in mammography, it is still challenging for radiologists and doctors 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. Considering this context, 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 spent, lower the likelihood of mistakes and errors, increase patient satisfaction, and streamline radiologists' workloads. The basic ideas of healthy breast tissue, breast cancer, mammography, and deep learning are briefly presented in this review. This paper delves into the latest advances in systems utilizing deep learning algorithms applied to breast cancer diagnosis using mammograms. Additionally, it provides a concise overview of publicly available mammogram datasets and explores the most widely used metrics for evaluating computer-aided breast cancer diagnosis systems.. Finally, issues and potential research objectives in this developing field are outlined. This paper presents a comprehensive examination of the topic and intend to inspire and direct medical professionals, researchers, scientists, and other healthcare workers who are interested in creating cutting-edge applications toward early breast cancer diagnosis using mammographies 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, indicates a noteworthy change in global cancer trends in 2020. According to these statistics, female mammary carcinoma becoming the preeminent chiefly 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

E-mail:author's email

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 related also to the 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, late menopause, late primiparity, use of contraceptives) increasing the susceptibility to developing breast cancer [3].

Consequently, breast cancer is a serious illness, although it typically has a fair prognosis when diagnosed early. In order to enhance the prognosis, elevate the patient's chance of survival with 50% [4], 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 is currently relies on various widely recognized imaging techniques, including mammography with X-rays [5], Ultrasonography [6], computerized tomography, also known as CT, [7], and MRI, which stands for magnetic resonance imaging [8]. Mammography continues to be the modality that radiologists utilize the most frequently to appropriately diagnose this illness [9]. Radiologists should, in practice, identify aberrant lesions on mammograms during the diagnosing process in order 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 [10] 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) along with 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 is one that has undesirable outcomes. False negatives are more significant when they put the patient's life in peril and happen when the radiologist misses an abnormality. Additionally, studies have showed that lesions with a greater than 2% potential of being malignant will be advised for biopsy in order to decrease the likelihood of FN diagnosis. Only 15–30% of those who get a biopsy are ultimately found to have cancer [11, 12].

Sophisticated algorithms, categorized as CADe for computer-aided detection and CADx for computer-aided diagnosis systems, are developed to assist medical specialists with the interpretation of 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 are continuously developing and evaluating CADe/CADx systemsbased 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 prevent unneeded biopsies and preserve a lot of time for human professionals who would otherwise have to manually review medical images. The concept of CAD systems was originally developed in the 1960s to screen for breast cancer using mammograms [13,14]. As of right now, it is among the most important study areas for clinical image processing [15]. CADe uses compute findings to pinpoint the exact location of concerning lesions while leaving the radiologist to make sense of these anomalies. In contrast, CADx produces quality information that assists the radiologist to make decisions regarding the observed anomalies, in particular to further identify and classify lesions [16].

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 [17]. 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 [18]. Machine learning often requires the manual extraction of characteristics, whereas deep learning is totally automatic.

This study aims to depict the current trends in utilizing systems based on deep learning (DL), aiming for pre-symptomatic identification as well precise breast carcinomas classification, by analyzing mammography images.

We intend to address the subsequent research questions (RQs):

- 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 algorithm 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?

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 existing literature. Section 6 then lists the limitations of CAD systems that employ deep learning methods along with recommendations for future research. Lastly, Section 7 brings the paper to its conclusion.

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 of which produces a number of tiny milk-secreting bulbs (alveoli). The lobes and lobules that gather the milk are connected by milk ducts. 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. All non-fatty tissue is referred to by doctors as fibro-glandular tissue. Ligaments are also 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 also found in each breast [19]. 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 [20]. 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 [20].

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. In order to begin therapy, the specialist must be able to identify between two forms of breast cancer during the diagnosis. Both benign as well 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 [21]. Breast cancers that are non-invasive (also known as in situ), invasive (also known as infiltrating), and metastatic are all examples of malignant tumors, which are cancerous growths [21,22]. 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 [21]. 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, the 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 show up in mammograms as irregularly shaped masses with spikes projecting from them. The non-cancerous masses usually have borders that are well defined and circular or oval in form. [23]. Both macrocalcifications and microcalcifications of the breast can occur [24]. Macrocalcifications, which look as sizable white dots randomly dispersed across the breast on a mammography, are thought to be benign cells. In contrast, in mammography, microcalcifications manifest as minute deposits of calcium, 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 within 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 as well digital mammography (DMM), which areused for different tasks in breast cancer analysis, such 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) [25]. Present practices need for 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 trying to spot possible abnormalities in a mammogram.

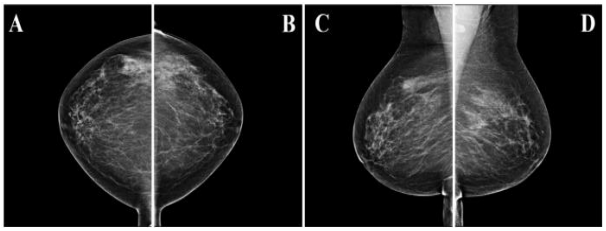

Figure 1. An illustration of the various points of view from a breast mammogram: (A) Right CC view, (B) Left CC view, (C) Right MLO view, and (D) Left MLO view are the four possible views.

