College of Engineering, Design and Physical Sciences

Machine Learning Methods for AI-Enhanced DCE-MRI Breast Cancer Diagnosis

Marco Berchiolli

Acknowledgements

This work has been a rather tumultuous journey of self-discovery and self-improvement. I had the privilege of sharing this time with great people that have been generous and patient to bring me up to where I am today. First and foremost, I would like to express my deepest appreciation to my principal supervisor Professor Wamadeva Balachandran, for his outstanding support and humanity. I would also like to extend my deepest gratitude to the rest of the committee, Professor Tat-Hean Gan and Doctor Nenad Djordjevic, for their trust and valuable insight. I would like to express my sincerest gratitude to Dr Susann Wolfram and Dr Naveed Altaf for the medical knowledge they provided and for their support in dataset labelling. The project would not have been possible without the support from the Brunel Innovation Centre staff, both past and present, especially Maria, Mario, Sergio, Jamil, Ioanna, Evelyne, Ali, Makis, Anurag, and Anuj. They have made my stay in Cambridge an absolute pleasure and I am honoured to have been one of their colleagues.

I would not have been able to even enrol in this adventure without the unwavering and relentless support of Alessandra, who has endured my great lows with extraordinary grace, and has shown true happiness for my highs. I am especially grateful to my (large) family, particularly my parents, my brothers, sisters, my nephews, and my nieces. They always supported and nurtured me throughout my studies, and I hope they are proud of me as I am of them. Special thanks to the wondrous group of friends I grew up with: Domenico, Fabio, Gianluca, Gilberto, Scoiattolo, and Uberto. The heterogeneity in life experiences has allowed me to learn from different contexts and extract valuable life lessons. Thanks should also go to the newer friends, especially Dr Marco Zennaro, which was an endless source of helpful advice and profound belief in my abilities.

Finally, I'd like to recognize the assistance that I received from Maria Grazia and Angela, which have helped a clueless teenager to get into an UK university out of the kindness of their heart.

Abstract

Breast cancer remains one of the most prevalent and challenging diseases affecting women globally. Early and accurate diagnosis plays a pivotal role in improving patient outcomes and reducing mortality rates. This project aims to improve the current state of the art by employing several deep learning methodologies to be used as diagnostic support in Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI).

The work has been carried out as part of an InnovateUK project named "Intelliscan" (project reference: 104192), funded by UK Research and Innovation. UK Research and Innovation did not have any involvement in the study design, or the collection, analysis, and interpretation of data. The data collection was carried out in collaboration with consultant radiologist Dr Naveed Altaf (North Tees and Hartlepool NHS Foundation Trust) and Dr Susann Wolfram (Teesside University).

The research begins by providing an overview of neural networks, including structure, activation, regularization, training, architectures, loss functions, and performance metrics for model evaluation.

The first contribution to knowledge is provided as the diagnostic process is then presented from a clinician's point of view, along with empirical evidence of the subjectivity of the process prevents the application of ground truth data for the development of algorithms in this area.

The core contributions of this thesis lie in the development of several deep learning methodologies that are aimed at reducing human errors and increasing the speed of the diagnosis in clinical settings. These methodologies include: the first lesion detection algorithm based on unsupervised deep learning, which matches the performance of the current state of the art, while not relying on manually annotated data, significantly lowering the cost of research in the field; the development of a state of the art deep learning methodology to segment the organs within the chest wall from the breast; a novel application of deep learning in lesion morphology characterisation.

Overall, this thesis contributes to advancing the state-of-the-art in breast DCE-MRI diagnosis by introducing innovative deep learning methodologies that offer enhanced accuracy, efficiency, and interpretability.

Table of Contents

Table of abbreviations

ADAM	Adaptive Moment Estimation
ANN	Artificial Neural Network
BC	Bottleneck Connection
BEIT	Bidirectional Encoder representation from Image Transformers
CA	Contrast Agent
CAD	Computer Aided Diagnostics
CNN	Convolutional Neural Networks
CR	Clinical Radiology
CT	Computer Tomography
DCE-MRI	Dynamic Contrast Enhanced MRI
DL	Deep Learning
FCN	Fully Connected Network
FGT	Fibro-Glandular Tissue
GELU	Gaussian Error Linear Unit
JSC	Jaccard Similarity Coefficient
LTFL	Less Than Full Time
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NHS	National Health Services
NME	Non-Mass Enhancement
ReLU	Rectified Linear Unit
ResNet	Residual Network
ROI	Region Of Interest

SA	Sale-Attention layer
SELU	Scaled Exponential Linear Unit
SGD	Stochastic Gradient Descent
VQ-KD	Vector-Quantized Knowledge Distillation
WHO	World Health Organization
XCiT	Cross-Covariance Image Transformers

Table of figures

Chapter 1. Introduction

National Health Services (NHS) are experiencing delivery difficulties in almost any country in Europe, and not only, due to the continuous ageing of the population and the shortfall of personnel; this situation is in general affecting all fields of NHS, but for this work I will focus on medical imaging, central to the diagnosis and treatment of many medical conditions, cancer first.

The demand for imaging services such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is growing at a rate of 7% (see Figure 1.1.1) per year, whereas the Clinical Radiology (CR) workforce is growing at a lower rate, in the order of 4% ([1]); this mismatch is likely to continue for the years to come: workforce growth is evaluated to be 24% lower than the required growth.

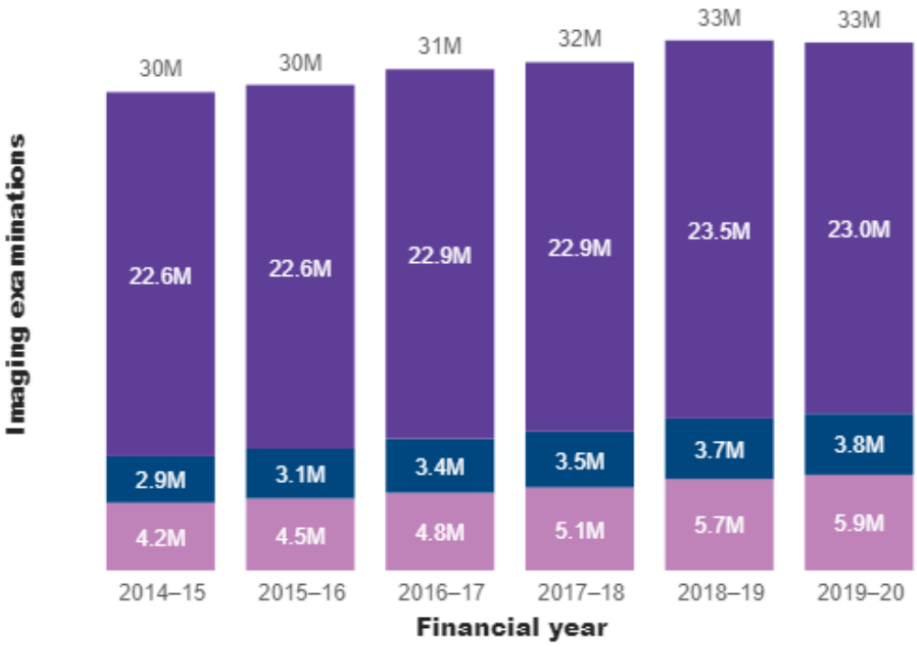


Figure 1.1: demand volumes for X-ray, CT, and MRI diagnosis [1].

Compared with the rest of Europe, the UK averages >30% fewer clinical radiologists per headcount (see Figure 1.2).

Figure 1.2: clinical radiologists per 100.000 population in UK countries, vs. European average [1].

In absolute values, the Royal College of Radiologists ([2]) estimates a shortfall of almost 2000 CR consultants across the UK, accounting for about 33% of the total needed workforce. Breast radiology is the second biggest radiology consultant specialist area but the first in terms of percentage shortfall, with 25% of the radiologists in service in the year 2020 expected to retire within 2025.

As of September 2020, well over 300.000 UK citizens were on a waiting list for CT or MRI, and one quarter had to wait for more than six weeks; this is unacceptable for a prompt cancer diagnosis. Such numbers were of course impacted by the COVID-19 pandemic, however, given the personnel shortage, recovery from such delays will be hard to achieve; nevertheless, reducing a lengthy waiting list is of paramount importance, since timely detection of serious medical conditions will positively impact the full NHS.

Let's then focus on breast radiology, the topic of this research project. According to the related report by The Royal College of Radiologists released in 2016 ([3]), the situation for breast radiologists is as follows:

- Breast radiologists comprise the larger subspeciality group of clinical radiologists.
- The breast radiologists' workforce accounts for a larger share of less than full-time (LTFT) consultants (35%) concerning general radiology (23%).
- Breast radiologists have a higher share of older professionals, with 42% being over 50 years old.
- The unfilled posts in breast radiology are doubling concerning general radiology.

These numbers suggest that, within an already critical situation for what concerns general radiology, the situation appears to be even worse for breast radiology, with a consistent number of consultants and professionals about to retire in the forthcoming years and a low substitution rate.

For this work, it is also interesting to consider the results, evidenced in the same report ([3]), of the breast screening activities carried out by the interviewed persons. All the reported charts are divided for professional figures in clinicians, consultant radiologists and practitioners. In Figure 1.3, Figure and Figure it can be seen that MRI is by far the least performed screening activity, despite its importance for high-risk patients.

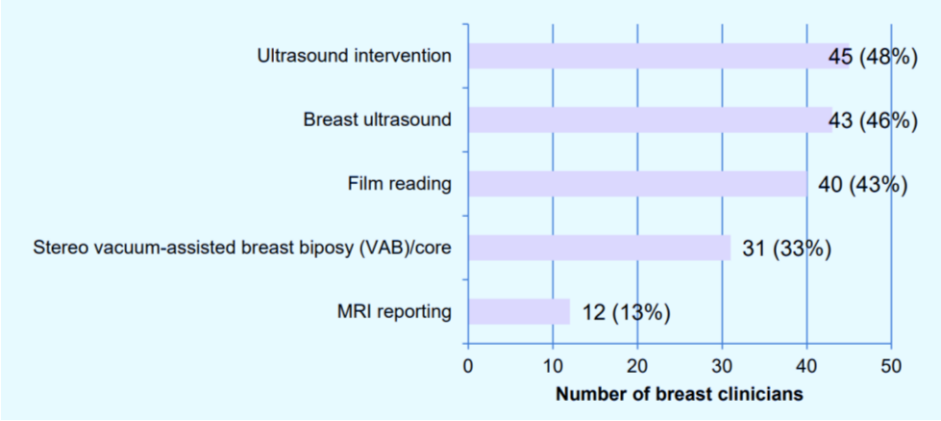


Figure 1.3: type of breast-screening activities performed by clinicians [3].

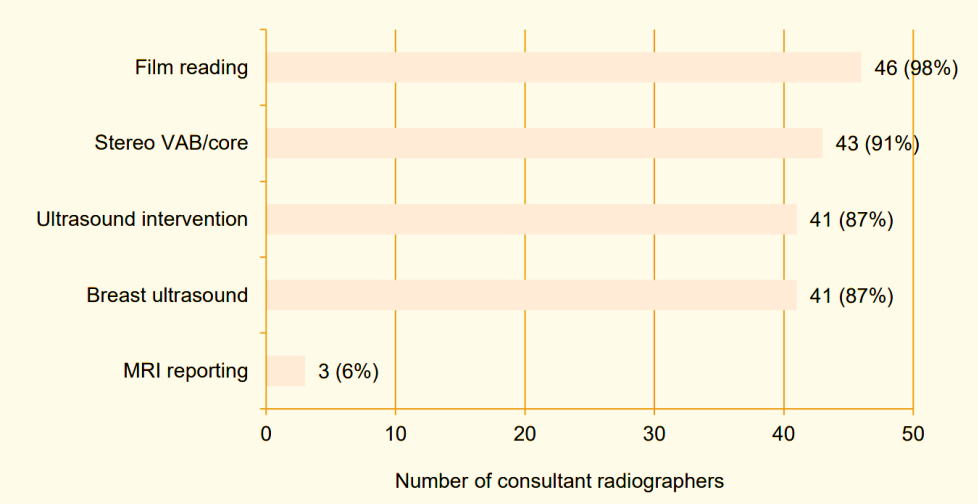


Figure 1.4: type of breast-screening activities performed by consultant radiologists [3].

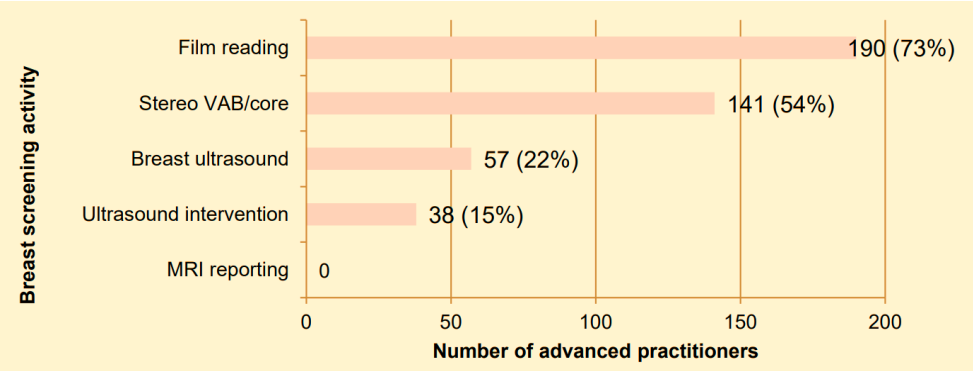


Figure 1.5: type of breast-screening activities performed by advanced practitioners [3].

According to the North Tees and Hartlepool NHS Foundation Trust, interpreting and reporting of breast MRI is time-consuming, requiring an average of 30 min. per patient. The introduction of automated detection and categorization of breast lesions could be of paramount importance for improving productivity, fostering the adoption of MRI techniques and help reducing waiting lists and cancer screening time, especially in a context in which breast radiology is experiencing a critical and constantly worsening shortfall of competencies.

1.1 Aim of the work

Breast cancer is the most frequent cancer among females, amounting to 24% of all cancer occurrences in 2018 ([4]), accounting for 684,996 deaths in 2020 worldwide ([5]). The World Health Organisation (WHO) indicates screening programmes aimed at early detection as one of the key factors in reducing mortality. Magnetic Resonance Imaging (MRI) is an increasingly popular procedure for the screening of high-risk groups ([7]) and evaluating the response to neo-adjuvant chemotherapy ([8]). MRI has several benefits over X-ray mammography; it does not utilise ionising radiation, generates high-resolution images and contains dynamic information. Moreover, recent developments in MRI image processing ([9]) indicate that MRI is gaining popularity even with its major drawbacks (time-consuming, stressful, and costly), and is actively being targeted by the research community. Dynamic Contrast-Enhanced MRI (DCE-MRI) is regarded as one of the main diagnostic tools for breast cancer. It outputs four-dimensional data (three spatial dimensions + one temporal dimension), consisting of images acquired before and after the intravenous injection of a contrast agent (CA). The change in tissue appearance in response to the CA is tissue-specific and, therefore, indicative of the presence of malignant breast lesions ([10]). These changes in tissue appearance are extremely like the ones of the internal organs such as the heart, making automatic lesion detection in breast DCE-MRI sequences challenging.

Manual delineation of the chest wall is an extremely time-consuming activity. Therefore, the automatic removal of the internal organs from the images by segmenting the chest cavity is instrumental to the development of an automatic lesion detection methodology. Figure 1.6 shows the results of an automatic lesion detection algorithm for a properly segmented image and a poorly segmented image of a breast DCE-MRI sequence. The algorithm is based on the statistical properties of the whole sequence and indicates areas with a high likelihood of containing a lesion, with red indicating the most likely candidate area. By not excluding the chest cavity, the system evaluates the heart as an area of high likelihood. As the chest cavity is segmented, the actual lesion is correctly highlighted.

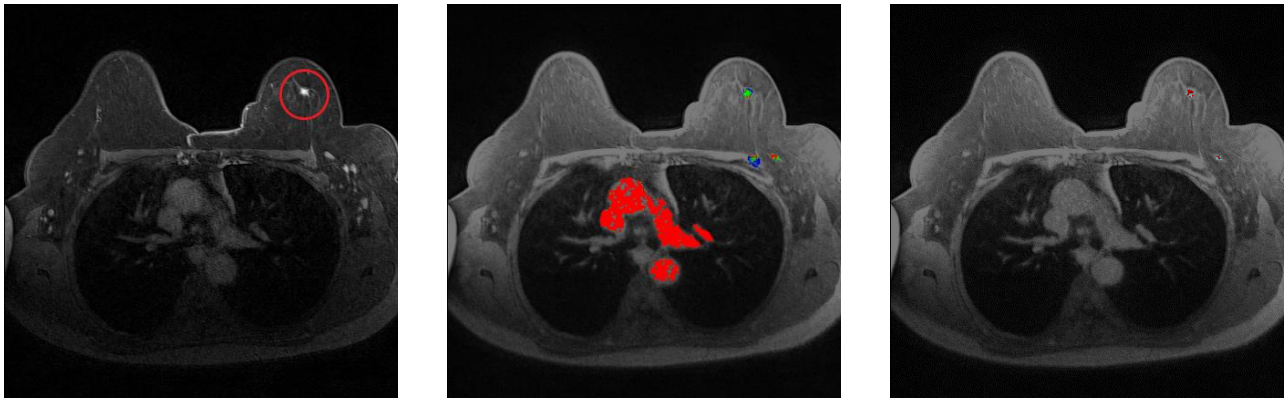


Figure 1.6: Results of an automatic lesion detection algorithm of a lesion. The original image with the lesion is visible on the left, with the lesion highlighted in the red circle. The colouring scheme of the other images (centre and right) is based on the statistical properties of the full breast volume. Red represents a high likelihood of a lesion, while green and blue represent a lower likelihood. In the centre image, the chest cavity is incorrectly segmented, and the lesion detection algorithm identifies the heart as an area of high likelihood to be a lesion. In the right image, the chest cavity is correctly segmented, and the area of high likelihood to contain a lesion is correctly identified.

Several methodologies have been developed to automatically localize suspect malignant breast lesions. Changes in tissue appearance in response to the injection of the contrast agent (CA) are indicative of the presence of malignant breast lesions. However, these changes are extremely like the ones of internal organs, such as the heart. Thus, the task of chest cavity segmentation is necessary for the development of lesion detection.

The clinical objective of this work is to enhance breast cancer diagnosis by providing AI-based tools that improve lesion detection and characterisation. Specifically, this methodology aims to automate the segmentation of the chest cavity in DCE-MRI sequences to exclude non-breast tissue, thus preventing false positives caused by the presence of internal organs (e.g., the heart). It also seeks to develop interpretable and trustworthy deep-learning algorithms to

detect and characterize lesions based on their statistical and dynamic properties. Furthermore, this work provides decision-support tools to radiologists for the assessment of breast lesions, aiding in distinguishing between benign and malignant findings. By improving detection accuracy, it aims to minimize unnecessary biopsies by accurately identifying lesions that are clearly benign, reducing patient anxiety, healthcare costs, and procedural risks.

The main purpose of the work described in the following chapters is the realization of a technique to support the diagnosis of breast cancer by choosing the best tools enabled by Artificial Intelligence, to compensate for the problems due at least partially to the shortfall of competencies and operators in the NHS. These tools include support for lesion detection in a trustworthy, interpretable way, followed by deep-learning-based lesion characterisation support.

1.2 Specific objectives

Given its aim to find the best possible solution for supporting the correct and timely detection of BC by supporting the analysis of DCE-MRI images, the main objectives of this work, which will be further detailed in the following chapters, are:

1. Identification of suspicious areas in DCE-MRI, using methodologies that would result in ease of integration in clinical practice. This objective was achieved, and the results are presented in chapter 4.
2. Minimisation of false positive results by the suspicious areas identification process to allow for increased productivity of the proposed solutions in clinical practice. This objective was achieved, and the results are presented in chapter 5.
3. Development of lesion characterization methodologies for faster and more accurate morphology classification of suspect lesions in clinical settings. This objective was achieved, and the results are presented in chapter 6.

