Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 12 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i12.49709

PAPER

# Deep Learning-Based Approaches Using Medical Imaging for Therapy Response Prediction in Breast Cancer: A Systematic Literature Review

Laila El Jiani[1](✉), Sanaa El Filali[1], El Habib Ben Lahmar[1], Ihsane Haloum[2]

[1]Faculty of Sciences Ben Msick, Hassan II University, Casablanca, Morocco

[2]Faculty of Medicine and Pharmacy of Casablanca, Hassan II University, Casablanca, Morocco

Laila.eljiani-etu@etu.univh2c.ma

**ABSTRACT**

The prediction of response to breast cancer therapy involves assessing the effectiveness of treatment by comparing biomarker levels before and after treatment. Deep learning (DL) models can provide a non-invasive and early way to evaluate the response to therapy based on medical imaging analysis. We conducted this systematic review to investigate the current DL based methods for predicting breast cancer therapy response using medical imaging. This review included 19 studies based on the PRISMA methodology. Some selected studies personalized the Convolutional Neural Network (CNN) architecture to improve its performance in handling medical images, while others used pre-trained models. The accuracy rates range from 0.73 to 0.90, and the Area Under the Curve (AUC) reaches 0.98. Our study's findings suggest that the performance of these approaches varies depending on various medical imaging modalities, the nature of the DL architecture used, and the fusion of training data sources. However, several challenges related to their explainability and generalizability arise. Therefore, it is necessary to develop larger datasets and broaden the scope of current studies to include multi-center studies.

**KEYWORDS**

deep learning (DL), therapy response prediction, medical image, breast cancer

## 1 INTRODUCTION

With approximately 2.3 million cases diagnosed in 2022, breast cancer is considered one of the most common cancers among women worldwide, and it is the primary cause of cancer-related mortality in many countries, with a rate of 6.9% [1]. After a breast cancer diagnosis, patients must begin a therapy protocol that varies depending on the cancer's stage [2]. Many women diagnosed with early-stage breast cancer are eligible for breast-conserving surgery followed by radiotherapy, partial

El Jiani, L., El Filali, S., Lahmar, E.H.B., Haloum, I. (2024). Deep Learning-Based Approaches Using Medical Imaging for Therapy Response Prediction in Breast Cancer: A Systematic Literature Review. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(12), pp. 37–54. https://doi.org/10.3991/ijoe.v20i12.49709

Article submitted 2024-04-17. Revision uploaded 2024-06-14. Final acceptance 2024-06-14.

mastectomy, or chemotherapy. The therapy can be adjuvant or neoadjuvant [3]. Clinical assessments, including physical exams, can assess a tumor's response to treatment in conjunction with pathologic measurements gathered from a biopsy or surgical resection [4]. Medical practitioners also use various imaging techniques, such as Magnetic Resonance Imaging (MRI), mammography, and ultrasound, to evaluate the effectiveness of treatment and monitor the response. Among them, MRI is known for having the highest accuracy rate. Currently, it's common practice to combine MRI with mammography to obtain complementary information about the effectiveness of the therapy [5]. Unfortunately, none of these techniques can replace a biopsy for accurately tracking response to treatment and providing crucial information about breast tumors [6].

The need for a non-invasive way to assist physicians in the diagnosis, monitoring, and prognosis of breast cancer patients led researchers to explore the path of artificial intelligence and, more specifically, the field of machine learning [7], [8]. To date, the field of radiology has the most widely implemented machine learning and deep learning. The in-depth analysis of medical imaging allows for the extraction of multiple imaging parameters, known as radiomics [9]. Many valuable radiomics-based approaches have been experienced in lesion detection, risk prediction, and the prediction of therapy response [10]. However, these radiomics-based approaches have some limitations, including the need for handcrafted feature extraction and manual segmentation of the lesions [11]. Recently, new applications have been developed taking advantage of advances in computer vision. This is a sub-field of deep learning that operates directly on the image data and extracts features automatically without the necessity of human intervention [12]. Convolutional Neural Networks (CNN) stand out as a prevalent method within the realm of image analysis. CNN is adept at tasks involving pixel data, making it particularly suitable for image processing, computer vision applications, and scenarios requiring object recognition, such as facial recognition and medical image analysis. Notably, CNN methods can eliminate the need for radiologists to manually outline tumors in images [13].

The prediction of response to breast cancer therapy refers to the early evaluation of treatment effectiveness through a comparison of the levels of biomarker acceptance before and after one or many cycles of systemic therapy. Thus, the sooner the response information is obtained, the better it is for therapy adjustment [14]. We conducted this systematic literature review to investigate the existing approaches based on DL and medical imaging aimed at predicting the response to therapy in breast cancer patients. In the literature, many reviews have been carried out to answer many research questions relating to the use of DL and medical image analysis to predict therapy response. However, most of the existing reviews have limitations in scope, focusing on one image modality, such as MRI, ultrasound, or mammography, and a specific therapy. Moreover, they do not differentiate between DL models that require manual segmentation and those that use the whole image as input [15–18]. Therefore, we conducted a review of the current methods for predicting therapy response using medical imaging that take a broader approach and encompass multiple modalities and therapies to develop a deeper comprehension of the topic. Our focus was on Deep Learning and its related practices, including the use of pre-trained models, transfer learning, and techniques for enhancing data quality. We aimed to gain an updated and comprehensive overview of recent approaches through this review, as well as identify new

research questions, challenges, and areas that require further research. We also focused on the use of whole-breast images acquired by different imaging modalities (ultrasound, MRI, and mammography). Additionally, our review involved a comparative analysis of performance metrics across different studies focused on predicting therapy response in breast cancer.

## 2 METHODOLOGY

To conduct our literature review, we followed the PRISMA guidelines, which stand for preferred reporting items for systematic reviews and meta-analyses. PRISMA is one of the most popular and standard review methodologies since it offers a reproducible and standardized approach for the identification, selection, and evaluation of existing studies. Additionally, it guides on selecting, recognizing, and evaluating studies consistently [19]. The key steps for conducting reviews using PRISMA guidelines are as follows:

- Defining the research question
- Developing inclusion and exclusion criteria
- Conducting a systematic search
- Selecting eligible studies
- Assessing quality
- Extracting and synthesizing data
- Interpreting results

Following the previous process, the PRISMA flowchart in Figure 1 summarizes the details at each step.

### 2.1 Research question and objectives

We formulate the following research question to guide our literature review: "What are the current deep learning-based approaches that use medical imaging to predict therapy response in breast cancer patients?" The objective is to gather comprehensive information about the popular architectures, their performance, the used data, and their eventual limitations.