D. Deep Learning

Deep learning (DL) operates by automatically deriving feature representations from input data [26,27]. 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 labor-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 use of these algorithms are enabled by the availability of extensive datasets, as they demand substantial volumes of training data. Within the domain of image data, deep learning (DL) models acquire hierarchical attributes. 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 [28]. The adoption of DL techniques has found application across a spectrum of medical specializations, most notably in radiology and pathology [29]. Deep learning, as an emerging technique, is surpassing traditional machine learning methods and is increasingly integrated into Computer-Aided Diagnosis (CAD) systems [30]. Deep learning techniques have recently showcased their potential in diagnosing breast cancer approximately one year earlier than traditional clinical methods [31]. Convolutional Neural Networks (CNNs) are widely utilized as one of the predominant architectures in deep learning. With enough training data, CNNs possess the capacity to grasp intricate and well-structured hierarchical attributes within an image. They are widely favored for neural network-based image classification and have demonstrated impressive performance for medical image analysis and categorization [32]. As depicted in Fig. 2, a basic CNN architecture involves the integration of one or more layers for convolution and pooling, subsequently complemented by one or more layers that are fully connected [33].

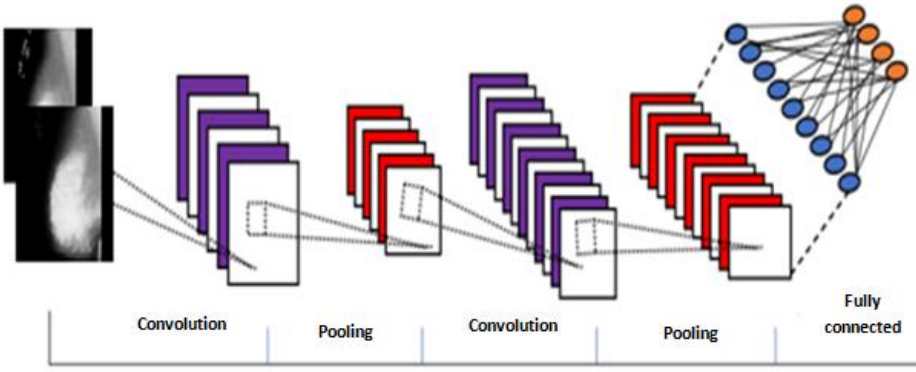

Figure 2. An illustration of a basic CNN architecture used for image diagnosis and classification.

3. PUBLIC AVAILABLE MAMMOGRAPHY DATASETS

Within this section, we make an effort 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, and 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. Table 1 displays a quick description of these collections.

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	Web Link
MIAS [101]	322	FM	CaD, Mass-C, Mass-S, Mass-I, Arch, Asym, Norm, Ben, Mal, BenWC	.PGM	MLO	1024 × 1024 pixels	https://www.kaggle.com/datasets/kmader/mias-mammography
CBIS-DDSM [102]	10239	FM	NORM, B, M .	.DICOM	MLO/ CC	16 bit	https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset
DDSM [103]	10480	FM	B, C, NORM, BWC	.JPEG	MLO/ CC	8-16 bit	https://www.kaggle.com/datasets/skooch/ddsm-mammography
INbreast [104]	410	DM	B, M, NORM	.DICOM	MLO/ CC	14 bit	https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset
BCDR [105]	7315 (3703 FFDM - 3612 FM)	DM	NORM, B, M	.TIFF	MLO/ CC	8-14 bit	https://www.medicmind.tech/cancer-imaging-data

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.

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

E-mail:author's email

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.

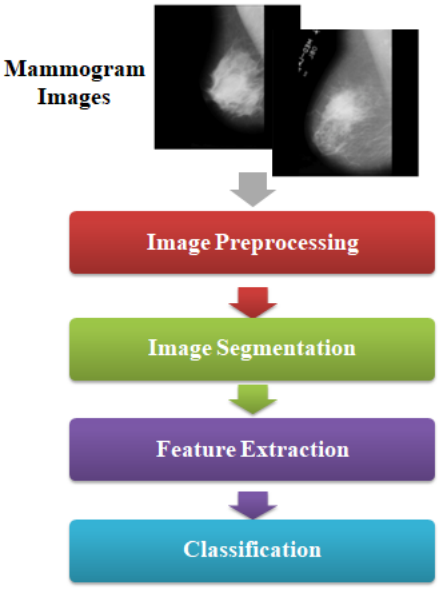

Figure 3. Illustration of the general layout of a CAD system for diagnosing breast cancer.

Table 2 below presents the key steps of a CAD system.