1.3 Summary of the Data

The data collected is comprised of 113 DCE-MRI scans, acquired on a 1.5T scanner (MAGNETOM Avanto, Siemens Healthcare GmbH, Erlangen, Germany) with the patient positioned lying face down. TR/TE/flip angle was 4.33s/1.32s/10° for each scan with a slice thickness of 1.1 mm with no gaps between slices. The resolution of each slice was 448 x 448 pixels. Each breast DCE-MRI protocol consisted of one pre-contrast T1-weighted sequence and

seven post-contrast T1-weighted sequences collected at intervals of 1:01 minutes between sequences. All data points were attached to a biopsy report, providing a ground truth on the correctness of the diagnosis.

An additional dataset, void of biopsy confirmation, was utilised in the classification portion of this work (chapter 6). In this, 86 scans were provided, of which 24 presented a benign lesion (no biopsy was performed), and the remaining 62 did not feature a lesion at all. The work was carried out solely on the 24 cases with benign lesions, with the sole purpose of decreasing the severe imbalance in the dataset.

1.4 Summary of methodology

The first objective was met through an unsupervised methodology for identifying suspect regions in DCE-MRI scans by treating them as outliers. The approach involved encoding the temporal intensity of a single pixel into a low-dimensional space, using neural networks, and utilizing algorithmic heuristics to determine the likelihood of a pixel being part of a cancerous region. The second objective is addressed through a deep learning segmentation model. The approach consisted of supervised training of state-of-the-art models to identify the inner portion of the chest wall. The third and final objective was addressed by training bleeding-edge deep learning algorithms for lesion classification on a labelled dataset based on biopsy results.

1.5 Organisation of the document

The document is organized as follows:

- **Chapter 2** will review the state of the art and will describe the related work that has been used, evaluated, and improved during the research activities for finding the best solution.
- **Chapter 3** will describe the material that has been used for the experimental activities, and will describe the structure of the datasets, the labelling approaches, etc., and will include recommendations about data structuring, data labelling and data standardization to replicate the workflow adopted during this project.
- **Chapter 4** will describe the solution as a first-ever truly explainable AI approach to breast lesion detection in DCE-MRI.
- **Chapter 5** will describe in detail the approach followed for the chest cavity segmentation, which is the core of the work. The complete research methodology is duly described, and experimental evidence of the

effectiveness of the technique is reported, therefore granting consistent progress concerning the state of the art.

- **Chapter 6** will describe the method that was implemented for lesion classification based on their characterization and will present an algorithm for pseudo-interpretation of the results.
- **Chapter 7** will derive conclusions and propose steps for future work based on the outcomes of this research project.

1.6 Contribution to New Knowledge

The main outcomes of the project are:

1. A novel, interpretable approach to automatic lesion detection and kinetic behaviour characterisation. The technique aims to address the growing concerns over the trustworthiness of artificial intelligence systems while providing practitioners with diagnostic tools for enhanced and safer diagnosis. The four-dimensional data from dynamic contrast-enhanced MRI scans of breasts were analysed using matrix decomposition techniques based on principal component pursuit to extract the transient behaviour of tissues and highlight potential lesions. Several signal-processing techniques were there applied to limit the number of false positives obtained from the main algorithm and to create composite images that would be easy to interpret. The output was aligned with the standard diagnostic process to allow fast and reliable validation of the resulting images.
2. A novel, DL-based methodology for the segmentation of the chest cavity from breast DCE-MRI scans. The solution addressed the main challenges in chest cavity segmentation, especially sternum detection, by using a data-based approach. To showcase the potential of the solution, the target area to segment was selected to be the upper half of the chest cavity. The purpose of the work was to define a data-efficient approach, to automatically segment breast MRI data. Specifically, a study on several UNet-like architectures (Dynamic UNet) based on ResNet has been extensively performed and is presented in detail in the following sections. Experiments quantify the impact of several additions to baseline models of varying depth, such as self-attention and the presence of a bottlenecked connection. The proposed methodology is demonstrated to outperform the current state of the art both in terms of data efficiency, as well as in terms of similarity index when compared to manually segmented data.

3. A novel, DL-based methodology to classify the morphology of the suspicious regions highlighted by the previous outcomes. The work proves the feasibility of utilising an automatic system as support to practitioners in the challenging phase of determining the regularity (or lack thereof) of a lesion, which can guide to appropriate treatment, reducing the number of biopsies that are carried out, hence greatly benefitting patients. The main technique uses several state-of-the-art deep learning classifiers based on ResNet to predict the irregularity of a lesion. Moreover, a technique to interpret the results is implemented, leading to allowing access to practitioners in the inner workings of the algorithm and providing them with an immediate and intuitive tool to gauge the accuracy of the prediction. The methodology uses the hidden activations of the neural network to construct a heatmap to be overlayed on top of the input image, thus visualising the attention map of the algorithm.

1.7 List of publications

Part of the work reported in this thesis resulted in the following publication:

- Berchiolli, M.; Wolfram, S.; Balachandran, W.; Gan, T.-H. Fully Automatic Thoracic Cavity Segmentation in Dynamic Contrast-Enhanced Breast MRI Using Deep Convolutional Neural Networks. *Appl. Sci.* **2023**, *13*, 10160. https://doi.org/10.3390/app131810160

Other publications are in the process of being prepared for submission to the same journal.

Chapter 2. Background

2.1 Deep Learning in DCE-MRI Breast Cancer Diagnosis

Recent advancements in deep learning for breast cancer detection using magnetic resonance imaging are making progress, though challenges remain due to limited datasets and suboptimal image quality. Unlike mammography, only a few public datasets for ultrasound and MRI have been released, often with small sample sizes. Notable datasets for DCE-MRI include Duke Breast Cancer [111] and BreastDM [112].

A review by Adam et al. in July 2023 highlighted the use of Convolutional Neural Networks (CNNs) for tasks like classification, detection, and segmentation in breast cancer diagnosis via MRI [113]. Although small studies show promising results, larger, well-designed studies are still lacking to assess deep learning performance in real-world clinical settings.

Recent studies have focused on automated segmentation of breast cancer using DCE-MRI, which offers precise lesion segmentation for staging, treatment planning, and response evaluation. For example, a 2023 study by Janse et al. trained an nnU-Net segmentation pipeline for local breast cancer and showed significant correlation between automated volumetric measurements and functional tumour volume [114].

Another emerging area is the generation of synthetic post-contrast MRI images using Generative Adversarial Networks (GANs). This technology, still in its early stages, could help improve breast cancer staging and treatment evaluation, especially for patients unable to receive intravenous contrast. Studies by Chung et al. [115] and Osuala et al. [116] explored the feasibility and accuracy of generating synthetic post-contrast MRI images, demonstrating potential but also revealing challenges in segmentation performance.

Finally, deep learning is being applied to radiomics for predicting treatment responses, such as neoadjuvant chemotherapy. A 2023 study by Li et al. [117] developed a deep learning-based radiomic model for predicting pathological complete response (pCR) to chemotherapy, outperforming traditional radiomic methods and showcasing its potential to improve treatment prediction accuracy.

2.2 Introduction to Neural Networks

In recent years, the field of artificial intelligence has witnessed remarkable progress, primarily propelled by advancements in neural networks. Neural networks, often referred to as artificial neural networks (ANNs), have

revolutionized various applications, such as computer vision, natural language processing, speech recognition, and more. These systems mimic the neural structure of the human brain, enabling them to learn from data and make accurate predictions. At its core, a neural network is a collection of interconnected computational units, or nodes, organized into layers. The network typically consists of an input layer, one or more hidden layers, and an output layer. Each node in a layer is connected to every node in the subsequent layer, forming a network of weighted connections. The architecture's depth (number of layers) and width (number of nodes in each layer) can vary based on the complexity of the problem at hand. The mathematical representation of a neural network's output can be formulated as follows:

Let x be the input to the neural network, W_i be the weight matrix of the connections between layer i and layer i+1, and b_i be the bias vector of layer i+1. The output of the neural network, denoted as $\hat{y}$, can be computed as:

$$\hat{y} = f\big(W_n * f\big(W_{n-1\}} * \dots * f(W_2 * f(W_1 * x + b_1) + b_2) + \dots\big) + b_n\big) \quad (2.1)$$

Here, f represents the activation function, which introduces non-linearity to the network and is essential for learning complex patterns.

2.3 Activations

Activation functions play a pivotal role in introducing non-linear transformations to the neural network, allowing it to model intricate relationships within the data ([11]). Different types of activation functions exist, and each of them brings unique properties to the neural network, impacting its learning speed, convergence, and ability to handle vanishing or exploding gradients. Examples of classic activation functions that have been widely used are, for example, the Sigmoid ([12]), ReLU ([13]), and tanh ([14]); more recent functions, such as GELU ([15]), Leaky ReLU ([16]), SELU ([17]), have demonstrated promising results.

More in detail:

Sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

Rectified Linear Unit (ReLU):

$$f(x) = max(0, x) \tag{2.3}$$

H**yperbolic tangent (tanh):**

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.4}$$

Gaussian Error Linear Unit (GELU):

$$f(x) = 0.5x\left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.044715 * x^{3})\right)\right) \tag{2.5}$$

The GELU function introduces smoothness, which aids in gradient propagation during training. Its close approximation to the identity function for positive values of x makes it an attractive choice in various deep-learning architectures, particularly in transformer-based models.

Leaky ReLU:

$$f(x) = \begin{cases} x, if\ x \geq 0 \\ \alpha * x, if\ x < 0 \end{cases} \tag{2.6}$$

where α is a hyperparameter (typically set to a small positive value, such as 0.01). The inclusion of α allows Leaky ReLU to prevent neurons from becoming inactive during training, mitigating the "dying ReLU" problem and improving the flow of gradients.

Leaky ReLU strikes a balance between the linearity of ReLU, and the saturation issue faced by the Sigmoid and Tanh functions, making it a popular choice in deep neural networks, especially in scenarios where the rectification of negative values is essential.

Scaled Exponential Linear Unit (SELU):

$$f(x) = \begin{cases} \lambda * (\alpha * e^x - \alpha), if\ x < 0 \\ \lambda * x, if\ x \geq 0 \end{cases} \quad (2.7)$$

where α and λ are hyperparameters. The SELU function exhibits a unique property of maintaining the mean activation close to zero and the standard deviation close to one across layers during training. This self-normalization effect helps mitigate the vanishing and exploding gradient problems often encountered in deep neural networks, contributing to more stable and efficient learning.

SELU is particularly well-suited for deep architectures, as its self-normalizing property reduces the need for extensive parameter tuning and enables better convergence even without batch normalization techniques. However, it is essential to ensure the proper initialization of weights and biases to fully exploit the benefits of SELU.

In this work, the choice of activation functions was made upon examination of ample empirical evidence ([18]). The selection of Leaky ReLU with $\alpha = 0.01$ for convolutional architectures and GELU for self-attention-based networks was justified by their respective abilities to address architectural challenges and improve the performance of the models in their specific domains. These choices have been widely adopted and validated, supporting their effectiveness in deep learning applications.

2.4 Regularisation techniques

Regularization is a crucial technique in deep learning that helps prevent overfitting and improve the generalization performance of neural networks. Overfitting occurs when a model performs exceptionally well on the training data but fails to generalize to unseen data, resulting in poor performance on test or validation sets. Regularization methods add constraints to the model during training, discouraging it from becoming too complex and fitting noise in the training data. This regularization encourages the model to learn more robust and generalizable patterns from the data. The combination of several techniques is vital to a successful application of deep learning. Examples of possible techniques to be combined are:

L1 Regularization:

L1 regularization ([19]) adds a penalty term to the loss function proportional to the absolute value of the model's weights. It encourages sparsity in the model, forcing some of the weights to be exactly zero. Loss with L1 regularization term is defined as follows:

$$L_{L1} = L_0 + \lambda * \sum_{i=1}^{n} |w_i| \quad (2.8)$$

Here, λ is the regularization parameter that controls the strength of the L1 penalty, and w represents the model's weights at each layer *i*.

Weight Decay (L2 Regularization):

Weight decay ([20, 21]), or L2 regularization, is a regularization technique that penalizes the model's weights by adding a term proportional to the sum of their squared values to the loss function. This penalty term encourages the model to prefer smaller weight values, leading to a smoother and less complex decision boundary. Loss with L2 regularization term (for weight decay) is defined as follows:

$$L_{wd} = L_0 + \lambda * \sqrt{\sum_{i=1}^{n} |w_i|^2} \quad (2.9)$$

In this formula, λ is the weight decay coefficient, controlling the strength of the regularization, and w represents the model's weights at each layer *i*..

Data Augmentation:

Data augmentation ([22]) is a technique used to artificially increase the size of the training dataset by applying various transformations to the original data. These transformations do not change the underlying label or meaning of the data but create new variations that the model can learn from. Data augmentation helps the model generalize better by exposing it to diverse examples of the same class, reducing the risk of overfitting.

Common data augmentation techniques include rotation, flipping, scaling, cropping, and colour jittering for image data. For sequential or time-series data, augmentation methods may involve shifting, scaling, or adding noise to the input sequences. Data augmentation is typically performed on the fly during training, generating augmented samples from the original data on each epoch or batch. The augmented samples are then used as inputs for training the neural network.

Batch Normalization:

Batch normalization ([23]) normalizes the activations of a layer to have zero mean and unit variance, which helps stabilize training and reduces internal covariate shifts. It acts as a form of regularization and can accelerate convergence. The batch normalization formula is:

$$z = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{2.10}$$

$$y = w * z + b \tag{2.11}$$

where x is the input to the layer, μ and σ are the mean and variance of the mini-batch, ε is a small constant for numerical stability, and *w* and *b* are learnable parameters.

Dropout:

Dropout ([24]) is a popular regularization technique that randomly sets a fraction of neurons to zero during training, effectively dropping them out of the network for that iteration. This prevents neurons from relying too much on specific features, forcing them to learn more robust representations. The dropout formula (applied to the output of a neuron during training *y*) is:

$$y_{dropout} = y * Bernoulli(p) \tag{2.12}$$

where p is the dropout probability, representing the probability of dropping out each neuron, and Bernoulli refers to a Bernoulli distribution:

$$f(k; p) = pk + (1 - p)(1 - k) \quad for\ k \in \{0; 1\} \tag{2.13}$$

2.5 Training Neural Networks

The essence of neural networks lies in their ability to learn from data. The learning process involves updating the network's weights and biases iteratively, such that the model's predictions progressively improve. This optimization is typically achieved using algorithms based on backpropagation ([25]), where the gradients of the network's error concerning its parameters are computed and used to adjust the weights accordingly.

Mathematically, the weight update rule during backpropagation can be expressed as:

$$w_{i+1} = w_i - lr * \frac{\nabla(loss)}{\nabla(w_i)} \tag{2.14}$$

Where lr is the hyperparameter controlling the step size during optimization, and $\frac{\nabla(loss)}{\nabla(W_i)}$ represents the gradient of the loss function concerning the weights *w*.

2.5.1 Training Algorithm

The goal of training is to find the optimal set of parameters that minimize the difference between the model's predictions and the actual target values in the training data.

The training algorithm plays a central role in adjusting the model's parameters to optimize the loss function. The loss function measures the discrepancy between the predicted output of the model and the true target values. By minimizing this loss function, the model learns to make accurate predictions on unseen data, a process known as generalization.

The training algorithm aims to find the optimal parameters by iteratively updating them based on the gradients of the loss function concerning each parameter. These gradients indicate the direction in which the parameters should be adjusted to reduce the loss and improve the model's performance.

Training a deep learning model typically involves feeding the training data through the model, computing the loss, computing gradients through backpropagation, and then updating the model's parameters using the chosen training algorithm. This process is repeated for multiple iterations (epochs) until the model converges to a state where the loss is minimized, and the model generalizes well to unseen data.

Below are brief explanations of some of the popular learning algorithms used in deep learning.

Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent ([26]) is a fundamental optimization algorithm widely used in training neural networks. It updates the model's parameters based on the gradient of the loss function concerning each parameter. SGD computes the gradient using a randomly selected data point (or a mini-batch of data points) in each iteration, making it computationally efficient for large datasets. While SGD can converge to a minimum, it often exhibits noisy convergence and may suffer from slow convergence or getting stuck in local minima.

Adagrad (Adaptive Gradient Algorithm)

Adagrad ([27]) is an adaptive learning rate optimization algorithm that adjusts the learning rate for each parameter based on the historical gradients of that parameter. It gives more weight to parameters with infrequent updates and less weight to frequently updated parameters, making it effective for training models with sparse features.

The update rule for Adagrad is:

$$G_t = G_{t-1} + \left(\nabla(\theta_t)\right)^2 \tag{2.15}$$

$$\theta_{t+1} = \theta_t - \alpha * \frac{\nabla(\theta_t)}{\sqrt{G_t} + \varepsilon} \tag{2.16}$$

where:

- α is the learning rate.
- θ_t represents the parameters at time step t.
- $\nabla(\theta_t)$ is the gradient of the loss function concerning parameters θ_t.
- G_t is the sum of the squared gradients up to time step t.
- ε is a small constant to prevent division by zero.

RMSprop (Root Mean Square Propagation):

RMSprop adapts the learning rate on a per-parameter basis, allowing it to converge faster and more reliably than standard SGD, especially when dealing with sparse or noisy data. The exponential moving average of squared gradients helps to normalize the learning rate, reducing its sensitivity to different scales of gradients for each parameter. This property makes RMSprop particularly effective in training deep neural networks.

The update rule for RMSprop is:

$$G_t = \beta * G_{t-1} + (1 - \beta) * \left(\nabla(\theta_t)\right)^2 \tag{2.17}$$

$$\theta_{t+1} = \theta_t - \alpha * \frac{\nabla(\theta_t)}{\sqrt{G_t} + \varepsilon} \tag{2.18}$$

where:

- α (alpha) is the learning rate.

- θ_t represents the parameters at time step t.
- $\nabla(\theta_t)$ is the gradient of the loss function with respect to the parameters θ_t.
- G_t is the moving average of squared gradients up to time step t.
- β is a decay factor, typically set to 0.9 to 0.999, controlling the exponential decay of the moving average.
- ε is a small constant to prevent division by zero.

Adam (Adaptive Moment Estimation)

Adam ([28]) is an adaptive learning rate optimization algorithm that combines the benefits of both the momentum-based optimization methods (RMSProp), and adaptive learning rate techniques. It maintains an exponentially decaying average of past squared gradients (second moments) and past gradients (first moments) to adaptively adjust the learning rate for each parameter. Adam's adaptive learning rates allow it to perform well across various types of neural network architectures. The update rule for Adam is:

$$m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * \nabla(\theta_t) \tag{2.19}$$

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * \nabla(\theta_t)^2 \tag{2.20}$$

$$\theta_{t+1} = \theta_t - \alpha * \frac{m_t}{\sqrt{v_t} + \varepsilon} \tag{2.21}$$

where:

- α (alpha) is the learning rate.
- θ_t represents the parameters at time step t.
- $\nabla(\theta_t)$ is the gradient of the loss function with respect to the parameters θ_t.
- m_t and v_t are the first and second moment estimates, respectively.
- β_1 and β_2 are hyperparameters controlling the exponential decay rates of the moment estimates.
- ε is a small constant to avoid division by zero.

2.5.2 Loss Functions

Loss functions, also known as objective functions or cost functions, are mathematical measures that quantify the difference between the predicted values generated by a machine learning or deep learning model and the actual target

values or ground truth. These functions play a crucial role in training models by providing a way to assess how well the model is performing on a given task. The primary purpose of a loss function is to guide the optimization process during training by providing a signal that helps adjust the model's parameters to improve its performance.

Loss functions in the context of deep learning emerge as a response to the diverse requirements of quantifying errors across various tasks. These tasks span regression, classification, generative modelling, and beyond, each necessitating a distinct approach to measuring and managing errors. The intricacies of these tasks demand specialized loss functions that cater to their unique characteristics. In detail:

Regression

In regression tasks, the objective is to predict continuous numerical values. The discrepancy between predicted and actual values is quantified using loss functions that are sensitive to the magnitude of errors. Mean Squared Error (MSE) is a prime example, capturing the average squared difference between predictions and actual values. Its focus on the magnitude of errors suits regression tasks where accurate estimation of numeric values is crucial.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2 \tag{2.22}$$

Where ŷ is the predicted value, and y is the actual value.