### 2.2 Data sources and search strategy

To collect literature data for this review, three bibliographic databases were surveyed: Scopus, Web of Science, and Google Scholar. We selected Scopus and Web of Science due to their extensive coverage, credibility, and ability to provide cross-disciplinary access to high-quality, peer-reviewed content. Google Scholar served as a complementary tool to access a broad range of gray literature, such as theses and conference papers. This strategy allows us to capture the most extensive spectrum of relevant academic information available, providing a robust foundation for our analysis. To get the most updated studies, we limited our search to the last five years corresponding to the period between 2019 and December 2023, and we selected only articles written in English.

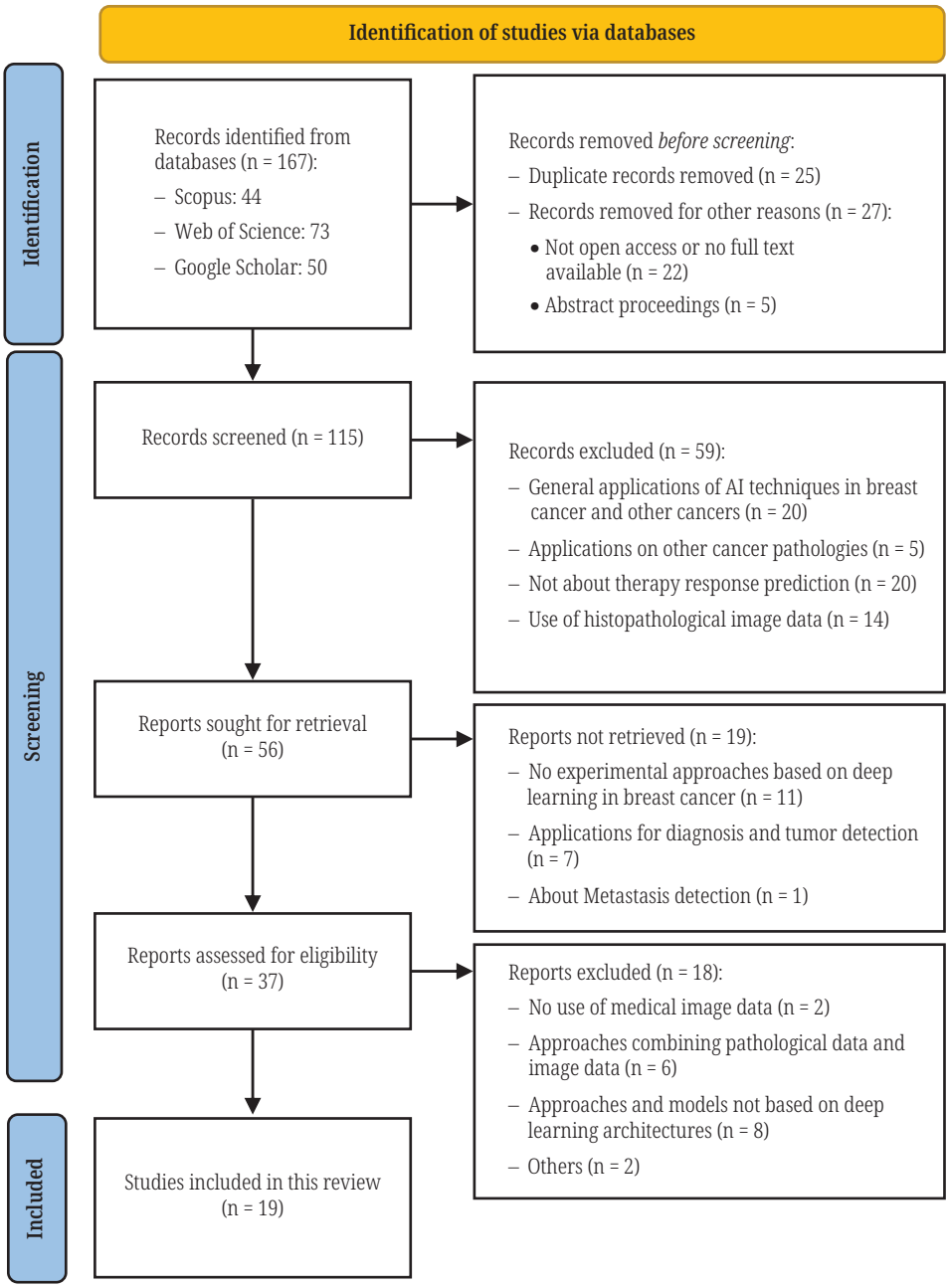


**Fig. 1.** PRISMA flowchart describing the study selection process for the review

The keywords used to perform the search are: "deep learning," "approaches," "medical image," "medical imaging," "therapy," "response prediction," and "breast cancer."

The search strategy used the following search strings (SS):

- **SS1:** "deep learning" AND "approaches" AND "Medical Image" AND "Therapy Response Prediction" AND "Breast Cancer"
- **SS2:** "deep learning" AND "approaches" AND "Medical Imaging" AND "Response Prediction" AND "Breast Cancer" AND "Therapy"

## 2.3 Selection (inclusion and exclusion) criteria

After performing the search strategies in the pre-cited databases, a total of 167 studies were obtained. At the identification step, the duplicated results, the non-open access and not full-text papers, and the abstract proceeding were excluded. As a result, 115 studies were retained for the screening step.

The screening of the 115 results based on title and abstract allowed the elimination of 59 other research papers; 20 of them were about general applications of artificial intelligence in breast cancer and did not include experimentation. At this stage, the main exclusion criteria considered are the application to cancers other than breast cancer, the use of histopathological imaging as a type of data for the approaches, and the treatment of another prediction problem other than therapy response.

After reading and deeply analyzing the full-text versions of the screened papers, we assessed eligibility and included 19 studies in this review. The different criteria for inclusion and exclusion have been summarized in Table 1.

**Table 1.** List of inclusion and exclusion criteria

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| Publication between 2019 and 2023 | Treating a subject other than that of predicting response to therapy |
| Use of English as a language | Considering a pathology other than breast cancer |
| Being a conference or journal paper | Absence of experimental approaches based on deep learning |
| Being about the prediction of therapy response in breast cancer patients | Use of histopathological image data |
| Use of medical image as a type of data for training models | No use of medical image data |
| Use of deep learning architecture-based models | Not open access or no full text available |
| Use of whole breast image datasets | Manual extraction of features from breast images |

All the included studies were about the development of new approaches based on deep learning with the objective of predicting breast cancer therapy response using medical imaging.