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

Stages	Description
Image pre-processing	Most automated image analysis systems depend on the pre-processing stage [34]. Basically, this is done to improve image quality, reduce noise, also removing unnecessary, unwanted artifacts [35]. At this stage, image contrast enhancement methods are used based on the equalization of the histogram and noise reduction techniques (mean and median filters [36], among others). In addition, other operations can be carried out during the pre-processing phase, covering tasks such as image resizing, data augmentation [37] 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 [38]. There are several works approaching the breast segmentation by using different methods, based on thresholding [39], active contour [40], edge-based and region based [41,42], gradient weight map [43], conditional network [44], support-pixel correlation and statistical method [45].
Feature extraction	Basically, the process of extracting feature sets from mammograms 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 [46]. Handcrafted features encompass a variety of options for extracting information from breast mammograms, including texture [47,48], morphological aspects, and descriptors [49], as well as shape, intensity, and hybrid features [50]. Additional possibilities include curvelet-based statistical features [51,52], local and global features [53]. Other methods involve the use of histogram of gradients (HOG), SIFT, and wavelets [54]. Moreover, features encompass contrast, geometrical aspects, location data, context, and patient-related information [46]. 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 are proposed for feature extraction [46,55].

Classification	Following feature extraction, the final phase involves the breast lesions classification, 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 are applicable to classifying breast cancer tissue, primarily including binary classes classification (distinguishing between cancerous and non cancerous) [56], multiple classes classification (encompassing categories like healthy tissue, non cancerous lesions, in situ malignancy, and invasive malignancy) [857], and the one-class classification (OCC) approach [58]. Statistical ML-based Classifiers and DL-based Classifiers are the two basic classification model types t used to diagnose breast cancer. Pathologists and doctors can utilize artificial intelligence based algorithms to diagnose breast cancer to aid in their decision-making. The statistical machine learning techniques are commonly used for the classification of breast cancer images [33]. Convolutional networks is one of the most effective used models for image analysis [59]. There are several DL architecture based on pretrained models such as AlexNet [33], VGG-16, ResNetXt50 [60,61], and Google Inception-V3 architecture [62].

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 [63-64, 65, 66, 67, 68, and69]. This section, outlines the main deep learning approaches, encompassing CNNs and RNNs (recurrent networks), and also we present transfer learning methods.

The CNN is composed of several layers where are applied convolutions and max-pooling operations [70]. In a recent study [71], a system called BMC was proposed by researchers 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 accuracy of 96% using DDSM and 95% using MIAS.

Other study [72], proposes an algorithm known as CNNI-BCC yielded impressive results, including a sensitivity of 89.47%, an accuracy rate of 90.50%, AUC of 0.901 ± 0.0314, and 90.71% of specificity. The utilization of this algorithm has the potential to be beneficial in the mammogram images classification into non-cancerous, cancerous, and normal classes, even without previous knowledge about the presence of a cancerous lesion.

Moreover, in a different study [73], researchers introduced a Two-perspective mammogram classification model that combines a CNN with a RNN, for the breast masses classification in mammographic images. Their approach reaches classification accuracy of 94.7%, recall of 94.1%, and AUC value of 0.968.

In separate research inquiries [74, 75], models based on CNNs were employed for theabnormalities classification in mammograms. It was used the MIAS dataset. CNN-based models has demonstrated encouraging outcomes, improoving the accuracy and of CAD systems for breast cancer diagnosis.

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 stack of restricted Boltzmann machine (RBM)[76]. Itis an effective tool for breast cancer diagnosis for several reasons. They can be used to reduce the input feature vectors dimensionality [77]. In [78], 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 capable of accurately identifying lesions in mammography images holds significant value for healthcare professionals. Consequently, researchers in [79] 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. [80, 81].

The YOLO effectiveness was assessed in [82] and [83] for detecting lesions in the breast. Subsequently, they made modifications to and evaluated:a traditional Multi-Layer Perceptron, 50-Layer Convolutional Model, as well Inception 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%.

Transfer learning has become a widely adopted technique, it serves to address 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 [84]. Recent studies have increasingly embraced this approach [85,86]. As an example, authors in [87] introduced a modified AlexNet architecture for mammogram classification of masses as benign/malignant. MIAS testing yielded 95.70% accuracy for the final model.

Researchers in a study [88] 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 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 show a discriminative accuracy in distinguishing calcifications, using the SVM classifier.

In an independent study [89], 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. In order to produce features in multiscale and guarantee stability at higher resolutions, this generator used a refinement process. Importantly, the GAN was confined to infill exclusively lesions, including both masses and calcifications. To assess the generated images' quality, a ResNet-50 classifier was employed. The study compared the classification performance of data enhancement using GANs and traditional methods, revealing that synthetic augmentation enhances classification accuracy.

Lately, a number of investigations have adopted the End-To-End (E2E) training approach, which has demonstrated promising outcomes for breast cancer detection [90]. In this context, researchers in [91] 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 features extraction. The experiments achieved 96.47% of accuracy and a 0.99 AUC score using the mini-MIAS dataset.

Table 3 summarizes studies focusing on breast cancer diagnosis from mammogram images using DL techniques, along with their respective performance metrics.