Classification

Classification tasks entail assigning data points to predefined classes. Since the predicted outputs are categorical, specialized loss functions are needed. Cross-Entropy Loss serves as a natural choice here. It quantifies the divergence between predicted class probabilities and actual class probabilities, effectively penalizing incorrect predictions while encouraging high probabilities for the true class.

$$CE = -\sum_{i=1}^{N} q_i \log(p_i) \tag{2.23}$$

Where p represents the predicted class probabilities and q represents the actual class probabilities. In classification problems, the probability distribution of a single prediction is obtained by applying SoftMax to the last activation of the neural network.

The SoftMax function ensures that the values in the resulting probability distribution are between 0 and 1, and they sum up to 1, making it suitable for tasks where you want the model to assign a probability to each class.

Given a vector of $z = [z_1, z_2, \dots, z_n]$, where each z_i corresponds to the raw score for class i, the softmax function transforms these scores into probabilities $p = [p_1, p_2, \dots, p_n]$:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{2.24}$$

Here, e^{z_i} represents the exponential of the $i - th$ logit, and the denominator is the sum of exponentials of all logits in the vector. This denominator ensures that the resulting probabilities sum up to 1.

Image Segmentation

In image segmentation tasks, where pixel-wise categorization is vital, Dice Loss gains prominence. It quantifies the overlap between predicted and ground truth segmentation masks, favouring accurate segmentation and penalizing misalignments.

$$DICE = 1 - \frac{2\sum_{i=1}^{n} y_i \widehat{y_i}}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \widehat{y_i}} \tag{2.25}$$

Where ŷ is the predicted value, and y is the actual value.

2.6 Fully Connected Neural Networks

Fully Connected Networks (FCNs), also known as Dense Networks or Multi-Layer Perceptron (MLPs), stand as a cornerstone architecture within the annals of artificial neural networks. They are the naïve approach to neural networks, as they manifest as layered arrangements of neurons, wherein each neuron within a given layer establishes connections to every neuron in the subsequent layer. This dense interconnectedness empowers FCNs to sculpt intricate nonlinear mappings, rendering them proficient in tasks necessitating a holistic comprehension of input data.

Mathematically, the output y of a neuron in an FCN can be expressed as:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{2.26}$$

Where y is the neuron's output; f is the activation function, introducing nonlinearity; w_i are the weights connecting the neuron to its inputs x_i; b is the bias term, a learnable parameter that is added to each operation in the layer; n is the number of inputs to the neuron.

The standard FCN architecture encompasses an input layer, one or more hidden layers, and an output layer. Neurons within hidden layers incorporate activation functions, thereby infusing nonlinearity and endowing the network with the prowess to apprehend intricate input-output relationships. The process of backpropagation, harmonized with optimization algorithms, orchestrates the fine-tuning of network weights and biases throughout training, aligning predictive outcomes with verifiable labels.

However, despite their aptitude in capturing comprehensive relationships, FCNs are susceptible to overfitting in the presence of high-dimensional data. To counter this propensity, regularization techniques, including dropout and L2 regularization, are frequently enlisted to mollify overfitting's adverse effects. Additionally, FCNs' inherent lack of spatial hierarchies might limit their efficacy in tasks mandating localized feature disentanglement, exemplified by image analysis.

2.7 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) ([29]) are a specialized class of neural networks designed to excel at processing and analysing data with grid-like structures, such as images, video frames, and even sequential data. Unlike fully connected neural networks (also known as multi-layer perceptions), which treat each input feature independently, CNNs harness the power of spatial hierarchies and local correlations present in the data.

At the core of CNNs lies the convolutional operation, a cornerstone for their success. This operation applies learnable filters to local patches of input data, yielding feature maps that capture salient patterns. Mathematically, the convolution operation is expressed as:

$$S(i,j) = (I * K)(i,j) = \sum_{m}^{M} \sum_{n}^{N} I(i-m, j-n) \cdot K(m,n) \tag{2.27}$$

Here, I signifies the input data, K represents the filter kernel (of shape MxN), and S is the resulting feature map. The indices i and j refer to spatial positions within the feature map, while m and n denote positions within the kernel. As the operation convolves the kernel over the input data, it captures localized information, enabling the network to discern meaningful features. A fundamental paper in understanding how CNNs work has been [30], in which the increasing abstraction levels are show.

Furthermore, CNNs employ pooling layers for down sampling and non-linear activation functions to enhance abstraction. By stacking convolutional layers and pooling layers, the architecture can extract relevant information by building up increasingly complex representations of the data.

This architecture facilitates the construction of hierarchical representations, mirroring the hierarchical structures present in real-world data. Comparatively, while FCNs excel at capturing global relationships in data, CNNs excel at localized pattern recognition. In essence, this leads to CNNs being considerably better in unstructured data (i.e. images, text) than FCNs ([31]). In images, this fact is intuitively visualised: an FCN could not learn the relative importance of a pixel as the location of relevant information varies at each data point, leading to the necessity to learn patterns that would in turn provide information on the relative importance of the pixels. A CNN, in contrast, would not need a relevancy map, as the features would be extracted in a bottom-up approach.

2.8 Residual Networks

Residual Networks, often abbreviated as ResNets, represent a pivotal advancement in deep learning architecture that addresses the challenges of training very deep neural networks. Developed by Kaiming He et Al. in 2015 ([32]), ResNets introduced a groundbreaking concept that revolutionized the way deep networks are constructed and optimized.

The central innovation of ResNets lies in the introduction of residual blocks, which facilitate the training of extremely deep networks by alleviating the vanishing gradient problem. This problem arises when gradients become infinitesimally small as they backpropagate through many layers, hampering the training process. Residual blocks introduce skip connections, also known as shortcut connections, that allow the network to directly propagate information from one layer to a later layer, bypassing intermediate layers. This short-circuiting enables the network to learn the identity mapping more effectively when needed, providing a reference point for the network to optimize deviations from this identity.

Mathematically, a residual block can be represented as:

$$y = I + f(I) \tag{2.28}$$

Where I represents the input to the layer and $f(I)$ is the residual function, learned by the block, which represents the deviation to an identity mapping. This approach ensures that even if the learned function is close to zero, the network can still propagate gradients through the skip connection, mitigating the vanishing gradient issue. ResNets come in

various depths, from shallow to extremely deep architectures, containing numerous residual blocks. The skip connections are commonly implemented as convolutional layers with zero padding to match the dimensions of the input and output.

With the pivotal introduction of ResNets, it is possible to view deep neural networks as input modification machines, instead of feature extractors. In other words, a CNN would transform the input data into a high-dimensional representation of the information contained within the input, while ResNets merely enhance the features that are already present in the data. While in practice this difference is minute, as demonstrated by Hao Li et al. ([33]), the effect on stabilisation during training is massive.

2.9 Self-attention Networks

Self-attention, a transformative mechanism originating from the natural language processing domain ([34]), has been seamlessly integrated into the field of computer vision, reshaping how models perceive and analyse images. In the context of computer vision, self-attention provides a mechanism for capturing contextual relationships between different regions of an image, enabling models to weigh and emphasize pertinent information dynamically.

At its core, self-attention operates by computing attention scores for each position in an input feature map, signifying the relevance of that position with respect to others. The attention scores are subsequently used to create weighted combinations of feature vectors from all positions, fostering a refined representation that explicitly considers the interdependencies within the input data.

Mathematically, the self-attention mechanism involves three linear transformations: query (Q), key (K), and value (V) matrices. These matrices are multiplied together, yielding attention scores that are applied to the values to obtain the final attention-weighted representation. This process can be summarized as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \tag{2.29}$$

Where Q represents query vectors, K representes key vectors, V represents value vectors and d_k is the dimension of the key vectors. In the realm of computer vision, self-attention has proven particularly effective for tasks involving long-range dependencies, context understanding, and object relationships within images. Unlike convolutional operations that apply fixed filters, self-attention dynamically adapts its weights based on the relationships present in the input, thus capturing intricate patterns and relationships regardless of their spatial separation. Transformative

models such as the Transformer and its visual counterpart, the Vision Transformer (ViT) ([35]), leverage self-attention to enable image recognition without the need for traditional convolutional layers. By treating images as sequences of patches, these models apply self-attention to capture both local and global relationships, leading to state-of-the-art performance on various computer vision benchmarks.

2.10 Losses

In the context of deep learning, the training and validation loss serve as fundamental metrics for assessing the performance and convergence of a neural network model during the training process. The loss quantifies the discrepancy between the model's predictions and the actual target values. Mathematically, the training and validation loss can be expressed as follows:

$$L_{train} = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} \mathcal{L}(\hat{y}, y) \tag{2.30}$$

$$L_{valid} = \frac{1}{N_{valid}} \sum_{i=1}^{N_{valid}} \mathcal{L}(\hat{y}, y) \tag{2.31}$$

Where N is the number of samples in each dataset; $\mathcal{L}$ represents the error, based on the loss function of choice; $\hat{y}$ represents the predicted value of the model. During the training process, the model's weights and biases are iteratively adjusted to minimize the training loss (L_{train}). The validation loss (L_{valid}) is used to monitor the model's generalization performance and detect overfitting. Ideally, the training and validation loss should decrease in tandem during training, and if the validation loss starts increasing, it may signal overfitting.

By analysing the training and validation loss, it is possible gain insights into the model's learning progress, convergence, and potential issues. This information guides the decision-making process regarding hyperparameter tuning, model architecture adjustments, and early stopping strategies.

2.11 Performance Metrics

Evaluating the performance of deep learning models is essential to assess their effectiveness in solving specific tasks. A plethora of performance metrics are employed, depending on the task at hand. These metrics offer valuable insights into various aspects of model behaviour, aiding in model selection, optimization, and deployment.

In the realm of classification tasks, evaluating model performance involves various metrics that offer distinct insights into the effectiveness of predictions. In this work, the following metrics were used to evaluate the effectiveness of classifiers:

Accuracy:

Accuracy quantifies the proportion of correct predictions made by a model over the total predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.32}$$

Recall:

Recall (also known as Sensitivity or True Positive Rate – TPR) measures the ratio of true positive predictions (correctly identified positive instances) to the total actual positive instances.

$$Recall = \frac{TP}{TP + FN} \tag{2.33}$$

Precision:

Precision gauges the ratio of true positive predictions to the total positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \tag{2.34}$$

F1 Score:

The F1 score is the harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives.

$$F1 = \frac{2 \times Precision \times Recall}{Precison + Recall} = \frac{2TP}{2TP + FP + FN} \tag{2.35}$$

Where TP (True Positives) are the instances correctly classified as positive, FP (False Positives) are the instances classified as positive which are negative, FN (False Negatives) are the instances classified as negative which are positive, and TN are the instances correctly classified as negative.

These metrics are particularly valuable in diverse classification scenarios. Accuracy is a straightforward metric, but it might be misleading in cases of imbalanced classes, such as the one of this work. Recall emphasizes the ability to correctly identify all relevant instances, critical in scenarios where false negatives are undesirable, such as medical this one. Precision highlights the accuracy of positive predictions, essential when false positives have significant

consequences. The F1 score combines both precision and recall, providing a more comprehensive evaluation of model performance, especially when balancing trade-offs between precision and recall is important.

2.12 Summary of Chapter

This chapter provided an extensive overview of neural networks, covering their architecture, activation functions, regularization techniques, training algorithms, various types of networks, loss functions, and performance metrics. Overall, the text serves as a comprehensive introduction to neural networks, catering to readers interested in understanding the underlying principles, architectures, and applications of this rapidly evolving field in artificial intelligence.

Chapter 3. Materials

3.1.1 Data Collection and characteristics

The data for this project was sourced from T1-Weighted Dynamic Contrast Enhanced MRI (DCE-MRI) scans of female patients of North Tees and Hartlepool NHS Foundation Trust. The scans were of patients which were recommended for a biopsy, and all biopsy reports were made available for this work. All data was anonymised according to a data sharing protocol among North Tees and Hartlepool NHS Foundation Trust, Brunel University London, and Teesside University. In the following sections, the data and the labelling process will be presented.

3.1.2 Ethical Approval

The study was conducted by the Declaration of Helsinki and approved by the Institutional Review Board of HRA and HCRW (IRAS project ID 258617; latest amendment date: 18 December 2020).

3.1.3 Dynamic Contrast Enhanced MRI Scans

Magnetic Resonance Imaging (MRI) ([36, 37]) is a pivotal medical imaging technique that provides detailed and non-invasive insights into the internal structures of the human body. Utilizing powerful magnets and radiofrequency pulses, MRI creates high-resolution images that aid in the diagnosis, evaluation, and monitoring of a wide range of medical conditions ([38]). Unlike other imaging methods, MRI does not involve ionizing radiation, making it a safer option for repeated examinations. Its versatility allows it to visualize various soft tissues, such as the brain, muscles, joints, and organs, enabling healthcare professionals to detect abnormalities, assess disease progression, and plan appropriate treatments. The information-rich images produced by MRI play a crucial role in improving patient care across specialties, from neurology to orthopaedics, oncology, and beyond ([39]).

DCE-MRI is a crucial imaging technique for diagnosing and evaluating breast cancer. This method provides detailed insights into blood flow and vascularity within breast tissue, aiding in the detection, characterization, and staging of breast lesions. It has been proven to provide earlier and more accurate cancer detections in patients compared to the more popular techniques as of today, such as ultrasound and mammography, while providing lower long-term risk due to its lack of ionizing radiations ([40]).

Before the procedure, patients are typically asked to fast for several hours to reduce the likelihood of nausea caused by the contrast agent. Additionally, any metal objects such as jewellery or clothing with metal components need to be removed to prevent interference with the MRI scan.

During the DCE-MRI process, a gadolinium-based contrast agent is injected into a vein in the patient's arm. Gadolinium is a paramagnetic substance that enhances the visibility of blood vessels and lesions in MRI images. The patient then lies on the MRI table, and images are acquired over time, capturing the distribution of the contrast agent within the breast tissue. This series of images includes different phases: pre-contrast, arterial, venous, and delayed phases.

The acquired images are carefully analysed to create a dynamic series illustrating enhancement patterns as they evolve. Radiologists define regions of interest (ROIs) within the breast tissue and identify any suspicious lesions. These enhancement patterns are crucial indicators of lesion characteristics, such as malignancy, vascularity, and heterogeneity. For instance, rapid and intense enhancement in the early phases of DCE-MRI is often associated with malignancy due to the increased blood supply in cancerous tissue.

Furthermore, kinetic analysis is conducted to quantitatively assess the behaviour of the contrast agent within the tissue. Parameters like peak enhancement, wash-in rate, wash-out rate, and time to peak enhancement are evaluated. These metrics provide additional insights into lesion characteristics and contribute to a more comprehensive assessment.

Radiologists play a vital role in interpreting the DCE-MRI images, enhancement patterns, and kinetic analysis results. Their expertise is essential for differentiating between benign and malignant lesions based on the imaging findings.

It's important to understand that DCE-MRI is frequently used in conjunction with other imaging modalities, such as mammography and ultrasound, to provide a well-rounded evaluation of breast health. While DCE-MRI is highly sensitive in detecting lesions, it may not always provide a definitive cancer diagnosis. Biopsy or further evaluation might be necessary to confirm the presence of cancer.

Biopsy is a fundamental medical procedure used to diagnose and determine the nature of various diseases and conditions by obtaining samples of tissue or cells from the body for microscopic examination. This diagnostic technique plays a pivotal role in confirming or ruling out the presence of malignancies, infections, and other abnormalities. Biopsies provide invaluable insights into the cellular and molecular characteristics of the sampled tissue, aiding healthcare professionals in making accurate diagnoses and tailoring appropriate treatment strategies. Depending

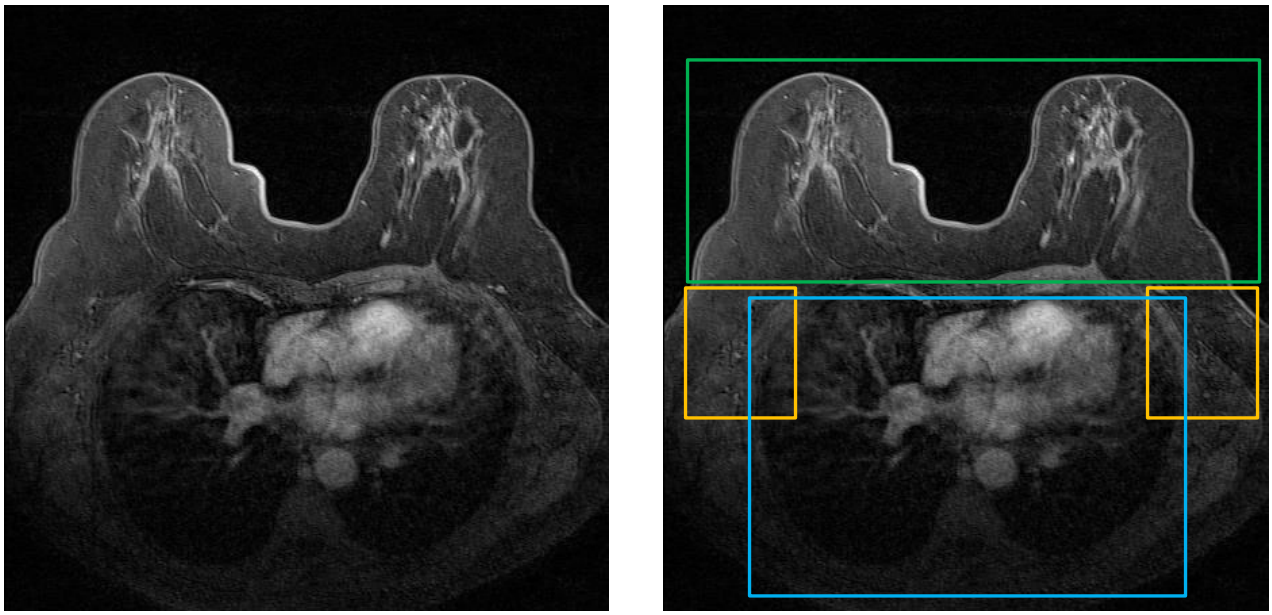
on the location and suspected condition, biopsies can be performed through minimally invasive methods such as needle biopsies, where a thin needle is used to extract tissue, or more invasive surgical procedures for larger or deeper tissue samples. The collected specimens are then sent to a pathology laboratory, where specialized experts analyse the samples under microscopes and through other advanced techniques to provide definitive information about the underlying condition.

3.1.4 Dataset Overview

The project benefitted from accessing a total of 120 T1-weighted DCE-MRI scans. All data was from patients with a biopsy-confirmed diagnosis. Each DCE-MRI scan featured 160 slices, which are 2D images obtained from sectioning the plane perpendicular to the spine every 1.1 cm. Each of the 160 slices was captured 8 times, once before contrast injection, and once every minute after contrast injection for 7 times. The data was stored in DICOM format ([41]), and the extraction of the raw pixel data resulted in images with a resolution of 448x448 pixels. In total, the project had access to 153600 images.

3.1.5 General Data Presentation

It is essential for the reader to have a basic understanding of the dataset's structure, as the subsequent sections will use specific terminology and reference morphological features that may not be immediately clear to those without prior knowledge. To aid comprehension, Figure 3.1 below provides an example of a single image.



(a) (b)

Figure 3.1: (a) single image. (b) breasts (green), lymph nodes (orange) and chest cavity (blue).

First and foremost, it is possible to notice the breasts in the upper part of the image (green area in Figure 3.1b). As the images are taken from below, the left breast is on the right side of the image. In this case, it is possible to notice an unusual bend on the inner part of the right breast (left side of the image). This is due to how the examination is conducted, during which the patient is laying down on their stomach, with the breast contained in a cage. Part of the breast is then resting on the edge of the cage. The orange area in is where the lymph nodes are usually located. A radiologist would carefully examine this part for anomalies in the lymphatic system. The blue area represents the chest cavity, in which it is possible to see the spine (grey circle at the bottom of the image) and internal organs (in this case, the heart, in the centre of the image). The chest wall separates the breast area and the chest cavity.

Each image is part of subset of images taken at the same location at each of the eight timesteps. The evaluation of how the tissue reacts to the contrast agent overtime is fundamental in the diagnostic process. A set of 8 images at any timestep will be referred to as a slice. The variation of the relative pixel brightness overtime is the quantification of how tissues are reacting to the contrast agent. The resulting patterns are referred to as enhancement.