## 2.4 Quality assessment

The objective of quality assessment in a systematic literature review (SLR) is to meticulously examine the methodological robustness and reliability of the included studies. As a result, it serves as a meticulous filter to sift through full-text articles, representing the concluding phase in the preparation of the study pool for subsequent data extraction and synthesis [20]. According to Yu Xiao et al. [21], there is no consensus on how reviewers should deal with quality assessment in their reviews, but they should collaboratively determine their stance on quality assessment, taking into account the specific circumstances and nuances that characterize their study.

To assess the quality of the included studies, two control levels were performed. The first control focuses on methodology quality through an elaborated checklist of nine quality assessments (QA). All the included studies must cover all the

criteria in Table 2. The second level of control aims to precisely meet the inclusion/exclusion criteria. For this purpose, another checklist with five questions (Q) was developed, as presented in Table 3. To be considered in this review, the study must have a yes answer to each of the questions; otherwise, it will be excluded from the final literature list.

**Table 2.** List of quality assessment questions

| ID | Quality Assessment Questions |
|---|---|
| QA1 | Has the research objective been clearly defined? |
| QA2 | Did the study have a clear research question? |
| QA3 | Is the target population well described? |
| QA4 | Is the methodology and experimental process clearly outlined? |
| QA5 | Is the data collection process well described? |
| QA6 | Were the results demonstrated by experimentation? |
| QA7 | Are the results clearly presented and interpreted? |
| QA8 | Are the reporting coherent and transparent? |
| QA9 | Are the findings of the study generalizable? |

## 2.5 Data extraction

From the included studies, we extracted general data such as title, authors, year of publication, and the objective of the research. More specific data was also extracted, such as image modality, type of therapy, subtype of breast cancer, level of response to the therapy, dataset, DL architecture, and performance metrics.

**Table 3.** List of inclusion/exclusion criteria assessment questions

| ID | Inclusion/Exclusion Criteria Assessment Questions |
|---|---|
| Q1 | Are breast cancer patients the target population of the study? |
| Q2 | Is therapy response prediction the objective of the study? |
| Q3 | Is there a medical image-based and deep learning-based approach developed in the study? |
| Q4 | Does the developed approach use the breast cancer image without a prior manual feature selection step? |
| Q5 | Does the developed approach predict outcomes without relying on biopsy data? |

## 2.6 Data synthesis

According to the selected studies' data analysis, there has been a noticeable trend in publication numbers from 2019 to 2023. The chart in Figure 2a displays the publication trends during this period. The overall trend shows an upward trajectory in publications, with 2022 being the peak year. The data suggests a growing interest in research and scholarly work during this period. Regarding the datasets used for training, we notice that there is a prevalence of private datasets in research. Only a

small number of public datasets are being utilized, with approximately five public datasets and 14 private datasets being used (see Figure 2b).

In our analysis, we observe that the prevalence of custom CNN architectures remains evident as being the most frequently employed in the selected studies. Also, we notice that there are eight instances of pre-trained CNNs that have been utilized, such as AlexNET, VGG16, ResNet50, and Unet. This finding confirms the practicality of transfer learning, which lets researchers adapt and fine-tune existing architectures, saving valuable time and computational resources (see Figure 3a).

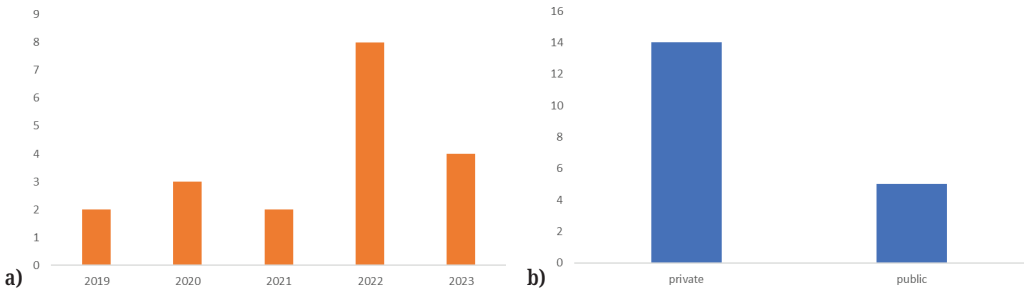


**Fig. 2.** (a) Distribution of selected studies by year of publication (b) Distribution of training datasets used in the selected studies by type

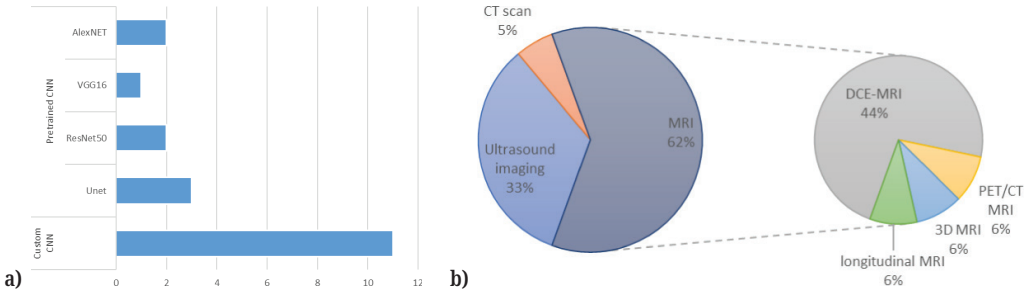


**Fig. 3.** (a) Deep learning architectures used in the selected studies (b) Different image modalities used in the relevant datasets

The retained research studies have utilized various medical imaging modalities, as illustrated in Figure 3b. We noticed that the most commonly used imaging modality is MRI, which includes various subtypes. Ultrasound follows as a close second. This can be attributed to the greater availability of MRI datasets. On the other hand, CT scans are not as frequently used, despite their capacity to perform several functions, but they are still a costly imaging modality.