TABLE III. COMPREHENSIVE ANALYSIS OF STATE-OF-THE-ART RESEARCH IN BREAST CANCER DIAGNOSIS ON MAMMOGRAPHY UTILIZING DEEP LEARNING. ACCURACY (ACC); AREA UNDER THE CURVE (AUC); SEGMENTATION (S); CLASSIFICATION (C).ION

Paper	Year	Application	Model	Dataset	Evaluation Metric
[71]	2021	Classification	k- mean clustering + Long Short-Term Memory + RNN+ CNN + Random Forest	DDSM MIAS	DDSM: Acc=96% MIAS: Acc=95%
[72]	2019	Classification	CNNI-BCC	MIAS	Sensitivity = 89.47% Acc = 90.5% AUC = 0.90 ± 0.03 Specificity = 90.7%
[73]	2021	Classification	CNN-RNN	DDSM	Acc = 94.7% Recall = 94.1% AUC = 0.968
[74]	2021	Classification	GNN + CNN	MIAS	Acc = 96.1%
[75]	2021	Classification	Convolutional Neural Network with Knowledge transfer	MIAS	Acc = 98.87% F-score = 99.3%
[78]	2020	Classification	DBN	DDSM	Acc = 84.5%
[79]	2020	Mass Segmentation	Faster Region-based Convolutional Neural Network	CBIS-DDSM INbreast	CBIS: TP = 0.93 INbreast: TP = 0.95
[92]	2022	Breast lesions Segmentation	CNN	DDSM	dice-coefficient = 65%

[86]	2020	Breast lesions Segmentation/Classification	YOLO Feedforward CNN ResNet-50 InceptionResNet-V2	DDSM INbreast	S: DDSM: F1-score = 99.28% INbreast: F1-score= 98.02% C: DDSM: Acc = 97.5% INbreast: Acc = 95.32%
[87]	2020	Classification	AlexNet + Augmentation	MIAS	Acc=95.70%
[88]	2016	Detection and classification of lumps and deposits of calcification	Stacked autoencoder	Private	Acc =87%
[89]	2018	Detection and classification of non cancerous and cancerous deposits of calcification and lumps	GAN + ResNet50	DDSM	AUC = 0.896
[91]	2020	Classification	End-to-End CNN	mini-MIAS	Acc = 96.47% AUC = 0.99

5. Evaluation Metrics

Within this section, we present the evaluation measures employed for assessing the performance of methods for diagnosis and detection of Cancer of the breast. A summarized overview of the calculation formulas and explanations for the most frequently employed evaluation metrics in the literature can be found in Table 4.

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

TABLE IV. Evaluation metrics commonly used for Breast Cancer Diagnosis

Metrics	Description	Formula		References
Accuracy *(Acc)*	The accuracy is computedby taking the proportion of correct predictions and dividing it the overall predictions generated. Essentially, it provides insight into the proportion of the model's predictions that were accurate.	$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)}$	(1)	[93]
Precision *(Pr)*	Precision 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 labeled as positive by the classifier.	$Pr = \frac{TP}{(TP + FP)}$	(2)	[93]
Sensitivity (Sn)/ Recall (R)/TPR	Sensitivity (Sn) or Recall (R), 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)	[93]
Specificity (Sp)	Specificity (True-Negative Rate), 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)	[93]
F1-score	F1-score metric 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.	$F1 - score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)}$	(5)	[93]

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)	[93]

6. Challenges And Future Directions

This section discusses some of the challenges and research directions in DL-based systems diagnosing of 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. A large number of 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 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 as 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 entails generating 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 [94] and [95].

Another effective strategy involves utilizing transfer learning, which has demonstrated significant success in the analysis of mammography images, as exemplified in [96]. 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 is employed general imaging dataset for this purpose [97], 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 [98], 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 referred to as imbalanced learning. For instance, in a simulation study conducted to address this issue [99], researchers examined how well a CNN could classify breast masses into malignant or benign categories. They used a 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 [100], which could affect model generalization. Conversely, another study [101] 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 of abnormal samples in comparison to known class instances. This approach offers a promising solution to address challenges associated with severely imbalanced datasets [102], 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 determining 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 endeavors should consider the utilization of Deep Learning-based One-Class Classification models to increase the accuracy of 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 comprehensively gauge the effectiveness of each model.

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.

7. CONCLUSION

8. This study examined a number of studies that used mammography pictures and DL models to diagnose breast cancer. Based on the findings of this study, the majority of research projects compare the accuracy rates of different deep learning algorithms..

ACKNOWLEDGMENT

The authors have no acknowledgments.

REFERENCES

[1] Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/CAAC.21660.

[2] Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. Nat Rev Clin Oncol. 2021.

[3] Shaoyuan Lei, Rongshou Zheng, Siwei Zhang, Shaoming Wang, Ru Chen, Kexin Sun, Hongmei Zeng, Jiachen Zhou, Wenqiang Wei. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020.CA Cancer J Clin. 2021 May; 71(3):209-249. doi: 10.3322/caac.21660.

[4] Breast Cancer. 2021. Available online: https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed on 19 July 2021).

[5] Moghbel, M.; Ooi, C.Y.; Ismail, N.; Hau, Y.W.; Memari, N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. Artif. Intell. Rev. 2019, 53, 1873–1918.