Each slice has a Z dimension assigned to it, corresponding to the vertical location relative to an arbitrary Z = 0 point. The slices are separated by 1.1 mm among each other in this specific dataset, covering a total of 17.6 cm. Depending on the dimensions of the chest cavity and breasts of the patient, the real-world dimensions of a single pixel can vary

slightly. The deviation, however, is rather minimal and does not influence the study, thus the reader should consider the area covered by a single pixel to be 0.76mmx0.76mm.

In the following sections, the data will be presented as slices and no single images for clarity reasons and to provide necessary context.

3.1.6 Dataset Labelling & Exploration

The data was accompanied by a spreadsheet containing some relevant information on the radiologist's report and biopsy report. A simplified table was extracted to remove additional information about the biopsy record and surgical procedure. An example of a few records can be found in the table here below.

Table 3.1: example of records containing relevant information from radiologist's report and biopsy report.

Lesion Characterisation	Side	Location	Root Sign	Margin	Enhancement	Categorisation
Mass	Right	LOQ	No	Irregular	Heterogenous	Suspicious
Mass	Right	Lower central	No	Multifocal	Heterogenous	Suspicious
Mass	Right	Lower central	No	Multifocal	Heterogenous	Suspicious
NME	Left	Lateral	No	Irregular	Heterogenous	Suspicious

The spreadsheet was provided by expert radiologists and should be considered as the ground truth. In the following paragraphs, a brief overview of the available information is provided.

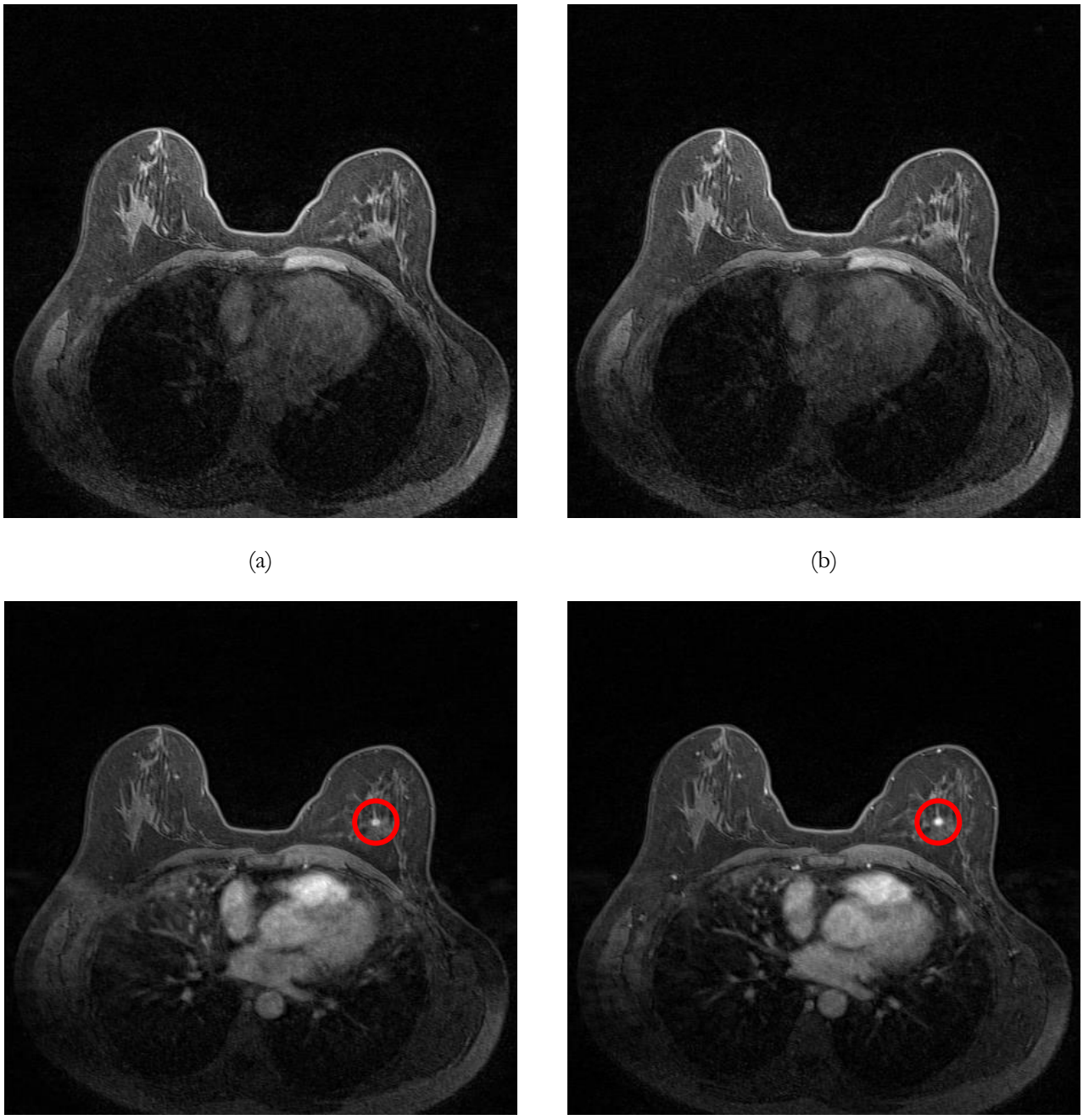
The meanings of the main columns in such spreadsheet are as follows:

- The "Lesion Characterisation" column refers to the nature of lesion as perceived by the radiologist during the diagnosis of the MRI. The possible values were "Mass", "NME", "Multifocal", and "Axillary Nodes". The value "Mass" refers to a well-defined, focal area of enhancement within the breast tissue. "Non-Mass Enhancement (NME)" refers to a finding that doesn't correspond to a distinct mass or lesion. "Multifocal" refers to a case in which there are multiple instances of suspect lesions in the breast volume. The scans featuring a "Axillary Nodes" diagnosis refer to the presence of cancer or metastasis in the axillary lymph nodes, but no breast cancer.

- The "Location" column in the table refers to where the suspect lesion has been identified. The standard procedure divides each breast in quadrants (Upper/Lower, Outer/Inner), hence a field with value "UOQ" refers to the fact that the location of the lesion was found in the upper outer quadrant, and so on.
- The "Root Sign" column, also named Spiculation, is a Boolean which refers to the appearance of the lesion.
- The morphology of the lesion is also characterised by the column dubbed "Margin", which characterises how the overall shape of the lesion is. The possible values for this field are "Smooth" and "Irregular". Lesions with smooth margins are more likely to be benign or non-cancerous. Irregular margins. On the other hand, can suggest a higher likelihood of malignancy or cancerous growth, as they may indicate invasive or aggressive behaviour of the lesion.
- The "Enhancement" field describes the uniformity of the relative reaction to contrast within the lesion itself. The possible values were "Heterogeneous" and "Homogeneous".

For better understanding the meaning of such information and the way they are evaluated by radiologists, it is important to associate them to images; therefore, a set of examples are reported in the following section and further explained. Every figure consists of 8 images creating a complete scan sequence, and for ease of reading they are split onto two pages.

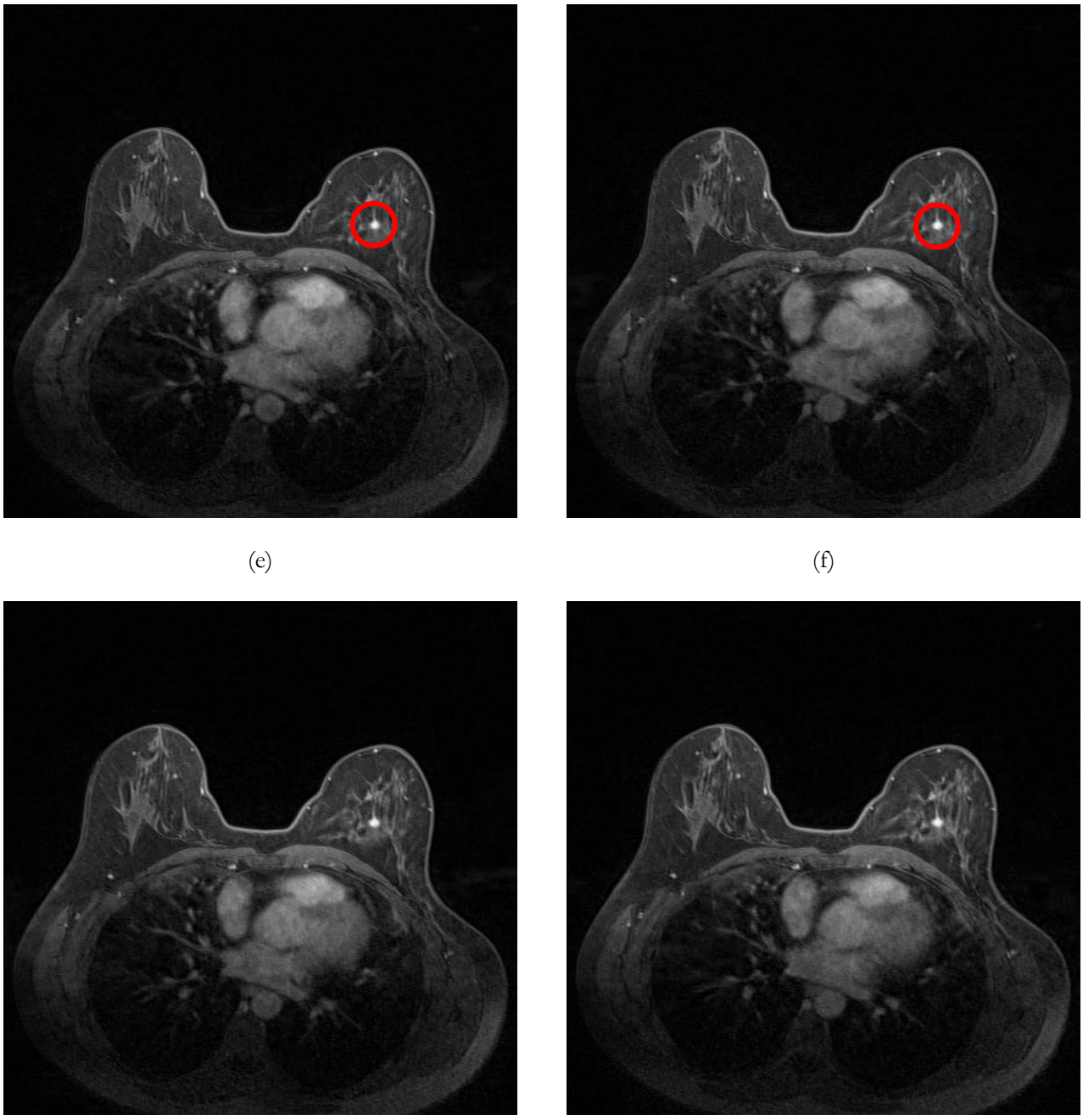
In the following figures 3.2 and 3.3, an example of a slice with a mass can be found. Such images will be used as a reference in the following of the section.



(a) (b)

(c) (d)

Figure 3.2: first batch of 4 timesteps from the same reference sequence in which a mass is visible. The red circles in (c) and (d) highlights the suspect area. It is important to note the enhancement behaviour of the area inside the chest wall, as well as other smaller regions in the axillary area and breast area.

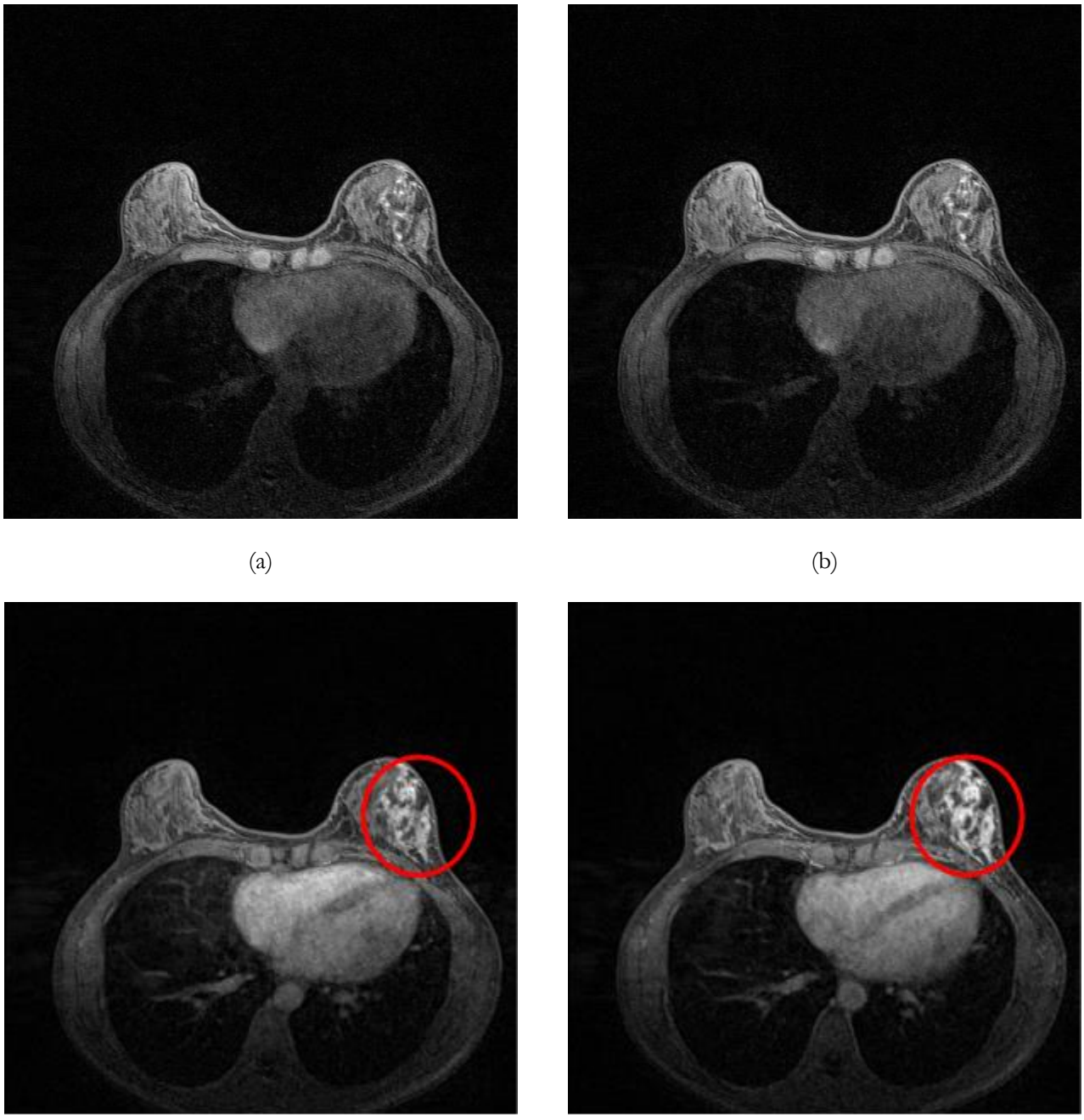


(e) (f)

(g) (h)

Figure 3.3: second batch of 4 timesteps from the same reference sequence in which a mass is visible. The red circles in (e) and (f) highlight the suspect area. It is important to note the enhancement behaviour of the area inside the chest wall, as well as other smaller regions in the axillary area and breast area.

As previously explained, NME refers to a finding that doesn't correspond to a distinct mass or lesion. Instead, it represents an area of enhancement in the breast tissue that doesn't have a well-defined border. An example of a slice showing NME can be found below in figures 3.4 and 3.5.



(a) (b)

(c) (d)

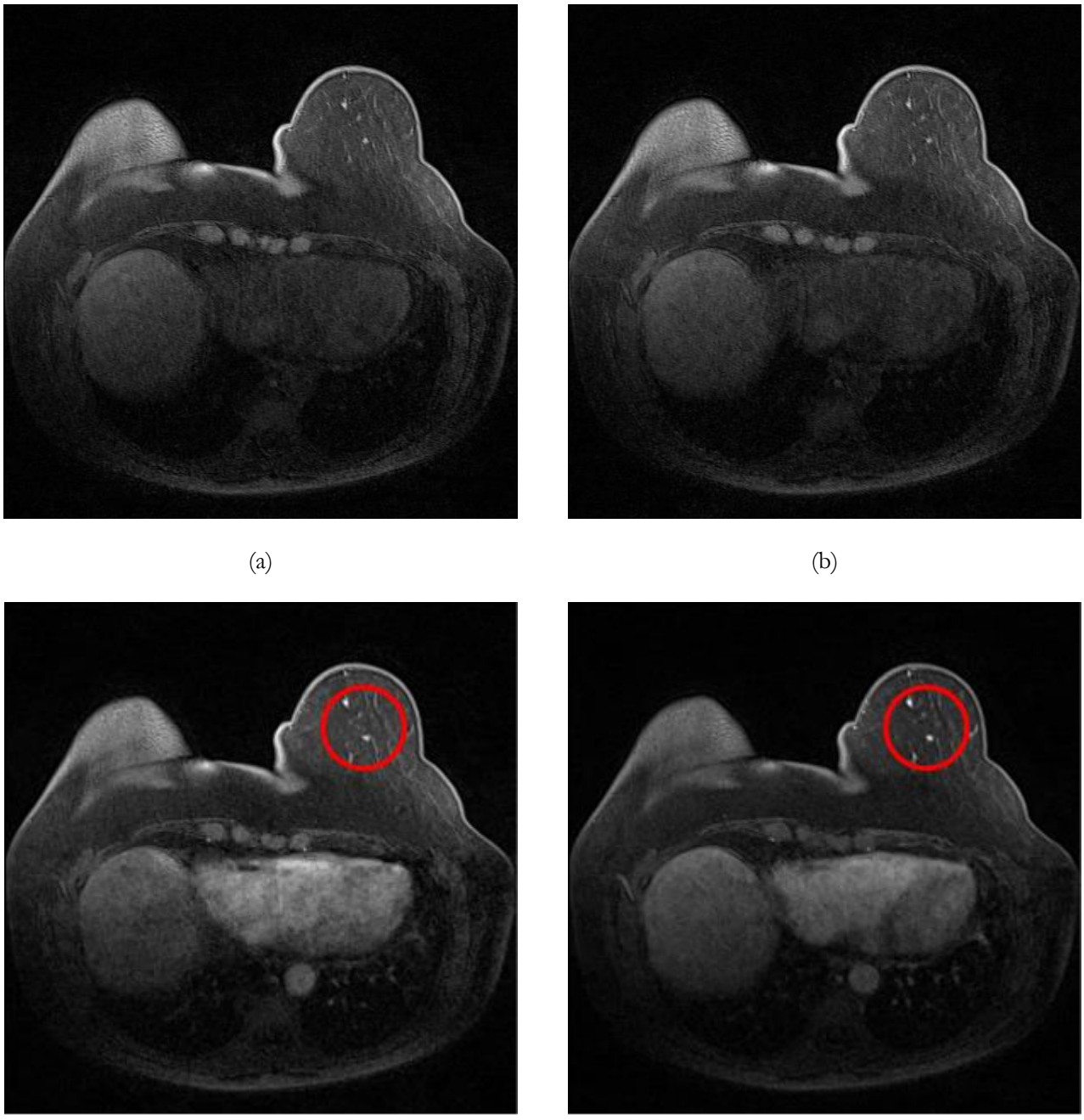
Figure 3.4: first batch of 4 timesteps from a sequence showing Non-Mass Enhancement. The red circles in (c) and (d) highlight the suspect area.

(e) (f)

(g) (h)

Figure 3.5: second batch of 4 timesteps from a sequence showing Non-Mass Enhancement. The red circles in (e) and (f) highlight the suspect area.