# 3 RESULTS AND DISCUSSION

## 3.1 Description of the selected studies

Different approaches emerged in the selected papers. Some of them had personalized the CNN architecture to enhance its performance when dealing with medical image, other authors preferred to harness the strengths of pre-trained models and build models based on them. We have also noticed that many studies used clinical, demographic, or molecular data as complementary sources to improve the accuracy of the prediction. Additionally, all the studies under consideration were focused on

neoadjuvant chemotherapy (NAC), and none of them examined the prediction of other types of therapies. Many researchers consider that neoadjuvant chemotherapy is the typical treatment for breast cancer because it can diminish the size of tumors and improve overall outcomes even at advanced stages [22]. Neoadjuvant chemotherapy facilitates also downstaging large tumors, rendering previously inoperable ones suitable for surgery and enabling breast-conserving surgery in situations where mastectomy was formerly the single choice [23]. However, histopathological assessment following breast surgery is crucial for measuring the response to NAC. One of the metrics used for the evaluation of the efficiency of NAC is pathological complete response (pCR). Essentially, it indicates the absence of any detectable residual cancer cells in the tissue removed during surgery following a course of treatment, typically chemotherapy or radiation therapy [24]. In addition to pCR, there are other metrics to evaluate treatment response in breast cancer, which are based on changes in tumor cellularity and observation of regressive changes in residual tumor tissue [3].

After conducting our review, we observed that all the analyzed studies were limited to the prediction of pCR as a metric for response assessment and NAC as a targeted therapy. Only one study considered the partial response among the predicted classes. In the following, the retained approaches are classified into two families: the first is based on customized CNN (see Table 4), and the second is based on pre-trained architectures (see Table 5).

## 3.2 Therapy response prediction approaches based on custom convolutional neural network architecture

Based on a custom CNN, several approaches and models were developed in the selected papers to predict a complete response to NAC. We will discuss the findings by image modality hereafter.

**Approaches using magnetic resonance imaging datasets:** Magnetic resonance imaging is a suitable imaging technique to monitor disease progression. There are different types of MRI, including structural MRI, functional MRI, diffusion MRI, and DCE-MRI, which stands for Dynamic Contrast-Enhanced MRI. To face the absence of imaging metrics that can effectively predict the response to NAC before it starts, Ravichandran et al. [25] utilized the I-SPY1 dataset, which consists of DCE-MRI scans on a per-voxel basis for 166 patients, to train a CNN-based model to predict the pCR to NAC. The developed model consisted of six blocks and was evaluated through two case studies. The CNN achieved a high prediction accuracy of 0.82 and an area under the curve (AUC) of 0.77 using pre-contrast data. The study suggested that identifying areas with greater predictive value could improve DL-guided treatment planning. However, the small training dataset and mix of molecular subtypes are limitations that can affect the effectiveness of this approach. Likewise, Qu et al. [26] conducted a study where they used a dataset of 302 patients with locally advanced breast cancer (LABC) to train a CNN architecture. The network was fed with six phases of enhancement of pre- and post-NAC images. The study resulted in improved accuracy, achieving an AUC of 0.97. Nevertheless, supposing that T1 weighted image pre-contrast and post-contrast sequences of breast MRI are the most characteristic could be a limitation to this approach and may raise questions about its accuracy.

Early prediction of a patient's response to NAC in breast cancer treatment is essential for informing therapy decisions and enhancing the personalization and effectiveness of patient care. For this purpose, varying time points can provide a more comprehensive understanding of the longitudinal alterations in primary tumors.

**Table 4.** Summary of approaches based on custom CNN

| Authors | Dataset | | | BC Subset | Imaging Modality | Cycles | Arch | DA | Seg. | Performance Metrics | | | | Other Data Sources |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Name | S. Size | Labels Dist. | | | | | | | Acc. | Spec. | Sen. | AUC | |
| Ravichandran et al. [25] | I-SPY1 Clinical trial | 166 | 49/117 | Localized | DCE-MRI | NS | CNN | Yes | Manual | 0.82 | NS | NS | 0.77 | Clinical |
| Xie et al. [30] | Private | 114 | 39/75 | LABC | Ultrasound | 1 | CNN | Yes | Manual | 0.87 | 0.85 | 0.9 | 0.93 | |
| Qi et al. [35] | Private | 324 | NS | LABC | CT Scan | NS | CNN | No | Manual | NS | NS | NS | 0.75 | Clinical |
| Verma et al. [27] | I-SPY1 Clinical trial | 121 | 37/84 | Localized | DCE-MRI | 4 | 3D-CNN | Yes | None | 0.85 | 0.76 | 0.9 | 0.88 | Clinical |
| Zhou et al. [29] | Private | 210 | 101/109 | TNBC | DCE-MRI | 4 | CNN | Yes | Manual | 0.77 | 0.81 | 0.82 | 0.86 | |
| Duanmu et al. [29] | I-SPY-1 TRIAL (2002–2006) | 112 | NS | Localized | 3D MRI | 1 | CNN | No | None | 0.83 | 0.88 | 0.68 | 0.8 | Molecular, demographic |
| Taleghamar et al. [31] | Private | 181 | 138/43 | LABC | Quantitative Ultrasound | NS | CNN | Yes | None | 0.88 | 0.92 | 0.7 | 0.86 | |
| Dammu et al. [44] | I-SPY-1 TRIAL (2002–2006) | 155 | NS | Localized | Longitudinal multiparametric DCE MRI | 2 | CNN | No | Semi-automatic | 0.81 | 0.86 | 0.68 | 0.83 | Molecular, demographic |
| Gu, et al. [32] | Private | 168 | 86/82 | Localized | Ultrasound | 4 | CNN | No | Manual | NS | 0.9 | NS | 0.93 | |
| Qu, et al. [26] | Private | 302 | 132/170 | LABC | DCE-MRI | 4 to 6 | CNN | Yes | Manual | NS | NS | 0.96 | 0.97 | Molecular, demographic |
| Jiang, et al. [33] | Private | 592 | 188/404 | LABC | Ultrasound | NS | DCNN | No | Manual | 0.84 | 0.83 | 0.83 | 0.94 | Clinical |

*Notes:* **S. size:** Sample size; **Labels dist:** Labels distribution (responder/No responder); **Cycles:** Therapy cycles; **Arch:** DL architecture; **DA:** Data augmentation; **Seg:** Segmentation method; **Spec:** specificity; **Sen:** Sensitivity; **Acc:** accuracy.