[6] Kozegar, E.; Soryani, M.; Behnam, H.; Salamati, M.; Tan, T. Computer aided detection in automated 3-D breast ultrasound images: A survey. Artif. Intell. Rev. 2019, 53, 1919–1941.

[7] Domingues, I.; Pereira, G.; Martins, P.; Duarte, H.; Santos, J.; Abreu, P.H. Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET. Artif. Intell. Rev. 2019, 53, 4093–4160.

[8] Murtaza, G.; Shuib, L.; Wahab, A.W.A.; Mujtaba, G.; Nweke, H.F.; Al-Garadi, M.A.; Zulfiqar, F.; Raza, G.; Azmi, N.A. Deep learning-based breast cancer classification through medical imaging modalities: State of the art and research challenges. Artif.Intell. Rev. 2019, 53, 1655–1720.

[9] Welch HG, Prorok PC, O'Malley AJ, Kramer BS (2016) Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med 375(15):1438–1447.

[10] Percha B, Nassif H, Lipson J, Burnside E, Rubin D. Automatic classification of mammography reports by BI-RADS breast tissue composition class. J Am Med Inform Assoc. 2012 Sep-Oct;19(5):913-6. doi: 10.1136/amiajnl-2011-000607. Epub 2012 Jan 29. PMID: 22291166; PMCID: PMC3422822.

[11] Sickles EA. Periodic mammographic follow-up of probably benign lesions: results in 3184 consecutive cases. Radiology. 1991;179:463–468.

[12] Adler DD, Helvie MA. Mammographic biopsy recommendations. Curr Op Radiol. 1992;4:123–129.

[13] H. Sittek, K. Herrmann, C. Perlet, et al., Computer-aided diagnosis in mammography, Der Radiologe, 37 (1997), 610–616.

[14] R. Takahashi and Y. Kajikawa, Computer-aided diagnosis: A survey with bibliometric analysis, Int. J. Med. Inf., 101 (2017), 58–67.

[15] A. Mansoor, U. Bagci, B. Foster, et al., Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends, Radiographics, 35 (2015), 1056–1076.

[16] Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS (2003) Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. AJR Am J Roentgenol 181(4):1083–1088.

[17] D. Shen, G. Wu and H. I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng., 19 (2017), 221–248.

[18] Qiu, Y.; Yan, S.; Gundreddy, R.R.; Wang, Y.; Cheng, S.; Liu, H.; Zheng, B. A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. J. X-ray Sci. Technol. 2017, 25, 751–763.

[19] Bhateja, V., Misra, M., Urooj, S., 2020. Breast cancer and mammographic anomalies. In: Non-Linear Filters for Mammogram Enhancement: A Robust Computer-Aided Analysis Framework for Early Detection of Breast Cancer, Volume 861. Springer Singapore, Singapore, pp. 9–20. https://doi.org/10.1007/978-981-15-0442-6_2.

[20] Yuan, W.-H., Hsu, H.-C., Chen, Y.-Y., Wu, C.-H., 2020. Supplemental breast cancerscreening ultrasonography in women with dense breasts: a systematic review and meta-analysis. Br. J. Cancer 123, 673–688. https://doi.org/10.1038/s41416-020-0928-1.

[21] Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, Ji X, Liu W, Huang B, Luo W, Liu B, Lei Y, Du S, Vuppalapati A, Luu HH, Haydon RC, He TC, Ren G. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 2018 May 12;5(2):77-106. doi: 10.1016/j.gendis.2018.05.001. PMID: 30258937; PMCID: PMC6147049.

[22] Oskouei RJ, Kor NM, Maleki SA. Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges. Am J Cancer Res. 2017 Mar 1;7(3):610-627. PMID: 28401016; PMCID: PMC5385648.

[23] Berment H, Becette V, Mohallem M, Ferreira F, Chérel P (2014) Masses in mammography: what are the underlying anatomopathological lesions? Diagnostic and Interventional Imaging 95:124–133. https://doi.org/10.1016/J.DIII.2013.12.010.

[24] Nalawade YV (2009) Evaluation of breast calcifications. The Indian Journal of Radiology & Imaging 19(4):282–286. https://doi.org/10.4103/0971-3026.57208.

[25] Comstock, C.E.; Gatsonis, C.; Newstead, G.M.; Snyder, B.S.; Gareen, I.F.; Bergin, J.T.; Rahbar, H.; Sung, J.S.; Jacobs, C.; Harvey, J.A.; et al. Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women with Dense Breasts Undergoing Screening. JAMA 2020, 323, 746–756.

[26] Shah, S.M.; Khan, R.A.; Arif, S.; Sajid, U. Artificial intelligence for breast cancer analysis: Trends & directions. Comput. Biol. Med. 2022, 142, 15.

[27] Roslidar, R.; Rahman, A.; Muharar, R.; Syahputra, M.R.; Arnia, F.; Syukri, M.; Pradhan, B.; Munadi, K. A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection. IEEE Access 2020, 8, 116176–116194.

[28] Zhicheng Jiao, Xinbo Gao, Ying Wang, Jie Li, A deep feature based framework for breast masses classification, Neurocomputing, Volume 197, 2016, Pages 221-231, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2016.02.060.