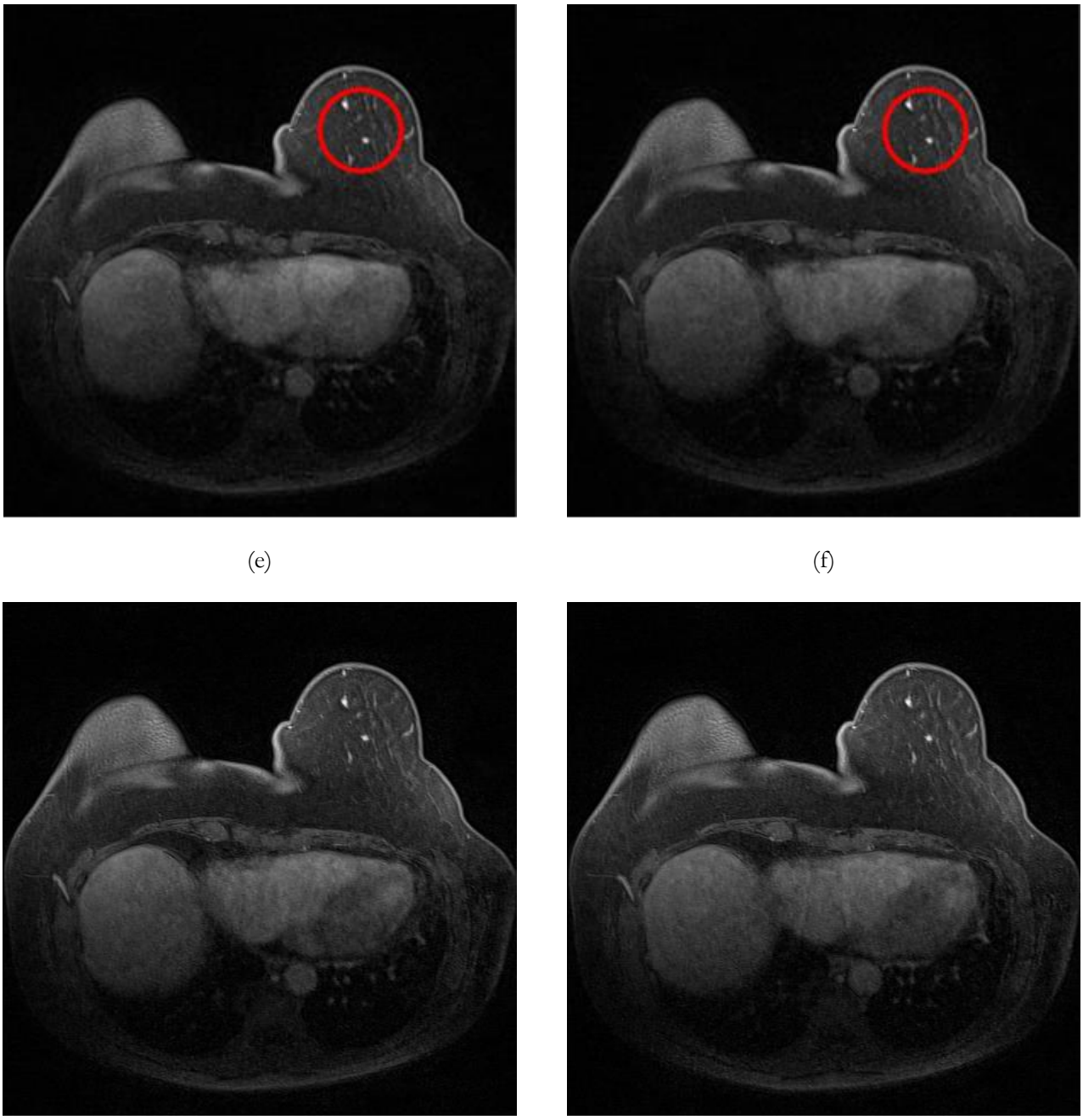
In the case of "Multifocal", there is the presence of multiple instances of suspect lesions in the breast volume. These can vary in shape, size, and appearance. The position of masses is also variable throughout the volume, hence often the diagnosis corresponds to the identification of multiple suspect lesions. An example of a multi-mass slice in which multiple distinct masses can be seen below in Figure 3.4. It is important to reiterate that, except for extreme cases, the appearance of a single slice of a scan resulting in a "Multifocal" diagnosis will appear like a "Mass" diagnosis.



(a) (b)

(c) (d)

Figure 3.6: first batch of 4 timesteps from a sequence showing Multifocal lesion. The red circles in (c) and (d) highlight the suspect area. The three visible lesions are quite small and are located close to the top left of the circle, bottom left and middle.



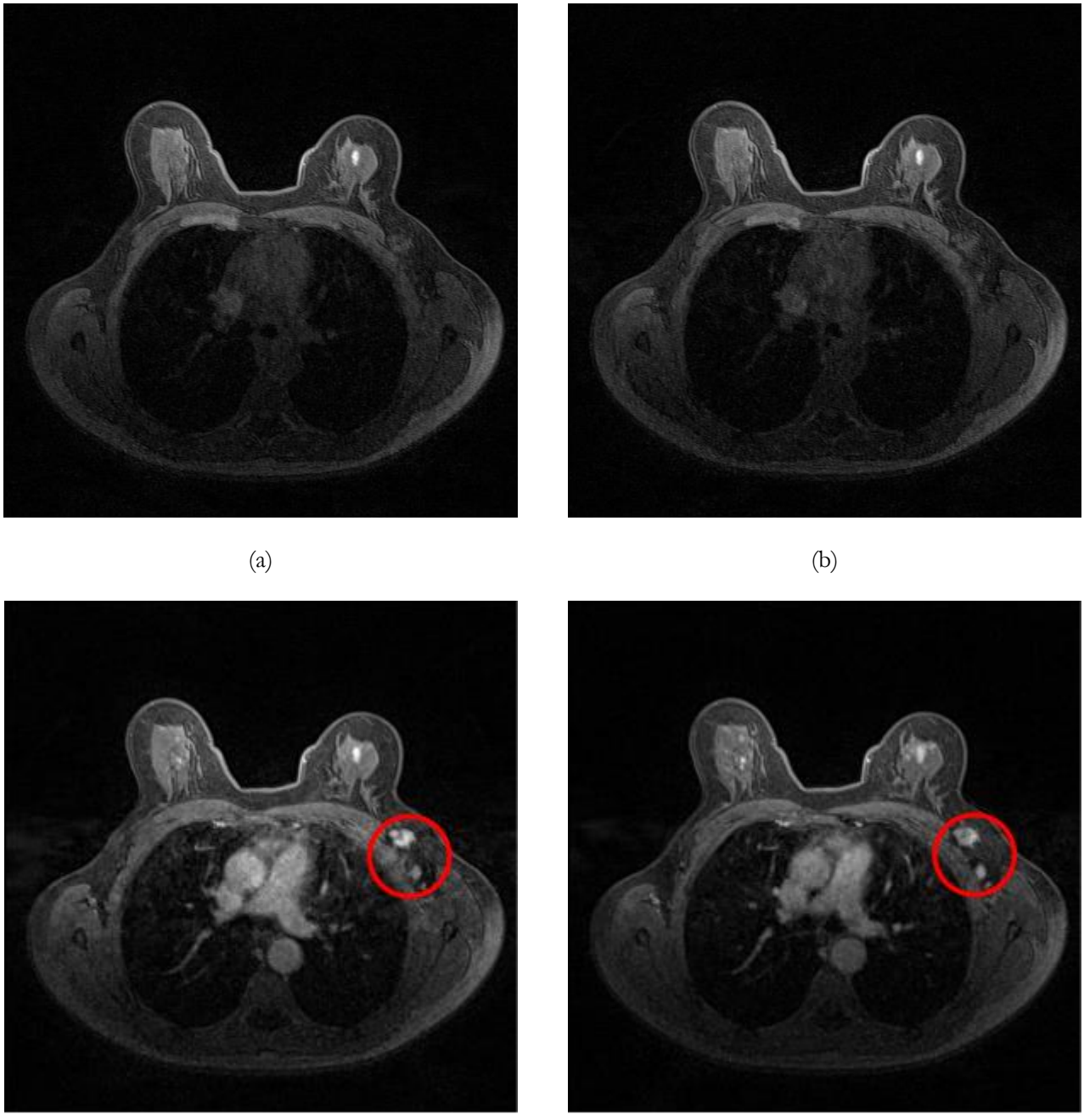
(e) (f)

(g) (h)

Figure 3.7: second batch of 4 timesteps from a sequence showing Multifocal lesion. The red circles in (e) and (f) highlight the suspect area. The three visible lesions are quite small and are located close to the top left of the circle, bottom left and middle.

Axillary nodes, also known as axillary lymph nodes, are a group of lymph nodes located in the axilla, which is the area under the arm or armpit. These nodes are an essential part of the lymphatic system and play a crucial role in filtering and draining lymphatic fluid from the breast, upper arm, and surrounding areas. The dataset contained one example

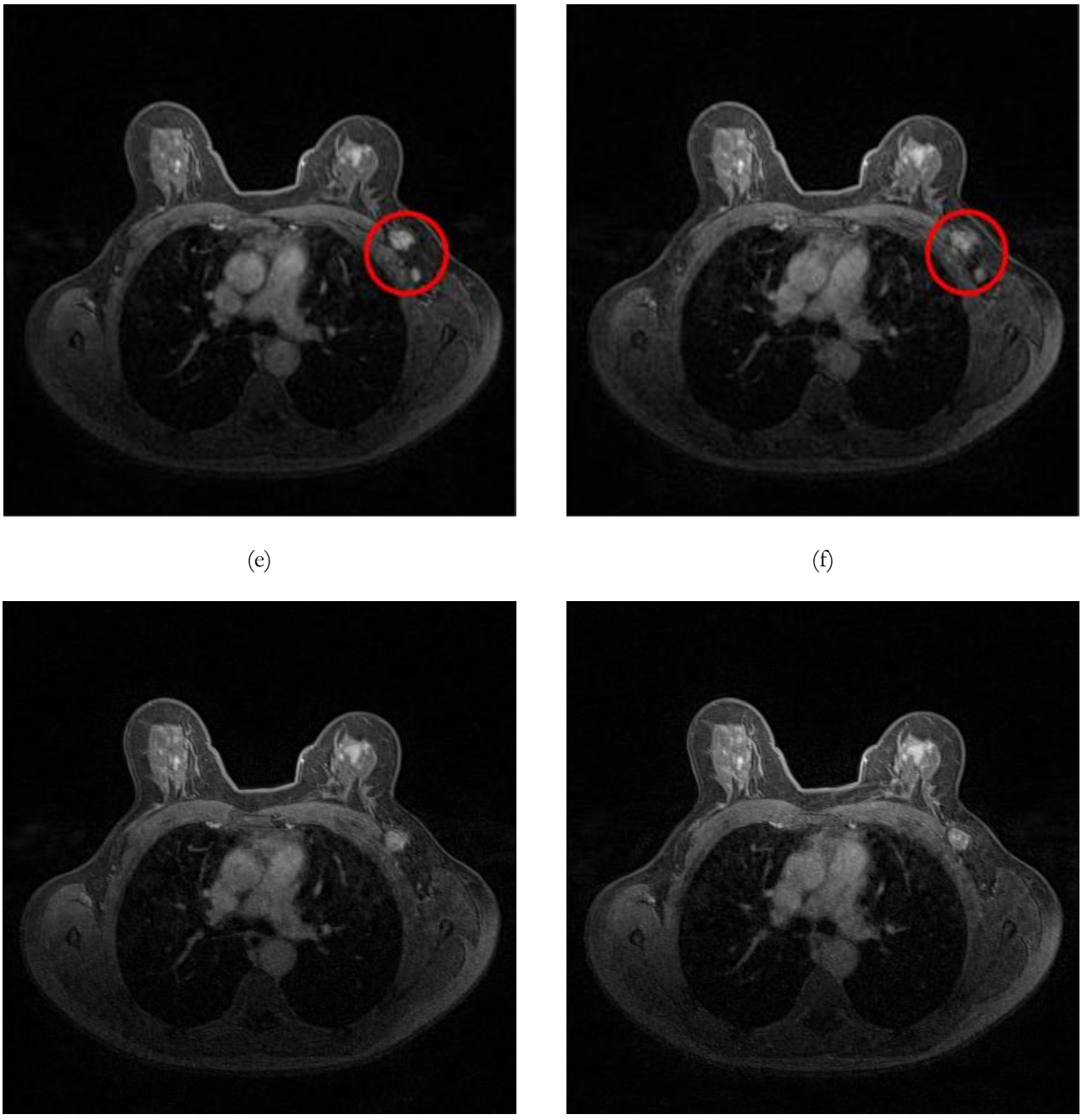
of this, and a representative slice is reported in figures 3.8 and 3.9. It is worth noting that an untrained eye might spot a mass in the centre of the left breast (right side of each image). The mass was confirmed to be benign, and more details on how one might have concluded this are provided in the following sections.



(a) (b)

(c) (d)

Figure 3.8: first batch of 4 timesteps from a sequence showing a visible Axillary nodes diagnosis. Red circles in (c) and (d) highlight the suspect area. The images show a clear abnormality in the middle of the left breast, especially in (a) and (b). This has been biopsy-confirmed to be a benign lesion.



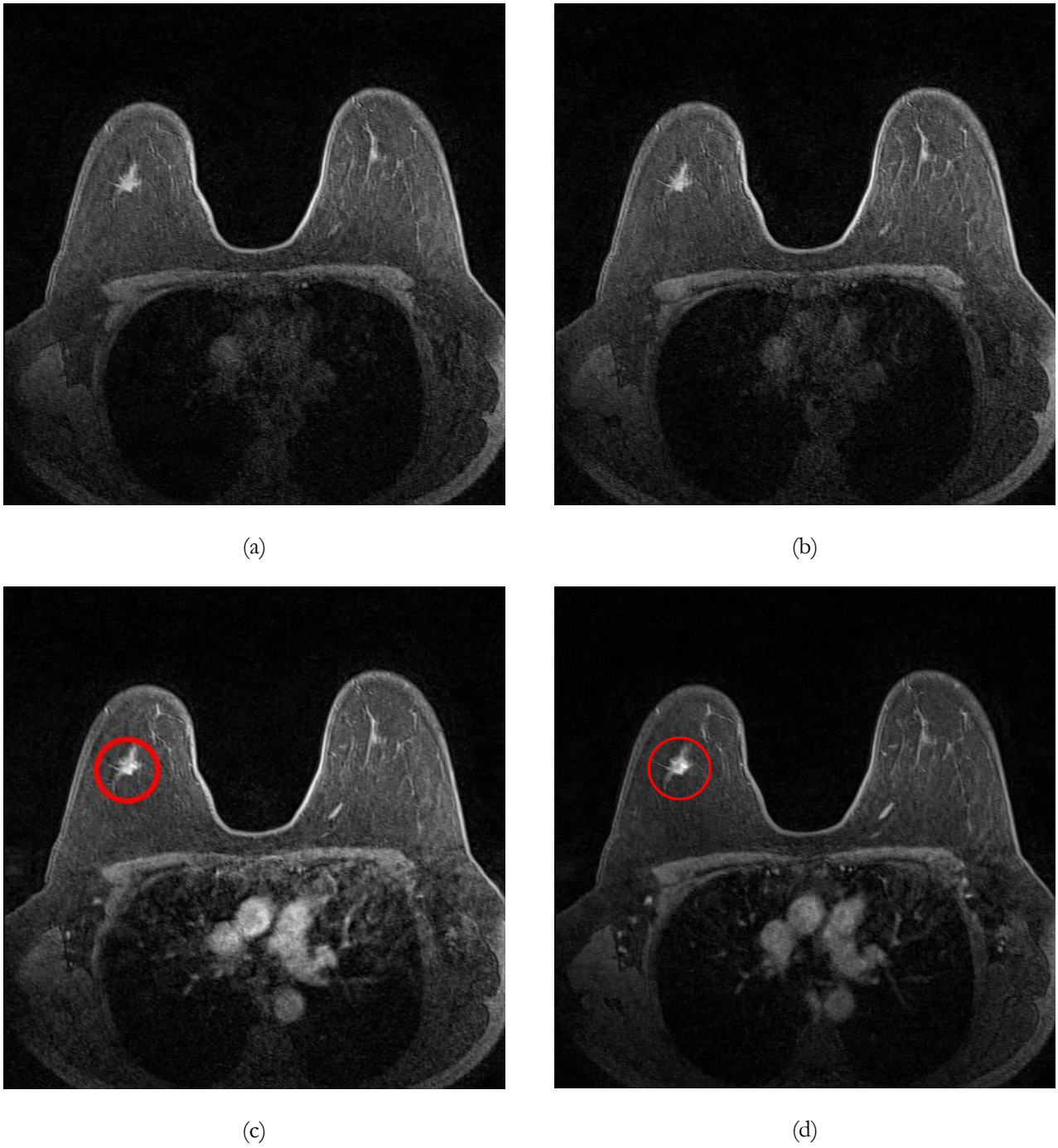
(e) (f)

(g) (h)

Figure 3.9: second first batch of 4 timesteps from a sequence showing a visible Axillary nodes diagnosis. Red circles in (d) and (f) highlight the suspect area. The images show a clear abnormality in the middle of the left breast. This has been biopsy-confirmed to be a benign lesion.

An example of a spiculated lesion, referring to the Root Sign column, is shown in figures 3.10 and 3.11 below. As it can be seen by the pictures, a spiculated lesion, also known as a spiky or irregular lesion, has irregular, jagged, or spiky edges. It often looks like it has tentacle-like extensions radiating out from the central mass. Spiculated lesions are more

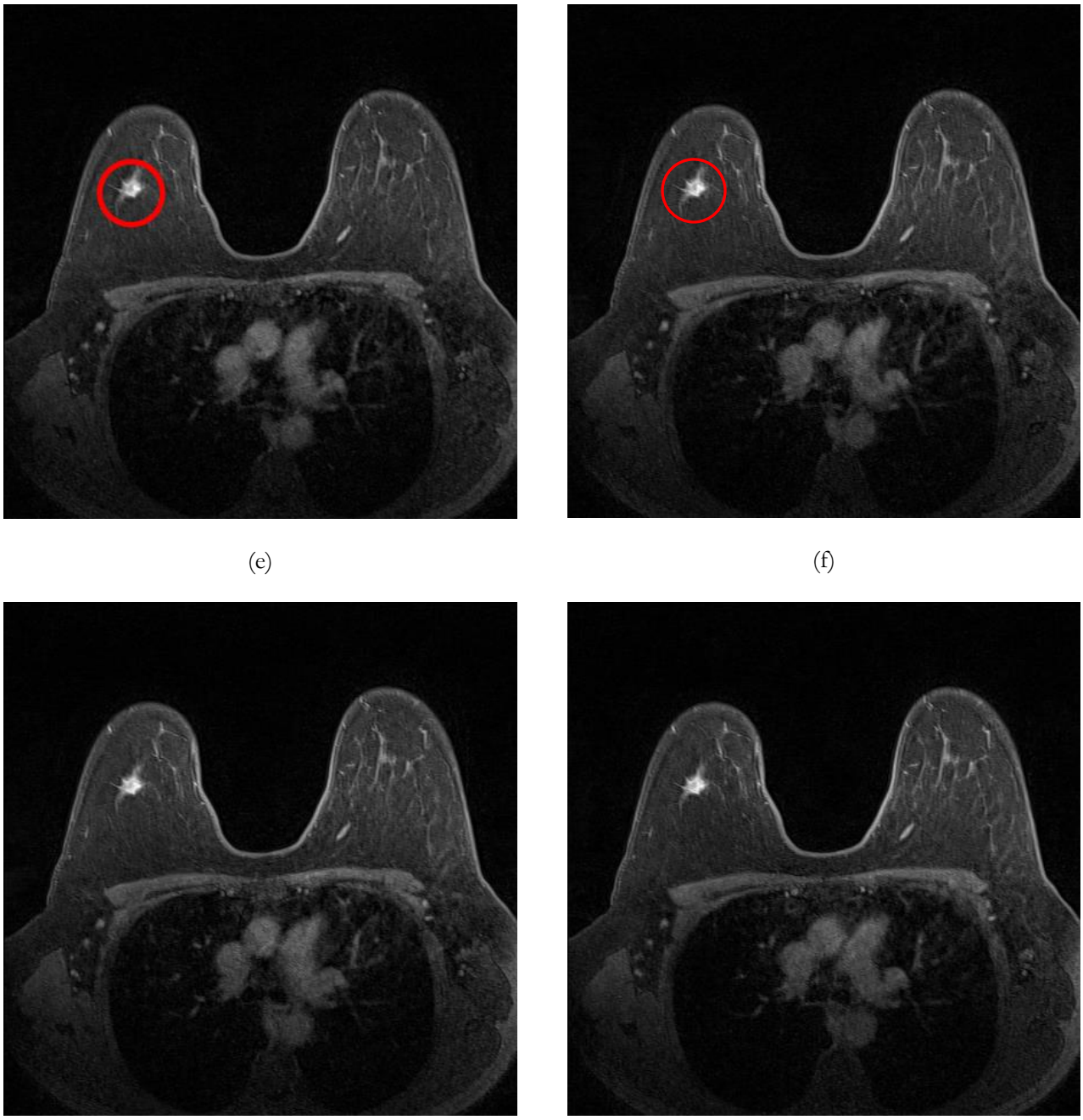
concerning because their irregular shape is often associated with malignant or cancerous growth. They may indicate invasive or aggressive behaviour of the lesion.