To reach this goal, Verma et al. [27] conducted experiments using the ISPy-1 trial dataset and proposed deep-NST, an end-to-end multimodal spatiotemporal DL framework. Deep-NST integrates four parallel 3D-CNN networks trained on DCE-MRI imaging data from different stamps. The developed framework includes molecular and demographical data that was fused using a cross-kernel feature fusion (CKFF) module. An AUC of 0.88 was registered. One of the strengths of this approach is its ability to predict pCR from pretreatment imaging alone, as well as the fusion of different data sources, which allows for improved prediction accuracy. Similarly, Duanmu et al. [28] developed three CNN models based on MRI imaging data in 3D format, demographic data, and molecular data from the I-SPY1 trial dataset. The models include an imaging data-only model, a parallel model, and an interactive model. One advantage of the proposed approach is that it operates on the entire three-dimensional MRI image without the need for prior manual tumor segmentation. The prediction was based only on the pre-treatment time point, and the best-reached accuracy was 0.80. However, more extensive testing on a larger sample size across multiple institutions is required.

Triple-negative breast cancer is the most aggressive type of breast cancer, characterized by the absence of three biomarkers: HER2, ER, and PR. To gain better insights into the progression of this type of cancer, Zhou et al. [29] laid emphasis on the use of multiparametric MRI, which includes additional imaging sequences or parameters such as T1-weighted, T2-weighted, and diffusion-weighted, in addition to DCE. The authors developed a CNN-based model to predict pCR, which achieved high accuracy in training and validation and reached an AUC of 0.86 in testing groups. The only limitations of this study are the exclusion of patients with axillary lymph node residual disease and the small size of the considered sample.

**Approaches using ultrasound datasets:** In addition to MRI, several studies have investigated the use of ultrasound image datasets to suggest non-invasive and practical methods for predicting personalized responses in breast cancer patients during NAC treatment. For this aim, a dual-branch CNN was proposed by Xie et al. [30] to predict early NAC response at different stages of chemotherapy in 114 patients with LABC. The best results were obtained when using a 9-block CNN, combining training data from pre-NAC and the first cycle of NAC, and using feature element sum as a sharing method. An AUC of 0.939 was achieved. Feature sharing enables the model to account for correlations between data at various stages of NAC during training.

Based on quantitative ultrasound multiparametric imaging (QUS) and deep convolutional neural network architecture (DCNN), Taleghamar et al. [31] developed an approach combining two cascaded networks: a residual network and a residual attention network to predict the response to therapy of LABC patients. The attention-guided network has shown better performance in extracting optimal quantitative features from QUS multi-parametric images, achieving an AUC score of 0.86. In [32], the authors proposed a novel deep-learning radiomics pipeline (DLRP) for response prediction in breast cancer patients who have undergone four courses of NAC. The dataset was composed of ultrasound images of 168 patients. The proposed model is based on CNN and enables sequential prediction of response at various time points during NAC administration. The best results were achieved when training with data from the four NAC courses, with an AUC of approximately 0.9. Working towards the same objective, Jiang et al. [33] developed and validated a deep learning radiomic nomogram (DLRN) trained on ultrasound images of 365 patients captured before the beginning of NAC and after the first cycle. The AUC was approximately 0.9. The common limitation of all these studies is their use of

**Table 5.** Summary of approaches based on pretrained architectures

| Authors | Dataset | | | BC Subset | Imaging Modality | Cycles | Pre. Arch | DA | Seg. | Performance Metrics | | | | Other Data Sources |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Name | S. Size | Labels Dist. | | | | | | | Acc. | Spec. | Sen. | AUC | |
| Choi. et al. [37] | Private | 56 | 6/50 | LABC | PET/CT MRI | 3 | Alexnet | Yes | | NS | NS | NS | 0.88 | |
| Liu Y, et al. [38] | Private | 393 | 114/279 | HER2-positive | Longitudinal Ultrasound | 2 | Unet | Yes | automatic | NS | 0.87 | 0.86 | 0.97 | Clinical |
| Joo et al. [39] | Private | 536 | 133/403 | Invasive | MRI | 6 | 3D Resnet50 | No | automatic | 0.85 | 0.93 | 0.66 | NS | Clinical |
| Peng et al. [43] | Private | 356 | 83/273 | Localized | DCE-MRI | 2 | ResNet50 | Yes | Manual | 0.77 | 0.76 | 0.78 | 0.83 | |
| Ha et al. [44] | Private | 144 | 46/57/38* | LABC | MRI | 2 | VGG16 | Yes | Manual | 0.77 | 0.81 | 0.82 | 0.86 | |
| Massafra et al. [36] | I-SPY-1 TRIAL (2002–2006) | 151 | 64/161 | Localized | DCE-MRI | 2 | Alexnet | No | semi-automatic | Sag: 0.88 Ax: 0.773 | Sag: 0.90 Ax: 0.80 | Sag: 0.69 Ax: 0.71 | Sag: 0.803 Ax: 0.78 | Molecular |
| Li et al. [41] | Private | 95 | 24/71 | Localized | DCE-MRI | 3 | Uctransunet | No | Manual | 0.89 | 0.90 | 0.87 | 0.91 | Clinical |
| Wu et al. [42] | Private | 801 | 242/595 | Localized | Ultrasound | 2 | Unet | No | Manual | NS | NS | NS | 0.87 | Molecular, demographic |

*Notes:* **S. size:** Sample size; **Labels dist:** Labels distribution; **Cycles:** Therapy cycles; **Pre Arch:** Pretrained architecture; **DA:** Data augmentation; **Seg:** Segmentation method; **Spec:** specificity; **Sen:** Sensitivity; **Acc:** accuracy. **Sag:** Sagittal; **Ax:** Axial; *(responder/Partial/No responder).

small datasets or their conduct in a single center, which underscores the necessity for larger and multi-institutional datasets to enhance their generalizability. Also, the pCR prediction models based on US imaging are often considered by practitioners as not sufficiently robust, therefore, they advise holistically evaluating treatment strategies using diverse sources of information.

**Approaches using CT scan datasets:** The CT scan imaging modality has the advantages of being more efficient in assessing soft tissue tumors and distinguishing between tissues, but it is not as commonly used as MRI and ultrasound modalities because of its high costs and its radiation exposure [34]. Tan Hong Qi et al. [35] developed a predictive AI model based on CT imaging and clinical parameters. The proposed deep learning model has been trained on a multi-center dataset containing CT scans of 342 patients with LABC. By including clinical parameters, the AUC of pCR prediction has improved from 0.743 to 0.772. However, the study acknowledged limitations, such as its retrospective nature and the relatively small size of the cohort.