[29] Huang, D.; Bai, H.; Wang, L.; Hou, Y.; Li, L.; Xia, Y.; Yan, Z.; Chen, W.; Chang, L.; Li, W. The Application and Development of Deep Learning in Radiotherapy: A Systematic Review. Technol. Cancer Res. Treat. 2021, 20, 15330338211016386.

[30] D. Shen, G. Wu and H. I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng., 19 (2017), 221–248.

[31] Maged Nasser, Umi Kalsom Yusof. Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction. Diagnostics 2023, 13(1), 161; https://doi.org/10.3390/diagnostics13010161.

[32] Hang Yu, Laurence T. Yang, Qingchen Zhang, David Armstrong, M. Jamal Deen, Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives, Neurocomputing, Volume 444, 2021, Pages 92-110, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.04.157.

[33] Gao J, Jiang Q, Zhou B, Chen D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. Math Biosci Eng. 2019 Jul 15;16(6):6536-6561. doi: 10.3934/mbe.2019326. PMID: 31698575.

[34] Weickert, J., Weickert, J., Schnörr, C., 2000. PDE-Based Preprocessing of Medical Images Introduction.

[35] Slavine NV, Seiler S, Blackburn TJ, Lenkinski RE (2018). Image enhancement method for digital mammography. Medical imaging 2018: image processing. https://doi.org/10.1117/12.2293604.

[36] Htay, T.T., Maung, S.S., 2018. Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image. Isc. 2018 - 18th Int. Symp. Commun. Inf.Technol. 345–348. https://doi.org/10.1109/ISCIT.2018.8587920.

[37] Taylor, L., Nitschke, G., 2017. Improving Deep Learning using Generic Data Augmentation. The Global Cancer Observatory, G., 2018. Breast Cancer. Source: Globocan 2018. World Heal. Organ. 876, 2018–2019.

[38] Ponraj, D., Jenifer, M., 2011. A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer. J. Emerg. 2, 656–664.

[39] Al-Bayati M, El-Zaart A (2013) Mammogram images thresholding for breast cancer detection using different thresholding methods. Adv Breast Cancer Res 02:72–77. https://doi.org/10.4236/ABCR.2013.23013.

[40] Rampun, A., Morrow, P. J., Scotney, B. W., & Winder, J. (2017). Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial intelligence in medicine, 79, 28–41.

[41] Maitra, I. K., & Bandyopadhyay, S. K. (2017). In Hybrid Intelligence for Image Analysis and Understanding (pp. 339–368). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119242963.ch15.

[42] Varughese LS, Anitha J (2013) A study of region based segmentation methods for mammograms. Int J Res Eng Technol 02:421–425. https://doi.org/10.15623/IJRET.2013.0212070.

[43] Shi, P., Zhong, J., Rampun, A., & Wang, H. (2018). A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms. Computers in biology and medicine, 96, 178–188.

[44] Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., Torrents-Barrena, J. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. expert systems with Applications, 139, 112855. https://doi.org/10.1016/j.eswa.2019.112855.

[45] Yang, S.-C. (2017). A robust approach for subject segmentation of medical Images: Illustration with mammograms and breast magnetic resonance images. Computers & Electrical Engineering, 62, 151–165.

[46] Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A., Karssemeijer, N., 2017. Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal. 35, 303–312. https://doi.org/10.1016/j.media.2016.07.007.

[47] Muramatsu, C., Hara, T., Endo, T., & Fujita, H. (2016). Breast mass classification on mammograms using radial local ternary patterns. Computers in biology and medicine,72, 43–53.

[48] Kulkarni DA, Bhagyashree SM, Udupi GR (2010) Texture analysis of mammographic images. Int J Comput Appl 5:12–17. https://doi.org/10.5120/919-1297.

[49] Llobet R, Paredes R, Pérez-Cortés JC (2005) Comparison of feature extraction methods for breast Cancer detection. Lect Notes Comput Sci 3523:495–502. https://doi.org/10.1007/11492542_61.

[50] Ede, F.P., n.d. A SURVEY FEATURE EXTRACTION TECNIQUES FOR MASS DETECTION IN DIGITAL MAMMOGRAM.

[51] Eltoukhy MM, Faye I, Samir BB (2010). Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram. 2010 international conference on intelligent and advanced systems.https://doi.org/10.1109/icias.2010.5716125.

[52] Eltoukhy MM, Faye I, Samir BB (2012) A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. Comput Biol Med 42:123–128. https://doi.org/10.1016/J.COMPBIOMED.2011.10.016.

[53] Nahid, A., Kong, Y., 2018. Involvement of Machine Learning for Breast Cancer Image Classification : A Survey 2017.

[54] Ergin, S., Kilic, O., 2014. A new feature extraction framework based on wavelets for breast cancer diagnosis. Comput. Biol. Med. 51, 171–182. https://doi.org/10.1016/j.compbiomed.2014.05.008.

[55] R. K. Samala, H. Chan, L. Hadjiiski, et al., Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets, IEEE Trans. Med. Imaging, 38 (2019), 686–696.