(a) (b)

(c) (d)

Figure 3.10: first batch of 4 timesteps from a sequence showing a spiculated lesion. The red circles in (c) and (d) highlight the suspect area.

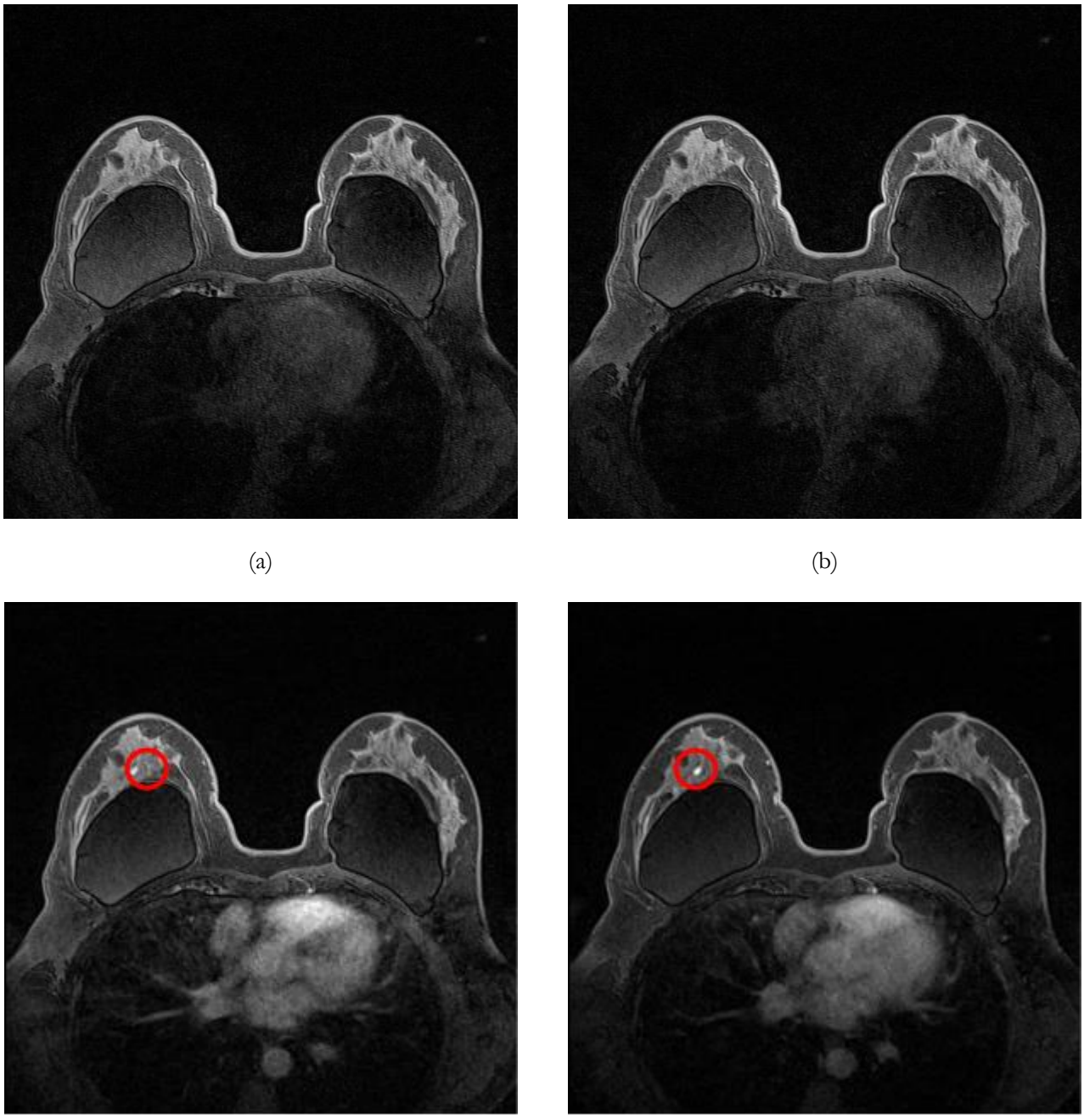


(e) (f)

(g) (h)

Figure 3.11: second batch of 4 timesteps from a sequence showing a spiculated lesion. The red circles in (g) and (h) highlight the suspect area.

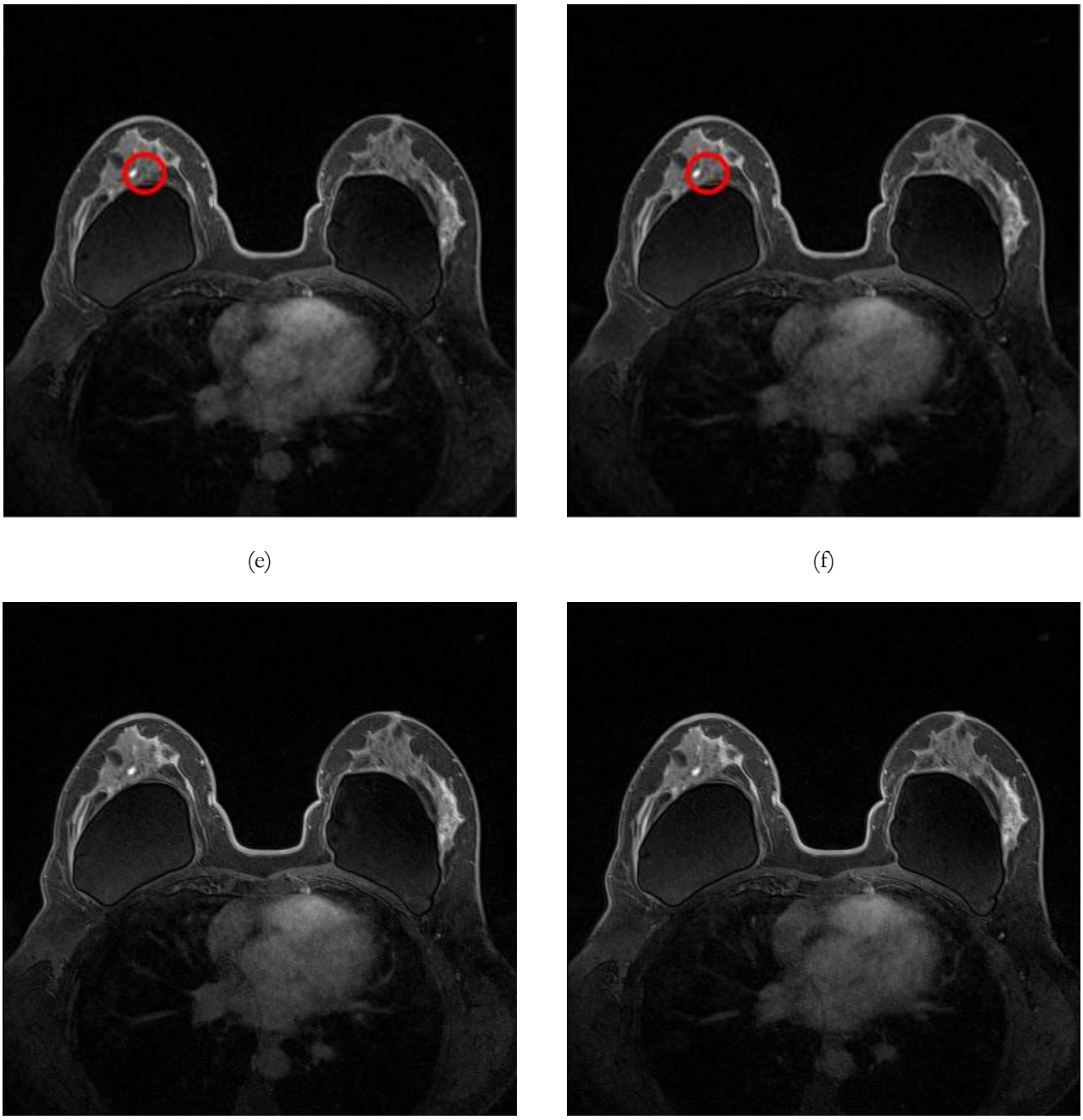
A smooth margin, also known as a well-defined or regular margin, refers to the presence of clear, smooth, and well-defined edges or boundaries around a lesion. It typically appears as a round or oval shape with a uniform border that is easy to distinguish from the surrounding tissue. Lesions with smooth margins are more likely to be benign or non-cancerous. An example of a smooth lesion is available in figures 3.12 and 3.13, which provide an example of the appearance of an MRI of a patient with breast implants.



(a) (b)

(c) (d)

Figure 3.12: first batch of 4 timesteps from a sequence showing a smooth lesion. The red circles in (c) and (d) highlight the suspect area. The image shows the appearance of breast implants in DCE-MRI.



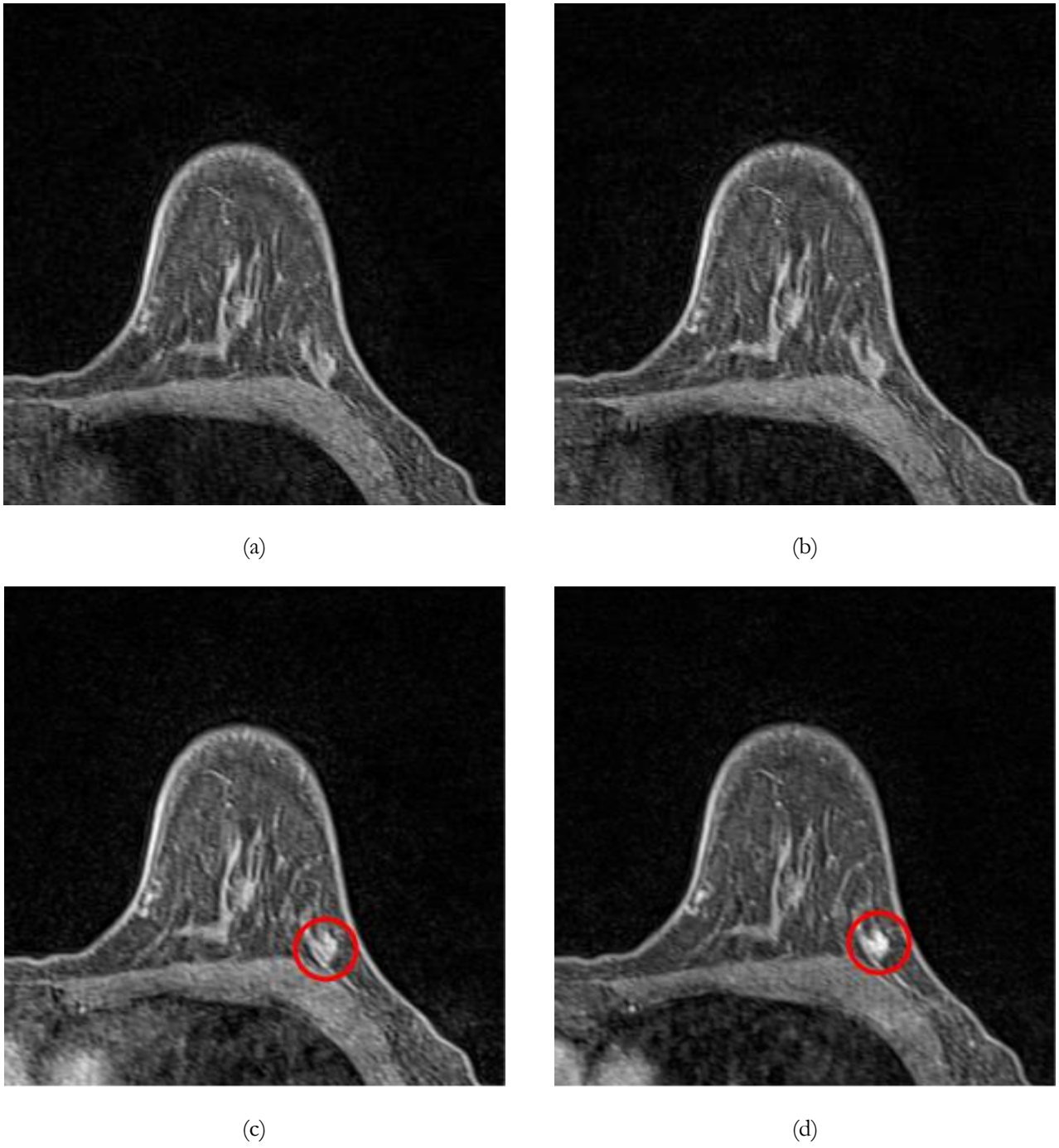
(e) (f)

(g) (h)

Figure 3.13: second batch of 4 timesteps from a sequence showing a smooth lesion. The red circles in (e) and (f) highlight the suspect area. The image shows the appearance of breast implants in DCE-MRI.

An irregular margin refers to the presence of edges or boundaries that are not well-defined and exhibit an uneven or jagged appearance. Irregular margins can suggest a higher likelihood of malignancy or cancerous growth, as they may

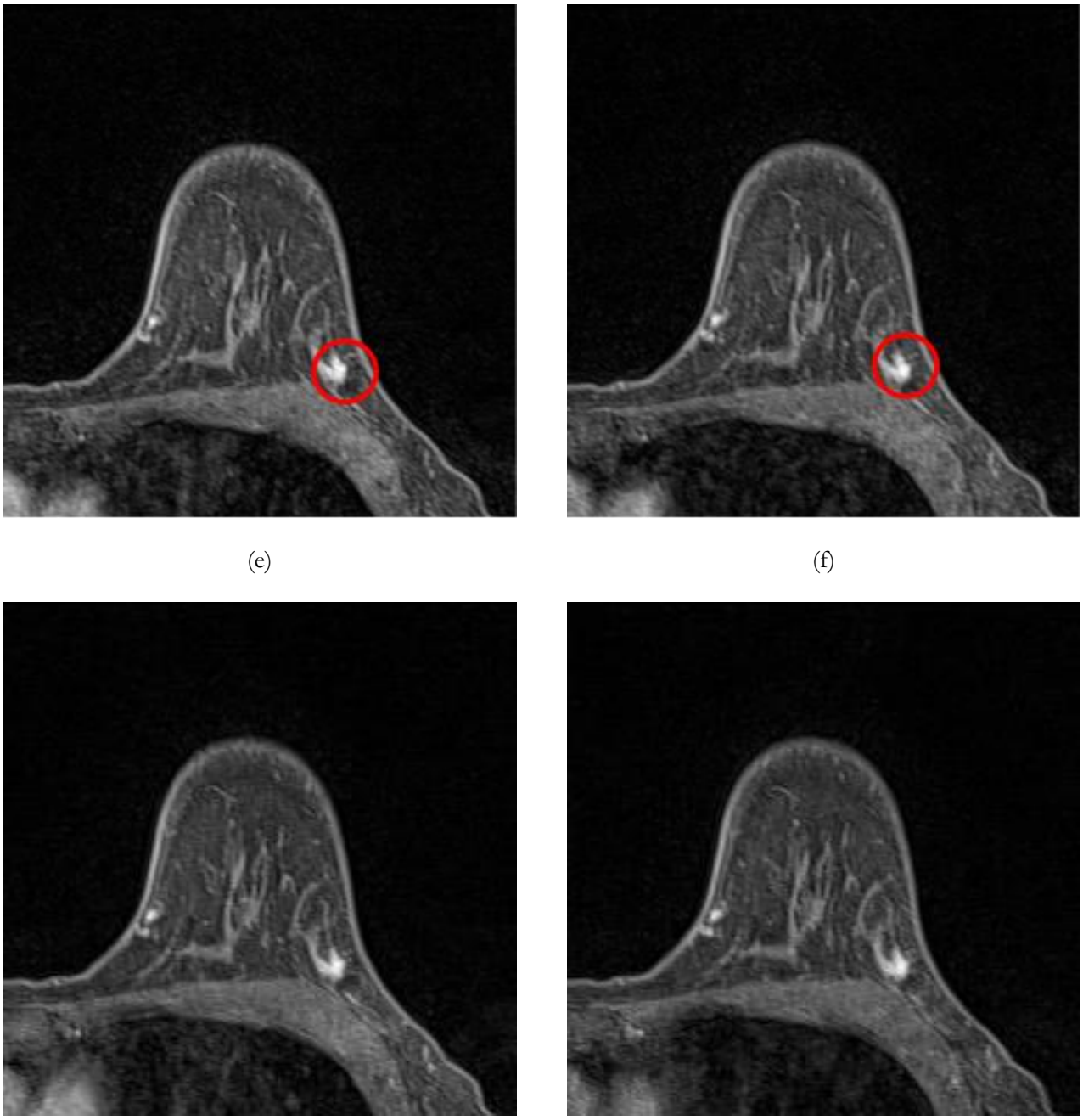
indicate invasive or aggressive behaviour of the lesion. An example of an archetypical irregular lesion is shown in figures 3.14 and 3.15.



(a) (b)

(c) (d)

Figure 3.14: first batch of 4 timesteps from a sequence showing an irregular lesion. The red circles in (c) and (d) highlight the suspect area. The images are cropped to allow the viewer to appreciate the shape of the lesion.



(e) (f)

(g) (h)

Figure 3.15: second batch of 4 timesteps from a sequence showing an irregular lesion. The red circles in (e) and (f) highlight the suspect area. The images are cropped to allow the viewer to appreciate the shape of the lesion.

Figure 3.12 and 3.13 (relative to the smooth lesion) show a homogenous enhancement behaviour, whereas figures 3.14 and 3.15 (relative to an irregular lesion) show a heterogenous enhancement behaviour.

3.1.7 Observations on Available Data

The available dataset is considerably different from the typical dataset found in literature, as all MRI scans resulted in a biopsy recommendation, hence the radiologist evaluated all of them to be potentially malignant. The result is an inherit bias towards malignancy which can be mitigated by the techniques shown in this work. However, evaluating the performances of the methodologies in terms of TNR would be statistically irrelevant. Moreover, the project aims to deliver diagnostic support to radiologists in identifying the lesions, limiting the misses, and speeding up the whole process. An overview of the distributions of the labels is shown in figures below.

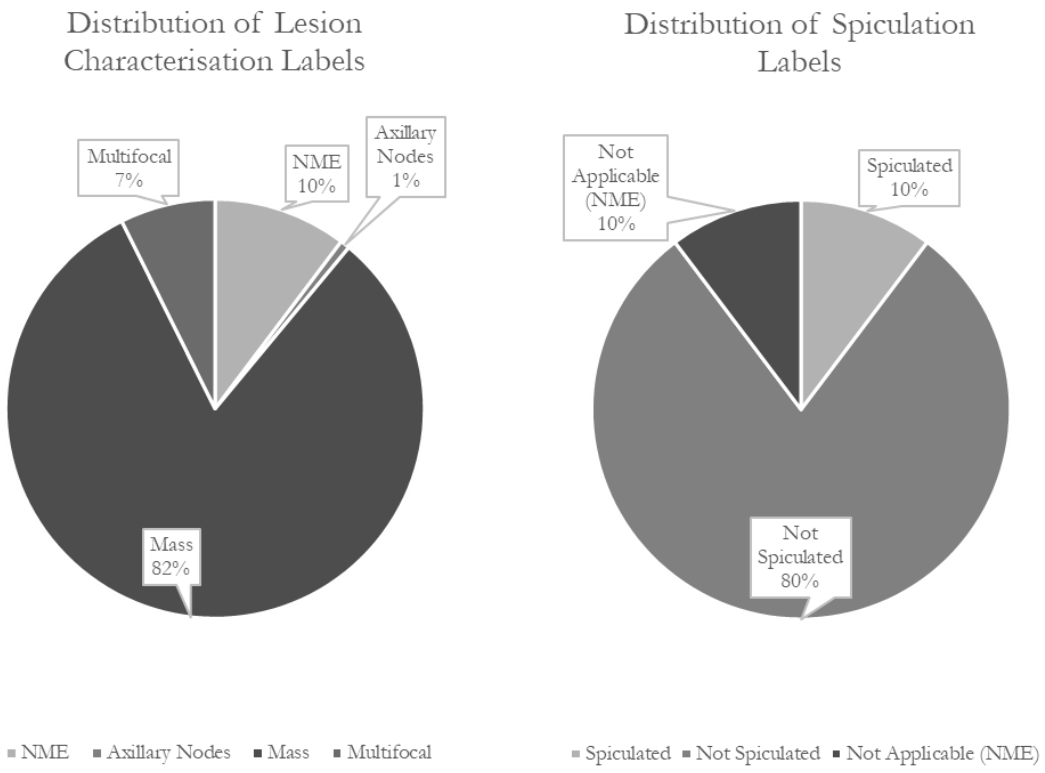


Figure 3.16: statistical distribution of lesion characterization labels and spiculation labels.