### 3.3 Therapy response prediction approaches based on pre-trained architectures

Deep learning models require large training datasets. To address the lack of data, transfer learning strategies offer a promising solution to address these constraints. In [36], Massafra et al. developed an AI method to predict early response to treatment by exploiting the data of 151 patients, including axial and sagittal DCE MRI. The approach follows a three-step process: First, automatic feature extraction is performed using the pre-trained CNN Alexnet; then, important features are extracted using the stratified feature selection method; and finally, the classification of responders is achieved. The results were enhanced using additional clinical data, reaching the best AUC of 0.80 on the sagittal MRI data. To improve the generalizability of the model, both private and public datasets were utilized for training. Based on the Alexnet architecture, Choi et al. [37] introduced a DL model tailored for PET and MRI images and compared its performance to conventional methods. The model was trained using a dataset of 56 advanced breast cancer patients who had undergone three cycles of NAC. Although the use of data augmentation contributed to the parametric improvement of the deep learning model, the imbalance rate between the two considered classes remains a limitation of this study due to the small size of the dataset.

Using U-net architecture, Lui et al. [35] proposed an approach predicting PCR to NAC by combining ultrasound imaging data and clinical data of 393 patients with HER2-positive breast cancer. The proposed architecture consisted of two subnetworks, one for tumor segmentation and the other for pCR prediction. The latter subnetwork uses the extracted features from the tumor segmentation subnetwork as its input. To enhance the prediction, a clinician model using multivariable logistic regression analysis was added. Hence, the achieved AUC was 0.90. However, there may be limitations due to the separate misanalysis of data from the two NAC cycles, which could impact the results. In [39], Joo et al. proposed a novel approach for predicting response to NAC by combining MRI imaging and a Resnet50 architecture. The approach involved a model consisting of three parts: a 3D volumetric CNN for extracting MR features, a layer for clinical information and feature concatenation, and a fully connected layer for predicting PCR. The authors experimented with different combinations of datasets, including clinical data and T1/T2-weighted MRIs, and found that the best results were obtained when all datasets were used together. The approach achieved an AUC of 0.88. Some limitations were identified for these

studies and concern the explainability of deep learning models and the generability of the results.

Based on the pre-trained VGG16 architecture, Ha et al. [40] developed a model to classify patients with LABC into three classes: response, partial response, and complete response, using an MRI breast cancer dataset of 144 patients. The model achieved an accuracy of 0.88. In [41], Li et al. developed a model combining multi-period image information and clinical characteristics to predict pCR. The DCE-MRIs of 95 patients were involved in the experimentation. The model was based on the pre-trained network Uctransunet to extract semantic segmentation features. The model achieved an AUC of 0.90 when using multi-period images and clinical data. The previous studies were based on a population from a single center, which can lead to biased results. To avoid this problem, Wu et al. [42] led a multi-center and retrospective study involving 801 patients dispatched in 4 cohorts and developed an auto segmentation-based ultrasonography assessment system to predict pCR. The system is based on Unet, which includes automatic tumor segmentation and a model called SUAS for predicting optimal therapeutic management after NAC. Even though a high AUC of 0.97 was achieved, there were some limitations to the developed approach, such as the imbalanced distribution of patients among the four cohorts. To compare the performance of deep learning-based models to radiomics analysis-based models, Peng et al. [43] developed four models based on Unet, using kinetic, imaging, and molecular data from a private dataset of 356 patients. The best-performing model was a DL model that utilized image, kinetic, and molecular data.

## 3.4 Discussion

Through the present systematic literature review, we attempted to answer our research question: "What are the current deep learning-based approaches that use medical imaging to predict therapy response in patients with breast cancer?" We identified several approaches with accuracy rates ranging from 0.73 to 0.90 and an AUC reaching 0.98. The difference in the performance of these approaches is highly dependent on the imaging modality used, the nature of the deep learning architecture used, and the fusion of training data sources. In addition to identifying DL approaches, this review enabled us to identify some factors that can influence the effectiveness of the models. These factors include:

- *The imaging modalities of the training datasets:* MRI and ultrasound imaging modalities were associated with most approaches with high accuracy. This can be explained by the high resolution of MRI images and their soft tissue contrast, which enhance the quality of training data. Additionally, ultrasound allows the evaluation of changes in tumor size and morphology throughout the course of NAC. In contrast, CT scans are less effective in detecting small changes in tumor size, especially during early treatment stages, and may struggle to distinguish between scar tissue and active cancer cells.
- *The use of pre-trained architectures:* Several pre-trained architectures, including U-Net, VGG16, and ResNet50, were utilized in the studies that were selected to improve tumor segmentation and feature extraction in the developed models. The use of pre-trained models improved the segmentation process's accuracy and reliability when delineating tumor boundaries. Furthermore, transfer learning has helped overcome the limitations of medical imaging datasets in terms of variations in image quality, acquisition protocols, and equipment.

- *The integration of multimodal data fusion:* Most retained studies evaluated the effectiveness of the proposed deep learning approaches when using only medical image data, as well as when combining other sources of data such as clinical, demographic, or molecular data. These studies have also explored the integration of data from different time stamps. To combine features from different modalities, approaches such as concatenation, summing, attention-based fusion, cross-kernel fusion, and the stack approach were used. The results have shown that combining different sources of data leads to the best outcomes.

## 4 CONCLUSION

In this systematic literature review, we have examined the recent DL-based methods for predicting therapy response using medical image datasets of breast cancer patients. To get an overall view, we did not limit the study to a single therapy or a specific imaging modality. The findings of this study cannot conclusively determine the best approach due to the varying methods involved, which makes it difficult to compare them. This study's limitation is due to the heterogeneity of the involved methods. However, we have identified some factors that affect the performance of the proposed models. These factors include the imaging modalities used in the training datasets, the utilization of pre-trained architecture, and the integration of multimodal data fusion. Moreover, there are significant challenges related to the explainability and generalizability of DL approaches in medical imaging. The complexity of DL models often makes it difficult to understand how decisions are made, which is a critical barrier in clinical environments where trust and transparency are paramount. Furthermore, these models frequently exhibit variability in performance when applied to different groups of patients or used with various imaging equipment, highlighting concerns about their reliability and broad applicability. Resolving these challenges is essential for the effective integration of DL into clinical practice.