[56] Wessam M. Salama, Moustafa H. Aly, Deep learning in mammography images segmentation and classification:

Automated CNN approach, Alexandria Engineering Journal, Volume 60, Issue 5, 2021, Pages 4701-4709, ISSN 1110-0168. https://doi.org/10.1016/j.aej.2021.03.048.

[57] Heenaye-Mamode Khan M, Boodoo-Jahangeer N, Dullull W, Nathire S, Gao X, Sinha GR, et al. (2021) Multi- class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). PLoS ONE 16(8): e0256500. https://doi.org/10.1371/journal.pone.0256500.

[58] Seliya, N., Abdollah Zadeh, A. &Khoshgoftaar, T.M. A literature review on one-class classification and its potential applications in big data. J Big Data 8, 12 (2021). https://doi.org/10.1186/s40537-021-00514-x.

[59] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

[60] J. G. Posada, D. M. Zapata and O. L. Q. Montoya, Detection and Diagnosis of Breast Tumors using Deep Convolutional Neural Networks, Conference Proceedings of the XVII Latin American Conference on Automatic Control (2016), 11-17. Available from: https://pdfs.semanticscholar.org/9566/d1f27a0e5f926827d3eaf8546dab51e40e21.pdf.

[61] N. S. Ismail and C. Sovuthy, "Breast Cancer Detection Based on Deep Learning Technique," 2019 International UNIMAS STEM 12th Engineering Conference (EnCon), Kuching, Malaysia, 2019, pp. 89-92, doi: 10.1109/EnCon.2019.8861256.

[62] Yemini, M., Zigel, Y., Lederman, D., 2019. Detecting Masses in Mammograms using Convolutional Neural Networks and Transfer Learning. 2018 IEEE Int. Conf. Sci. Electr. Eng. Isr. ICSEE 2018 1–4. https://doi.org/10.1109/ICSEE.2018.8646252.

[63] Zahra Rezaei, A review on image-based approaches for breast cancer detection, segmentation, and classification, Expert Systems with Applications, Volume 182, 2021, 115204, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.115204.

[64] Mridha, M.F.; Hamid, M.A.; Monowar, M.M.; Keya, A.J.; Ohi, A.Q.; Islam, M.R.; Kim, J.-M. A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis. Cancers 2021, 13, 6116. https://doi.org/10.3390/cancers13236116.

[65] Hassan, N.M., Hamad, S. & Mahar, K. Mammogram breast cancer CAD systems for mass detection and classification: a review. Multimed Tools Appl 81, 20043–20075 (2022). https://doi.org/10.1007/s11042-022-12332-1.

[66] Nusrat Mohiud din, Rayees Ahmad Dar, Muzafar Rasool, Assif Assad, Breast cancer detection using deep learning: Datasets, methods, and challenges ahead, Computers in Biology and Medicine, Volume 149, 2022, 106073, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2022.106073.

[67] Madani, M.; Behzadi, M.M.; Nabavi, S. The Role of Deep Learning in Advancing Breast Cancer Detection Using Different Imaging Modalities: A Systematic Review. Cancers 2022, 14, 5334. https://doi.org/10.3390/cancers14215334.

[68] KosmiaLoizidou, Rafaella Elia, Costas Pitris, Computer-aided breast cancer detection and classification in mammography: A comprehensive review, Computers in Biology and Medicine, Volume 153, 2023, 106554, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2023.106554.

[69] Nasser, M.; Yusof, U.K. Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction. Diagnostics 2023, 13, 161. https://doi.org/10.3390/diagnostics13010161.

[70] Nemade, V.; Pathak, S.; Dubey, A.K. A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques. Arch. Comput. Methods Eng. 2022, 29, 4401–4430.

[71] Malebary, S.J.; Hashmi, A. Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram. IEEE Access 2021, 9, 55312–55328.

[72] Ting, F.F.; Tan, Y.J.; Sim, K.S. Convolutional neural network improvement for breast cancer classification. Expert Syst. Appl. 2019, 120, 103–115.

[73] Li, H.; Niu, J.; Li, D.; Zhang, C. Classification of breast mass in two-view mammograms via deep learning. IET Image Process. 2020, 15, 454–467.

[74] Zhang, Y.-D.; Satapathy, S.C.; Guttery, D.S.; Górriz, J.M.; Wang, S.-H. Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network. Inf. Process. Manag. 2020, 58, 102439.

[75] Saber, A.; Sakr, M.; Abo-Seida, O.M.; Keshk, A.; Chen, H. A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. IEEE Access 2021, 9, 71194–71209.

[76] Arel, I.; Rose, D.C.; Karnowski, T.P. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE Comput. Intell. Mag. 2010, 5, 13–18.

[77] Abdel-Zaher, A.M.; Eldeib, A.M. Breast cancer classification using deep belief networks. Expert Syst. Appl. 2016, 46, 139–144.

[78] Boumaraf, S.; Liu, X.; Ferkous, C.; Ma, X. A New Computer-Aided Diagnosis System with Modified Genetic Feature Selection for BI-RADS Classification of Breast Masses in Mammograms. BioMed Res. Int. 2020, 2020, e7695207.