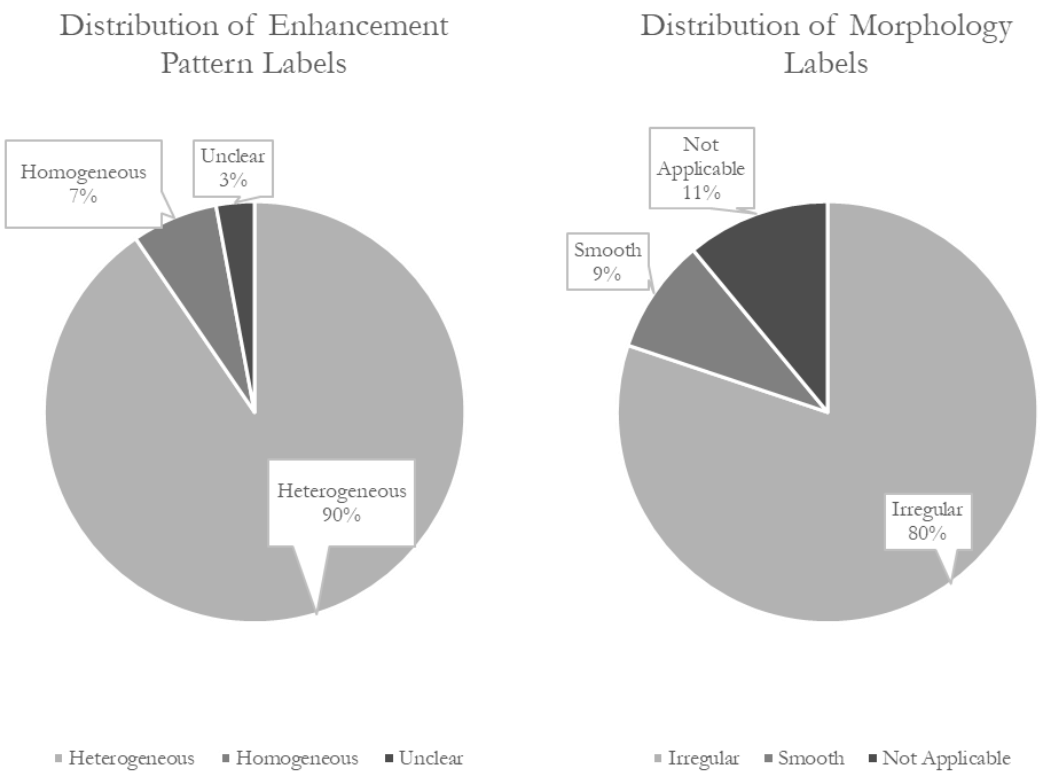


Figure 3.17: statistical distribution of enhancement pattern labels and morphology labels.

The distributions of data paint a partial picture of the challenging nature of this dataset compared to the ones available in literature. However, it allowed the development of robust algorithms that can cover the wide array of real-world examples of malignant lesions, thus augmenting the impact of this work.

3.2 Diagnostic Process

3.2.1 Traditional Diagnostic Workflow

Literature on the topic of breast cancer diagnosis is plentiful ([42 – 46]). For this work, a limited introduction will be provided, using figures 3.2 to 3.14 as reference and a simplified version of the schematic as in [47].

The diagnostic and reporting processes are standardized by BI-RADS (Breast Imaging Reporting & Data System) ([48]). The general procedure begins with a visual examination of the volume, with abnormalities being noted by the radiologist. As masses are identified, a standard flowchart is followed, a version of which is found in Figure below. An interactive flowchart such as this one is available at [49].

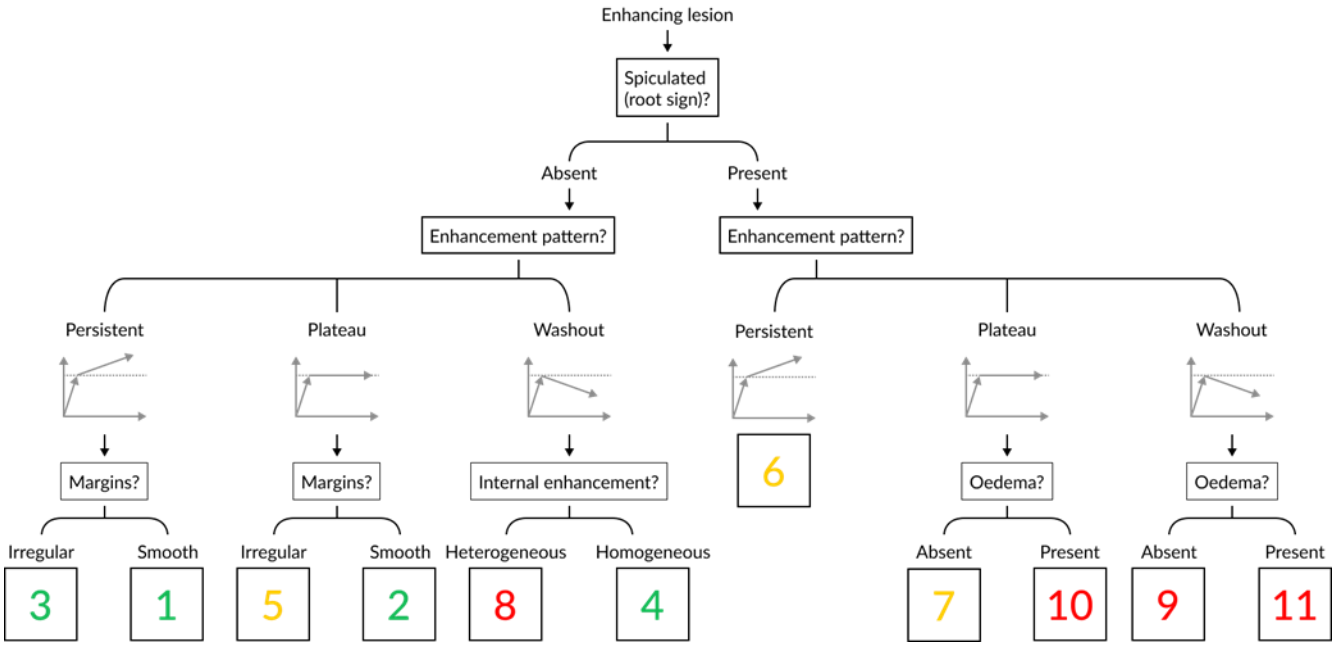


Figure 3.18: flowchart for Kaiser Score classification.

The output of this standard flowchart is a score, known as Kaiser score ([50]), which indicates the likelihood of malignant lesions and thus the recommendation for biopsy. The low likelihood scores (1-4) signify a low chance of malignancy; hence patients are usually asked to reschedule an exam in a span of a few weeks or months. All other scores (5-11) represent a likelihood of malignancy which warrants a biopsy. It is important to note that the flowchart is often not followed to the letter, as variance in these cases is extremely high, and doctors must rely on their great experience to make critical decisions.

The presence of spiculation is the first and most impactful observation, as a strong correlation between it and lesion malignancy has been shown. Spiculation is present in figures 3.4, 3.5, 3.10, 3.11, 3.14, 3.15.

The enhancement pattern (or delayed phase) is the second determinant factor and will be discussed in detail in the following paragraph, as the unfortunate truth is that often these enhancement patterns are highly subjective, and doctors rely on several years of experience to determine if a lesion "looks" to be enhancing in a certain way. Tentative categorizations for this parameter are shown in the following table below, which was confirmed by an expert radiologist.

Table 3.2: categorisation of enhancement patterns for the previously presented scans.

Figure	Enhancement pattern
3.3, 3.4	Plateau
3.5, 3.6	Persistent
3.7, 3.8	Persistent
3.9, 3.10	Washout
3.11, 3.12	Persistent
3.13, 3.14	Plateau
3.5, 3.16	Washout

The morphology classification is only relevant in a restricted number of cases; however, the low resolution of the images makes the task of classifying the spiculation of a lesion particularly hard. Hence, in practice, the lesion morphology is the easiest and fastest classification to be made, which in some cases excludes the possibility of spiculation. The only example of a smooth lesion is in figure 3.12 and 3.13. The homogeneity of the enhancement is relevant to non-spiculated, washout lesions. The examples of heterogeneous enhancement are figures 3.2, 3.3, 3.8, 3.9, 3.10, 3.11. It is worth noting that all of these are larger lesions: this is the cases for most heterogeneous lesions in the dataset. The presence of oedema, necessary for the right side of the chart, is determinable by observation of a T1-weighted MRI of the breast, which is not part of this study, as the results are invariably leading up to a biopsy.

As a conclusion to this paragraph, it should be noted that the subjective decisions made by radiologists to classify borderline cases make life-changing impact every day on real people. When confronted with a hard decision, it is human and ethically sound to err on the side of caution.

3.2.2 Kinetic cures and enhancement patterns

Kinetic curves in DCE-MRI represent the transient behaviour of tissues when reacting with the contrast agent. They're fundamental to the diagnostic process, as highlighted by the previous paragraph, because they provide information about the blood flow in a particular area. They are the main way to visually identify cancerous lesions and are of great help during the characterization of suspect areas. A radiologist would often observe the enhancement patterns by scrolling through the 4-dimensional MRI scan, and later use this information to determine a Kaiser score.

The enhancement patterns refer to how kinetic curves in a suspect lesion behaves. As the contrast agent is injected, the pixel intensity for all tissues rises. As the circulation stabilizes, suspect tissues will behave in one of three possible ways:

- Persistent behaviour: the intensity of the pixels rises slowly.
- Plateau behaviour: the intensity of the pixels stabilizes.
- Washout behaviour: the intensity of the pixels sharply drops.

A stereotypical representation of these curves is found in the following figure.

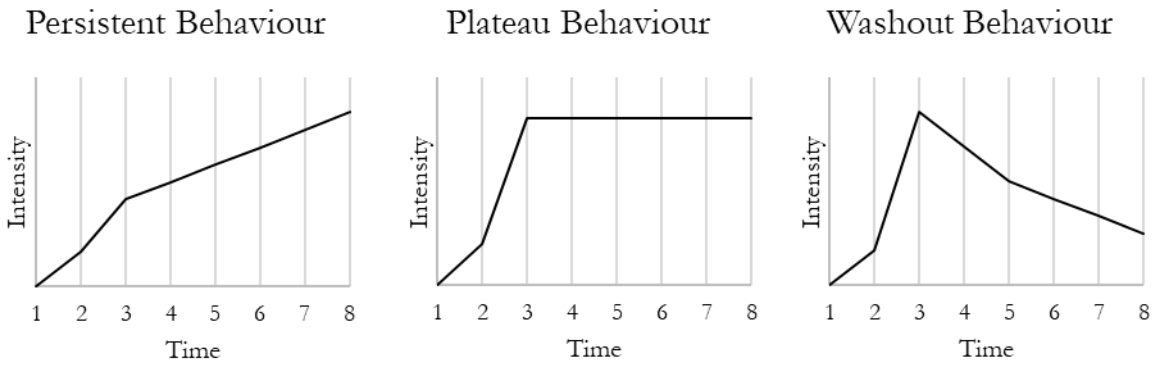


Figure 3.18: stereotypical representation of Persistent behaviour, Plateau Behaviour and Washout Behaviour curves.

The division in three distinct categories provides easily interpretable reports, while enhancing the ability to research on the topic at a clinical level. As an example, area of 10 pixels by 10 pixels was selected from the lesion in figures 3.14 and 3.15, as shown in figure 3.19 below.

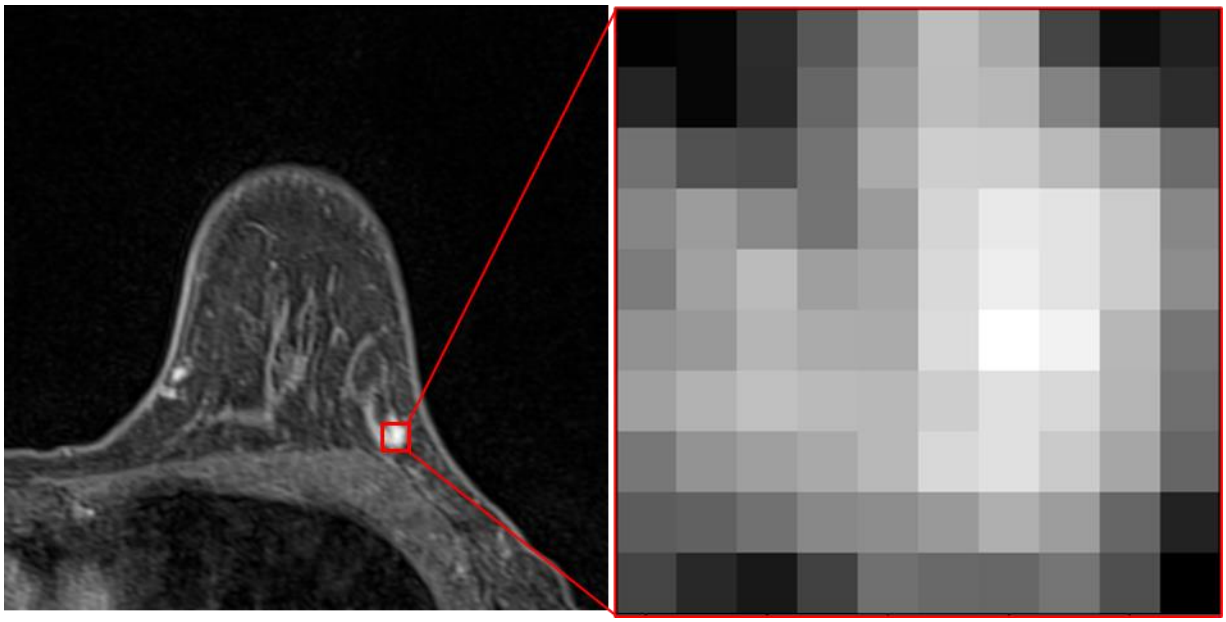


Figure 3.19: detailed view of a suspect lesion in exam. It is important to note how the resolution is insufficient to determine boundaries of the lesion.

The average enhancement curve relative to this area, along with the pixel variance, which is a suspect lesion, is shown in figure 3.20. The behaviour is persistent, with a steady increase in intensity.

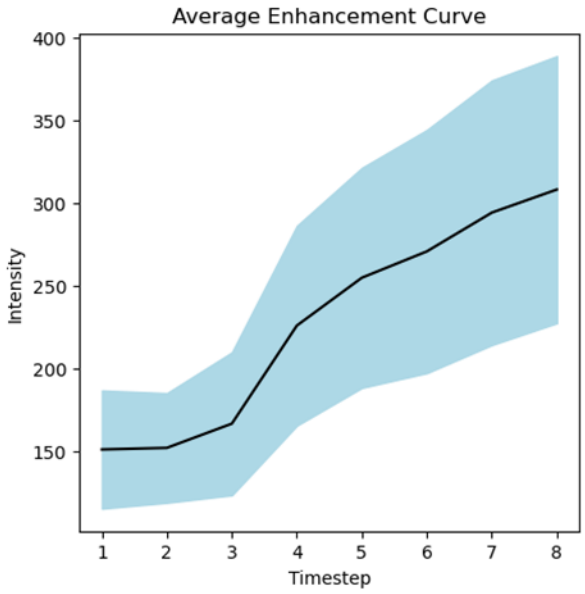


Figure 3.20: average enhancement curve related to figure 3.13, clearly showing a Persistent behaviour.

Classifying the enhancement pattern, however, is often extremely hard, as parts of a heterogeneous lesion can exhibit different enhancement patterns. The clinician is left to decide what category the lesion falls into based on their experience and professional evaluation on the likelihood of the lesion being malignant. As an example of this, the non-

mass enhancement in figures 3.4 and 3.5 was evaluated similarly to the previous example, shown below in Figure 3.21, with an area of 20 pixels by 20 pixels to account for the size of the enhancing region.

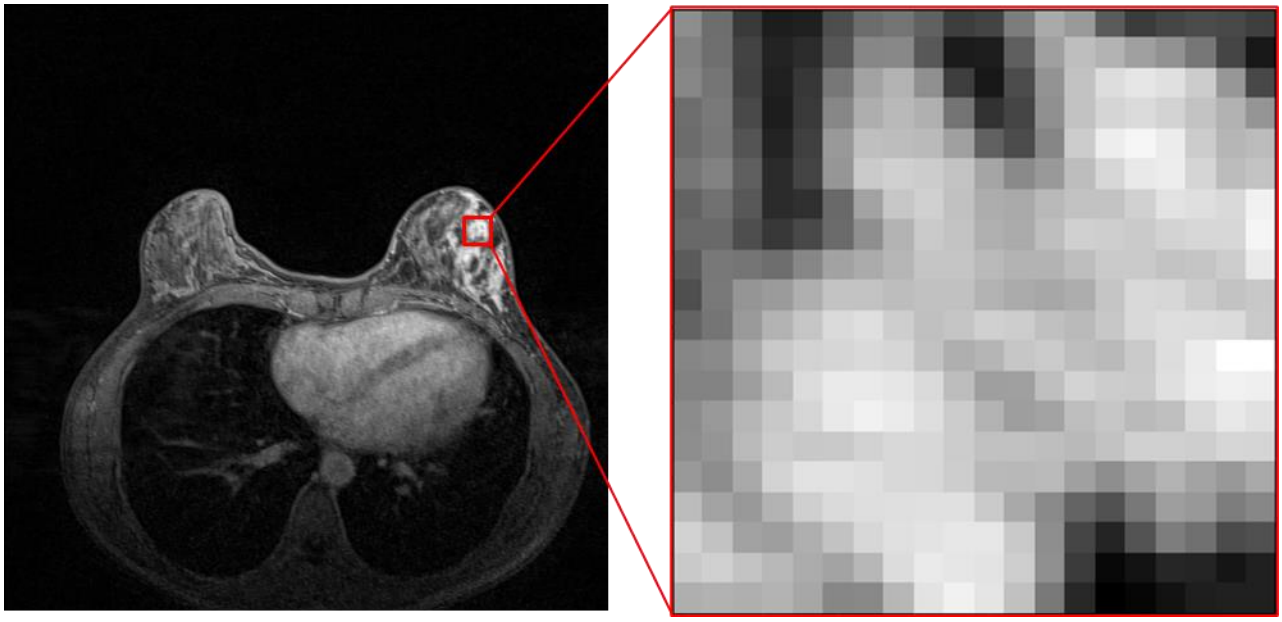


Figure 3.21: detailed view of a suspect lesion in exam. It is important to note how the resolution is insufficient to determine boundaries of the lesion.

The average enhancement curve relative to this area, along with the pixel variance, is shown here below. The behaviour is clearly plateau, with steady intensity levels throughout the suspect area.

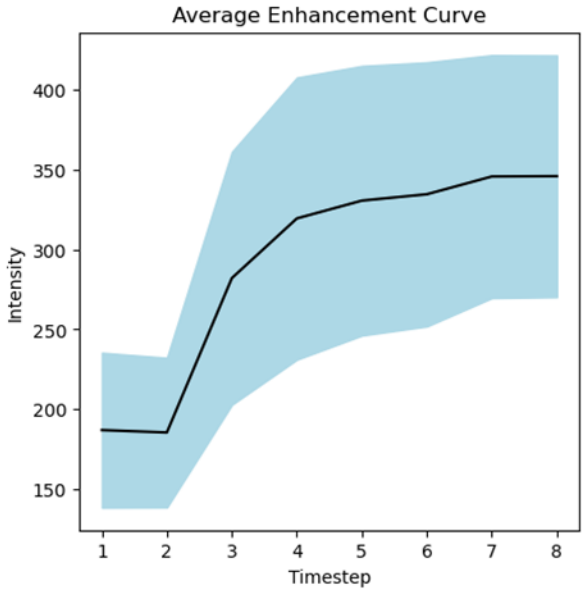


Figure 3.22: average enhancement curve related to figure 3.15, clearly showing a Plateau behaviour.