## 5 REFERENCES

[1] J. Zhang, J. Wu, X. S. Zhou, F. Shi, and D. Shen, "Recent advancements in artificial intelligence for breast cancer: Image augmentation, segmentation, diagnosis, and prognosis approaches," *Seminars in Cancer Biology*, vol. 96, pp. 11–25, 2023. https://doi.org/10.1016/j.semcancer.2023.09.001

[2] M. Tarousi, P. Bountris, P. Daskalakis, and D. D. Koutsouris, "Development of a cost-effective intelligent clinical decision support system for breast cancer early diagnosis and triage," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 5, pp. 43–64, 2022. https://doi.org/10.3991/ijoe.v18i05.29067

[3] T. A. Moo, R. Sanford, C. Dang, and M. Morrow, "Overview of breast cancer therapy," *PET Clinics*, vol. 13, no. 3, pp. 339–354, 2018. https://doi.org/10.1016/j.cpet.2018.02.006

[4] N. Avril, S. Sassen, and R. Roylance, "Response to therapy in breast cancer," *Journal of Nuclear Medicine*, vol. 50, (Suppl. 1), pp. 55S–63S, 2009. https://doi.org/10.2967/jnumed.108.057240

[5] H. P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," *The British Journal of Radiology*, vol. 93, no. 1108, p. 20190580, 2020. https://doi.org/10.1259/bjr.20190580

[6] W. L. Bi *et al.*, "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 2, pp. 127–157, 2019. https://doi.org/10.3322/caac.21552

[7] R. Cuocolo, M. Caruso, T. Perillo, L. Ugga, and M. Petretta, "Machine learning in oncology: A clinical appraisal," *Cancer Letters*, vol. 481, pp. 55–62, 2020. https://doi.org/10.1016/j.canlet.2020.03.032

[8] H. H. Muljo, A. S. Perbangsa, Y. Yulius, and B. Pardamean, "Mobile learning for early detection of cancer," *International Journal of Interactive Mobile Technologies*, vol. 12, no. 2, pp. 39–53, 2018. https://doi.org/10.3991/ijim.v12i2.7814

[9] P. D. Moyya and M. Asaithambi, "Quantitative analysis of breast cancer NACT response on DCE-MRI using Gabor filter derived radiomic features," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 12, pp. 106–122, 2022. https://doi.org/10.3991/ijoe.v18i12.32501

[10] A. Bitencourt, I. Daimiel Naranjo, R. Lo Gullo, C. Rossi Saccarelli, and K. Pinker, "AI-enhanced breast imaging: Where are we and where are we heading?" *European Journal of Radiology*, vol. 142, p. 109882, 2021. https://doi.org/10.1016/j.ejrad.2021.109882

[11] Z. Liu *et al.*, "The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges," *Theranostics*, vol. 9, no. 5, pp. 1303–1322, 2019. https://doi.org/10.7150/thno.30309

[12] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of deep learning in breast cancer imaging," *Seminars in Nuclear Medicine*, vol. 52, no. 5, pp. 584–596, 2022. https://doi.org/10.1053/j.semnuclmed.2022.02.003

[13] M. A. Jones, W. Islam, R. Faiz, X. Chen, and B. Zheng, "Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction," *Frontiers in Oncology*, vol. 12, 2022. https://doi.org/10.3389/fonc.2022.980793

[14] D. Vasudevan, P. S. Jayalakshmy, S. Kumar, and S. Mathew, "Assessment of pathological response of breast carcinoma in modified radical mastectomy specimens after neoadjuvant chemotherapy," *International Journal of Breast Cancer*, vol. 2015, 2015. https://doi.org/10.1155/2015/536145

[15] N. Khan, R. Adam, P. Huang, T. Maldjian, and T. Q. Duong, "Deep learning prediction of pathologic complete response in breast cancer using MRI and other clinical data: A systematic review," *Tomography*, vol. 8, no. 6, pp. 2784–2795, 2022. https://doi.org/10.3390/tomography8060232

[16] R. Granzier, T. van Nijnatten, H. Woodruff, M. Smidt, and M. Lobbes, "Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: A systematic review," *European Journal of Radiology*, vol. 121, p. 108736, 2019. https://doi.org/10.1016/j.ejrad.2019.108736

[17] X. X. Yin, S. Hadjiloucas, Y. Zhang, and Z. Tian, "MRI radiogenomics for intelligent diagnosis of breast tumors and accurate prediction of neoadjuvant chemotherapy responses-A review," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106510, 2022. https://doi.org/10.1016/j.cmpb.2021.106510

[18] X. Liang, X. Yu, and T. Gao, "Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis," *European Journal of Radiology*, vol. 150, p. 110247, 2022. https://doi.org/10.1016/j.ejrad.2022.110247

[19] H. A. Mohamed Shaffril, S. F. Samsuddin, and A. Abu Samah, "The ABC of systematic literature review: The basic methodological guidance for beginners," *Quality & Quantity*, vol. 55, no. 4, pp. 1319–1346, 2020. https://doi.org/10.1007/s11135-020-01059-6

[20] P. V. Torres-Carrion, C. S. Gonzalez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, "Methodology for systematic literature review applied to engineering and education, in *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2018. https://doi.org/10.1109/EDUCON.2018.8363388

[21] Y. Xiao and M. Watson, "Guidance on conducting a systematic literature review," *Journal of Planning Education and Research*, vol. 39, no. 1, pp. 93–112, 2017. https://doi.org/10.1177/0739456X17723971

[22] M. Asaoka, S. Gandhi, T. Ishikawa, and K. Takabe, "Neoadjuvant chemotherapy for breast cancer: Past, present, and future," *Breast Cancer: Basic and Clinical Research*, vol. 14, 2020. https://doi.org/10.1177/1178223420980377

[23] H. Wang and X. Mao, "Evaluation of the efficacy of neoadjuvant chemotherapy for breast cancer," *Drug Design, Development and Therapy*, vol. 14, pp. 2423–2433, 2020. https://doi.org/10.2147/DDDT.S253961

[24] M. Takaoka *et al.*, "Pathological complete response patients after neoadjuvant chemotherapy in breast cancer," *Acta Medica Okayama*, vol. 76, no. 2, pp. 105–111, 2022.