[79] Peng, J.; Bao, C.; Hu, C.; Wang, X.; Jian, W.; Liu, W. Automated mammographic mass detection using deformable convolution and multiscale features. Med. Biol. Eng. Comput. 2020, 58, 1405–1417.

[80] Dildar, M.; Akram, S.; Irfan, M.; Khan, H.U.; Ramzan, M.; Mahmood, A.R.; Alsaiari, S.A.; Saeed, A.H.M.; Alraddadi, M.O.; Mahnashi, M.H. Skin Cancer Detection: A Review Using Deep Learning Techniques. Int. J. Environ. Res. Public Health 2021, 18, 5479.

[81] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

[82] Mohammed A. Al-masni, Mugahed A. Al-antari, Jeong-Min Park, Geon Gi, Tae-Yeon Kim, Patricio Rivera, Edwin Valarezo, Mun-Taek Choi, Seung-Moo Han, Tae-Seong Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, Computer Methods and Programs in Biomedicine, Volume 157, 2018, Pages 85-94, ISSN 0169-2607. https://doi.org/10.1016/j.cmpb.2018.01.017.

[83] Al-Antari, M.A.; Han, S.-M.; Kim, T.-S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. Comput. Methods Programs Biomed. 2020, 196, 105584.

[84] Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. Journal of Big Data 3(1):1–40. https://doi.org/10.1186/S40537-016-0043-6.

[85] Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S (2018) Deep learning to distinguish recalled but benign mammography images in breast cancer screening. Clin Cancer Res 24(23):5902–5909. https://doi.org/10.1158/1078-0432.CCR-18-1115.

[86] Huynh BQ, Li H, Giger ML (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging (Bellingham, Wash) 3:034501. https://doi.org/10.1117/1.JMI.3.3.034501.

[87] E. L. Omonigho, M. David, A. Adejo and S. Aliyu, "Breast Cancer:Tumor Detection in Mammogram Images Using Modified AlexNet Deep Convolution Neural Network," 2020, International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), Ayobo, Nigeria, 2020, pp. 1-6, doi:10.1109/ICMCECS47690.2020.240870.

[88] Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. Sci Rep. 2016 Dec 7;6:27327. doi: 10.1038/srep27327. doi: 10.1038/srep27327.

[89] Wu E, Wu K, Cox D, Lotter W. Conditional Infilling GANs for Data Augmentation in Mammogram Classification. Proceedings of the Image Analysis for Moving Organ, Breast, and Thoracic Images; RAMBO'18; September 16-20, 2018; Granada, Spain. 2018. pp. 98–106.

[90] Glasmachers T (2017) Limits of end-to-end learning. In Asian conference on machine learning, pp 17-32. PMLR.

[91] Agnes SA, Anitha J, Pandian SI, Peter JD (2020) Classification of mammogram images using multiscale all convolutional neural network (MA-CNN). J Med Syst 44(1):1–9. https://doi.org/10.1007/S10916-019-1494-Z.

[92] Atrey, K.; Singh, B.K.; Roy, A.; Bodhey, N.K. Real-time automated segmentation of breast lesions using CNN-based deep learning paradigm: Investigation on mammogram and ultrasound. Int. J. Imaging Syst. Technol. 2021, 32, 1084–1100.

[93] Hicks, S.A., Strümke, I., Thambawita, V. et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep 12, 5979 (2022). https://doi.org/10.1038/s41598-022-09954-8

[94] Taylor L, Nitschke G. Improving Deep Learning using Generic Data Augmentation. 2017. arXiv:1708.06020.

[95] Wang J, Perez L. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 2017arXiv:1712.04621.

[96] Samala RK, Chan H-P, Hadjiiski LM, Helvie MA, Richter CD, Cha K. Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning using Deep Neural Nets. IEEE Trans Medical Imaging. 2019;38(3):686–696.

[97] Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei, 2010. ImageNet: A large-scale hierarchical image database 248–255. https://doi.org/10.1109/cvpr.2009.5206848.

[98] H. He, E.A. Garcia. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng., 21 (9) (2009), pp. 1263-1284.

[99] Richter CD, Samala RK, Chan H-P, Hadjiiski L, Cha KH. Generalization Error Analysis: Deep Convolutional Neural Network in Mammography. Proc SPIE. 2018;10575:1057520.

[100] Galar M., Fernandez A., Barrenechea E., Bustince H., Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2012;42(4):463–484. doi: 10.1109/tsmcc.2011.2161285.

[101] Sun Z., Song Q., Zhu X., Sun H., Xu B., Zhou Y. A novel ensemble method for classifying imbalanced data. Pattern Recognition. 2015;48(5):1623–1637. doi: 10.1016/j.patcog.2014.11.014.

[102] Seliya, N., Abdollah Zadeh, A. &Khoshgoftaar, T.M. A literature review on one-class classification and its potential applications in big data. J Big Data 8, 122 (2021). https://doi.org/10.1186/s40537-021-00514-x.

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 an Associated 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 computer graphics, geographical information systems, medical images system, and information systems.

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 Mahjoubiis 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.