[25] K. Ravichandran, N. Braman, A. Janowczyk, and A. Madabhushi, "A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI," in *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis*, 2018. https://doi.org/10.1117/12.2294056

[26] Y. Qu, H. Zhu, K. Cao, X. Li, M. Ye, and Y. Sun, "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method," *Thoracic Cancer*, vol. 11, no. 3, pp. 651–658, 2020. https://doi.org/10.1111/1759-7714.13309

[27] M. Verma, L. Abdelrahman, F. Collado-Mesa, and M. Abdel-Mottaleb, "Multimodal spatiotemporal deep learning framework to predict response of breast cancer to neoadjuvant systemic therapy," *Diagnostics*, vol. 13, no. 13, p. 2251, 2023. https://doi.org/10.3390/diagnostics13132251

[28] H. Duanmu *et al.*, "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using deep learning with integrative imaging, molecular and demographic data," in *Medical Image Computing and Computer Assisted Intervention (MICCAI 2020),* in Lecture Notes in Computer Science, Springer, Cham, vol 12262, 2020, pp. 242–252. https://doi.org/10.1007/978-3-030-59713-9_24

[29] Z. Zhou *et al.*, "Prediction of pathologic complete response to neoadjuvant systemic therapy in triple negative breast cancer using deep learning on multiparametric MRI," *Scientific Reports*, vol. 13, no. 1, 2023. https://doi.org/10.1038/s41598-023-27518-2

[30] J. Xie *et al.*, "Dual-branch convolutional neural network based on ultrasound imaging in the early prediction of neoadjuvant chemotherapy response in patients with locally advanced breast cancer," *Frontiers in Oncology*, vol. 12, 2022. https://doi.org/10.3389/fonc.2022.812463

[31] H. Taleghamar, S. A. Jalalifar, G. J. Czarnota, and A. Sadeghi-Naini, "Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy," *Scientific Reports*, vol. 12, no. 2244, 2022. https://doi.org/10.1038/s41598-022-06100-2

[32] J. Gu *et al.*, "Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: A prospective study," *European Radiology*, vol. 32, no. 3, pp. 2099–2109, 2021. https://doi.org/10.1007/s00330-021-08293-y

[33] M. Jiang *et al.*, "Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer," *European Journal of Cancer*, vol. 147, pp. 95–105, 2021. https://doi.org/10.1016/j.ejca.2021.01.028

[34] C. F. J. Kuo, H. Y. Chen, J. Barman, K. H. Ko, and H. H. Hsu, "Complete, fully automatic detection and classification of benign and malignant breast tumors based on CT images using artificial intelligent and image processing," *Journal of Clinical Medicine*, vol. 12, no. 4, p. 1582, 2023. https://doi.org/10.3390/jcm12041582

[35] T. H. Qi *et al.*, "Multi-center evaluation of artificial intelligent imaging and clinical models for predicting neoadjuvant chemotherapy response in breast cancer," *Breast Cancer Research and Treatment*, vol. 193, no. 1, pp. 121–138, 2022. https://doi.org/10.1007/s10549-022-06521-7

[36] R. Massafra *et al.*, "Robustness evaluation of a deep learning model on sagittal and axial breast DCE-MRIs to predict pathological complete response to neoadjuvant chemotherapy," *Journal of Personalized Medicine*, vol. 12, no. 6, p. 953, 2022. https://doi.org/10.3390/jpm12060953

[37] J. H. Choi *et al.*, "Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning," *Scientific Reports*, vol. 10, no. 1, 2020. https://doi.org/10.1038/s41598-020-77875-5

[38] Y. Liu *et al.*, "Early prediction of treatment response to neoadjuvant chemotherapy based on longitudinal ultrasound images of HER2-positive breast cancer patients by Siamese multi-task network: A multicentre, retrospective cohort study," *eClinicalMedicine*, vol. 52, p. 101562, 2022. https://doi.org/10.1016/j.eclinm.2022.101562

[39] S. Joo *et al.*, "Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer," *Scientific Reports*, vol. 11, no. 1, 2021. https://doi.org/10.1038/s41598-021-98408-8

[40] R. Ha *et al.*, "Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 693–701, 2018. https://doi.org/10.1007/s10278-018-0144-1

[41] Y. Li *et al.*, "Deep learning radiomic analysis of DCE-MRI combined with clinical characteristics predicts pathological complete response to neoadjuvant chemotherapy in breast cancer," *Frontiers in Oncology*, vol. 12, 2023. https://doi.org/10.3389/fonc.2022.1041142

[42] L. Wu *et al.*, "An integrated deep learning model for the prediction of pathological complete response to neoadjuvant chemotherapy with serial ultrasonography in breast cancer patients: A multicentre, retrospective study," *Breast Cancer Research*, vol. 24, no. 1, 2022. https://doi.org/10.1186/s13058-022-01580-6

[43] Y. Peng *et al.*, "Pretreatment DCE-MRI-based deep learning outperforms radiomics analysis in predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer," *Frontiers in Oncology*, vol. 12, 2022. https://doi.org/10.3389/fonc.2022.846775

[44] H. Dammu, T. Ren, and T. Q. Duong, "Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients," *PLoS One*, vol. 18, no. 1, p. e0280148, 2023. https://doi.org/10.1371/journal.pone.0280148

## 6 AUTHORS

**Laila El Jiani** is a PhD student and a member of the Laboratory of Modeling and Information Technology. She has an extensive background in computer engineering and networks, holding a Master's in Big Data and Data Sciences from the Faculty of Sciences Ben M'Sik. With over 10 years of teaching experience in higher education, her research is focused on leveraging computer vision, deep learning, and machine learning to drive advancements in precision medicine and image processing (E-mail: Laila.eljiani-etu@etu.univh2c.ma).

**Sanaa El Filali** is a Professor at Hassan II University. She conducts research in several fields including IoT, artificial intelligence, machine learning, deep learning, NLP, computer vision, and cybersecurity. She leads the AI & Dev Community at the Faculty of Sciences Ben Msick and has been internationally recognized with awards such as the Data-Driven Research Award MENA 2021 and the first prize in virtual reality at the Diginov 2024 competition. Her academic contributions are significant and widely respected with over 30 publications and numerous doctoral theses supervised.

**El Habib Ben Lahmar** is a Professor at Hassan II University's Faculty of Sciences Ben M'Sick, where he leads the Data Science and Big Data Master's program and the Semantic Web research team. He earned his Ph.D. from ENSIAS in 2007, receiving the best thesis award in computer science. His research focuses on the semantic web, natural language processing, artificial intelligence, search engines, knowledge engineering, data science, and digital education. Dr. Ben Ahmar has published over 100 peer-reviewed papers and holds several patents in AI.

**Ihsane Haloum** is a PhD researcher at the Laboratory of Immunogenetics and Human Pathologies (LIGEP), Faculty of Medicine and Pharmacy of Casablanca, Hassan II University. Supported by a CNRST excellence fellowship, she focuses on immunogenetics, oncology, and artificial intelligence. Ihsane earned her B.Sc. in Biomedical Technologies in 2019 from the Faculty of Sciences and Techniques of Mohammedia, and her M.Sc. in Medical Biotechnologies in 2021 from the Faculty of Medicine and Pharmacy of Rabat (E-mail: ihsane.haloum@gmail.com).