Inspecting the suspect area further, however, reveals a non-homogeneous behaviour. Figure 3.17 shows the pixel position and relative enhancement curves of the same area. As it is clearly shown all three enhancement behaviours would be correctly classified.

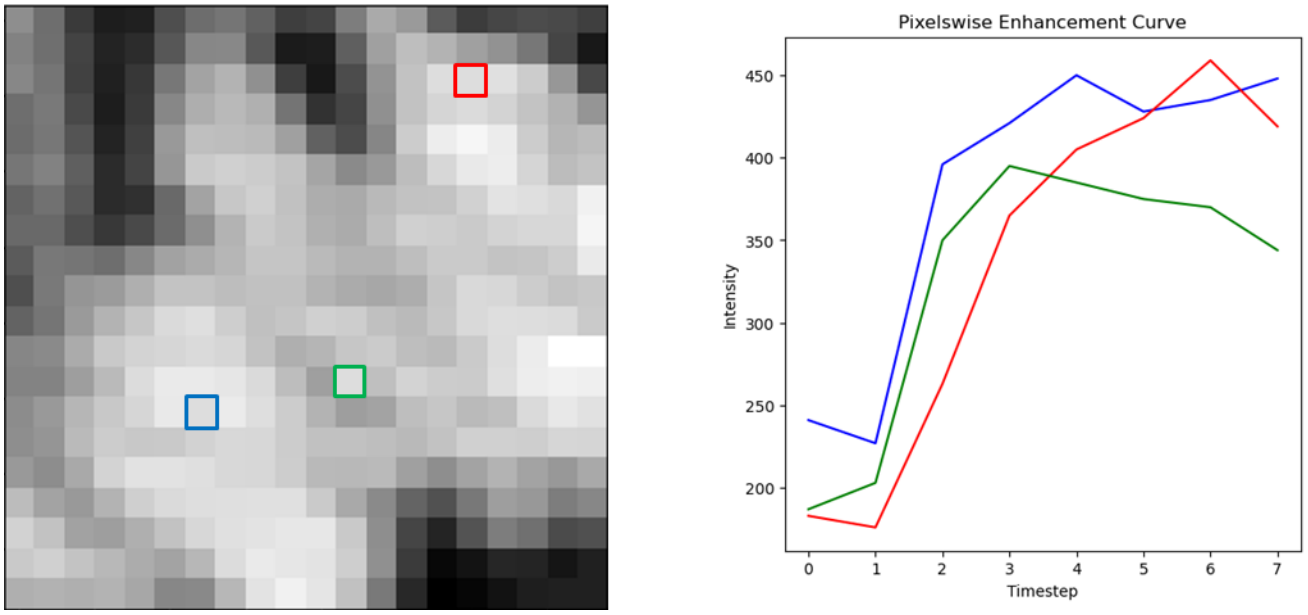


Figure 3.23: location and relative enhancement curves for the lesion presented in Figure 3.15. It is important to note their divergent behaviour.

This divergent behaviour is common in most lesions and renders the classification of enhancement behaviours a subjective task in which the result can vary with the size of the area that is taken into consideration during the diagnosis. An additional observation regarding enhancement curves is how these curves, typically associated with cancer, can appear in noncancerous areas. Figure 3.18 hereafter shows the enhancement curve of the top of the heart region in the slice featured in Figure 3.17. In a vacuum, the enhancement behaviour could be classified as plateau or washout, when this is a normal behaviour for the heart. This caveat is rarely mentioned in medical literature, as a radiologist would never analyse enhancement curves of the healthy heart. A fully automatic diagnostic system would need to be informed on the morphology of a human body, allowing for a selection of voxels to analyse.

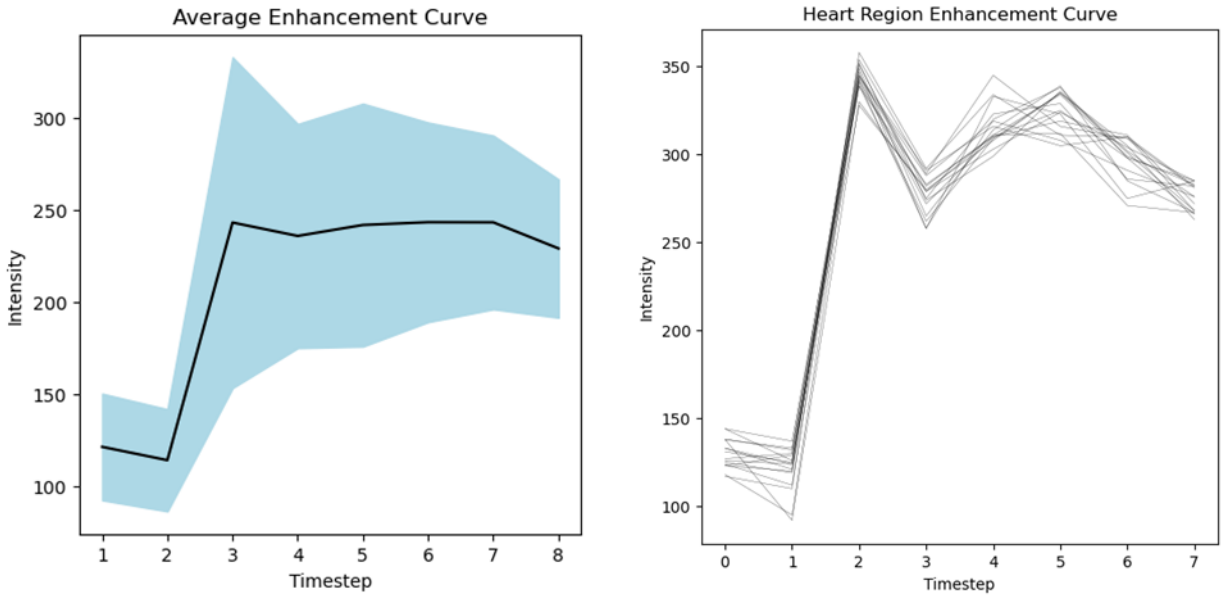
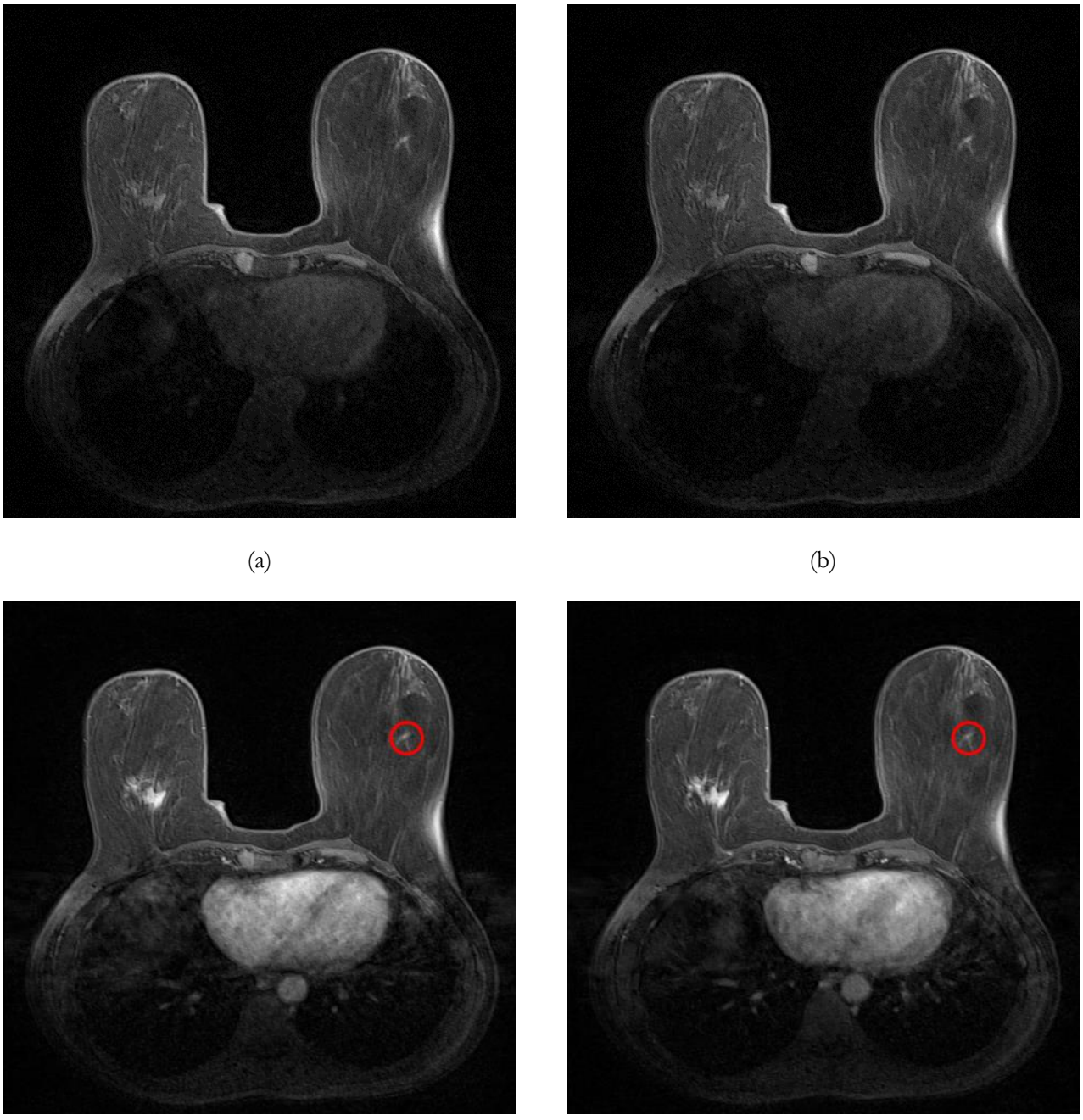


Figure 3.24: average enhancement curve of the heart region related to figure 3.15.

3.2.3 Notable exclusions and Diagnostic Ambiguity

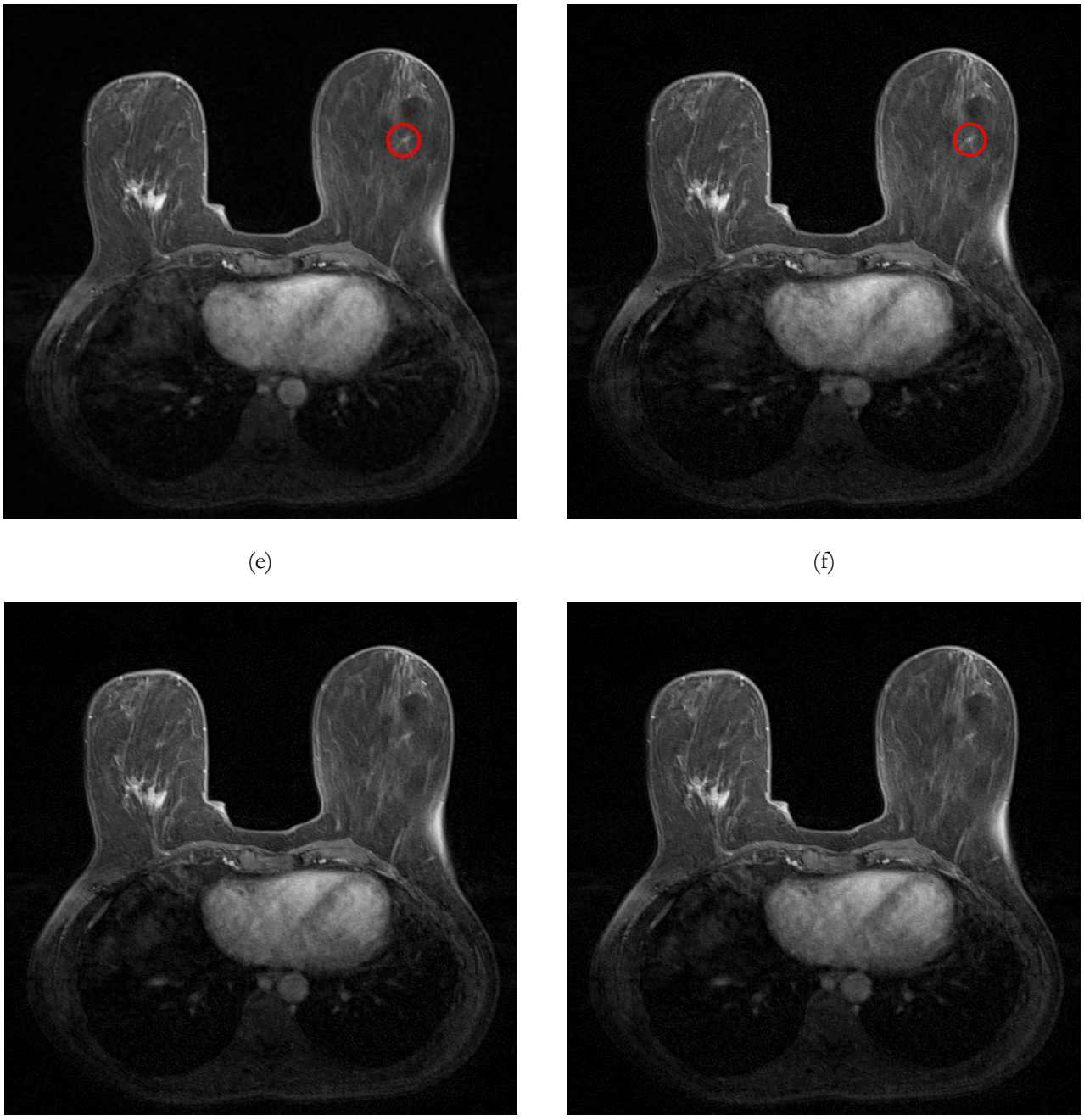
Radiologists often prioritize lesions larger than 5 mm in breast cancer diagnosis using DCE-MRI for several reasons. Smaller lesions, typically under 5 mm, can be more challenging to detect and characterize due to their size and the intricate breast tissue. Additionally, these tiny lesions may sometimes represent benign findings or artifacts, leading to unnecessary anxiety and further testing for the patient if they were closely examined. Radiologists focus on larger lesions as they are more likely to be clinically significant and are often better defined in DCE MRI images, allowing for a more accurate and confident diagnosis, which is crucial for patient care and treatment planning. An example of such exception is showcased in Figure 3.19, where the scan features a prominent lesion in the left breast, and a small lesion in the right breast. The report excluded the latter lesion; however, the behaviour of the area would suggest a suspicious lesion.



(a) (b)

(c) (d)

Figure 3.25: first batch of 4 timesteps from a sequence showing a lesion that was excluded due to size. The red circles in (c) and (d) highlight an area that could be classified as suspicious if it was larger than 5 mm.



(e) (f)

(g) (h)

Figure 3.19: second batch of 4 timesteps from a sequence showing a lesion that was excluded due to size. The red circles in (e) and (f) highlight an area that could be classified as suspicious if it was larger than 5 mm.

The figure above synthesizes the challenges with using the radiologists' reports as ground truth: the labelling in the report reflects the perception of an expert, derived from a process that is optimized for patient care and time effectiveness. Neither of these priorities allow for an objective labelling or areas as "cancer" or "not cancer", as small lesions could be a part of the "not cancer" areas.

An additional challenge in using the report data as ground truth comes from the diagnostic ambiguity regarding the dimensions of lesions. Given the low resolution of MRI scans, the decision on the extent of a lesion is made on the perception of a doctor, with subpixel accuracy. These decisions, however, are mostly to establish the need of a biopsy, hence the lesions are not measured with absolute and objective accuracy.

3.3 Chapter Summary

The section presented the challenge in diagnosing breast cancer from a clinician's perspective, and how the output of the process can be used in algorithmic ways. In essence, the processes are largely subjective, and the ground truth data cannot be used in algorithms that do not consider the anatomy of the human body. Henceforth, an unsupervised methodology is recommended for lesion detection, with knowledge integration supplied by supervised techniques. In the following chapter, the constraints in the data are leveraged in the development of an unsupervised methodology for lesion detection.

Chapter 4. Automatic suspect area identification through unsupervised deep learning

4.1 Introduction

Breast cancer diagnosis through dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has become a cornerstone in the detection and assessment of suspicious lesions. However, existing methodologies for lesion detection often face critical limitations, including reliance on supervised techniques, dependence on annotated datasets, and challenges in scalability. This chapter presents a novel, unsupervised deep learning approach aimed at addressing these limitations by facilitating diagnostic support without the need for labelled training data. The proposed methodology focuses on compressing temporal information from DCE-MRI sequences into a reduced dimensional space that is both interpretable and computationally efficient, enabling faster and more accurate identification of suspect lesions.

The primary goal of this research is to develop a methodology that highlights suspicious regions in DCE-MRI scans through an unsupervised neural network framework. Unlike prior approaches that depend on manually designed features and exhaustive search algorithms, this work leverages an encoder-decoder architecture to process temporal enhancement curves, distilling the data into a few interpretable dimensions. This dimensionality reduction not only preserves meaningful information but also simplifies the identification of abnormal enhancement patterns indicative of malignancies.

A key innovation of this study lies in the formulation of an "index of suspiciousness," derived from the low-dimensional representations. This index enables clinicians to prioritize regions for further examination while mitigating the impact of false positives. Additionally, the methodology is adaptable to diverse imaging protocols and hardware configurations, ensuring broader applicability in clinical settings.

Given the challenges associated with ground truth labelling in medical imaging, the proposed approach circumvents the need for extensive annotations by treating lesion detection as an anomaly detection problem. The algorithm identifies outliers based on reconstruction errors, leveraging patterns that deviate from typical enhancement behaviour. This unsupervised framework is designed to handle the high dimensionality of DCE-MRI data efficiently, making it scalable to large datasets without sacrificing performance.

The subsequent sections of this chapter detail the dataset preparation, network architectures, and experimental setups employed in this study. Comparative analyses with existing methods highlight the strengths of the proposed approach, particularly in terms of reconstruction error and true positive rates. The results underscore the potential of this methodology to streamline breast cancer diagnostics and improve clinical outcomes.

By introducing a framework that prioritizes interpretability, scalability, and flexibility, this chapter aims to bridge the gap between algorithmic performance and clinical usability. The findings presented here pave the way for future advancements in unsupervised diagnostic tools, offering a robust foundation for enhancing breast cancer detection through automated imaging analysis.

4.2 Dataset and labelling strategy

The whole dataset of 19200 sequences was used in this study, with a total of over 153000 images. The data collection took place on a 1.5T scanner (MAGNETOM Avanto, Siemens Healthcare GmbH, Erlangen, Germany), with patients lying face down. Each scan utilized a TR/TE/flip angle of 4.33s/1.32s/10°, employing a slice thickness of 1.1 mm with no inter-slice gaps. The resolution per slice stood at 448 x 448 pixels. The DCE-MRI protocol encompassed a pre-contrast T1-weighted sequence and seven post-contrast T1-weighted sequences, separated by intervals of 1:01 minutes.

The selected methodology was unsupervised, as ground truth labelling would not be possible, as mentioned in Chapter 3. The labelling provided in the attachments to the dataset was used to subjectively gauge at the effectiveness of the methodology a posteriori, however it was not used in the training process.

The sequence data (4-dimensional: x, y, z axis and time) was collapsed in two dimensions by stacking the height, depth, and length dimensions into one. The resulting dataset was in tabular form, with records detailing the transient behaviour of each pixel at each point in the breast volume. A simple thresholding algorithm was applied to allow compliance with relatively limited computational resources, shown in figure 4.1. All data from a single MRI scan was standardized (mean equal to 0 and standard deviation equal to 1), then for each slice, all pixels with value lower than 0 at either timestep 1 or timestep 3 were filtered out. This approach is meant to save computational resources and has no impact on the outcome of the solution.

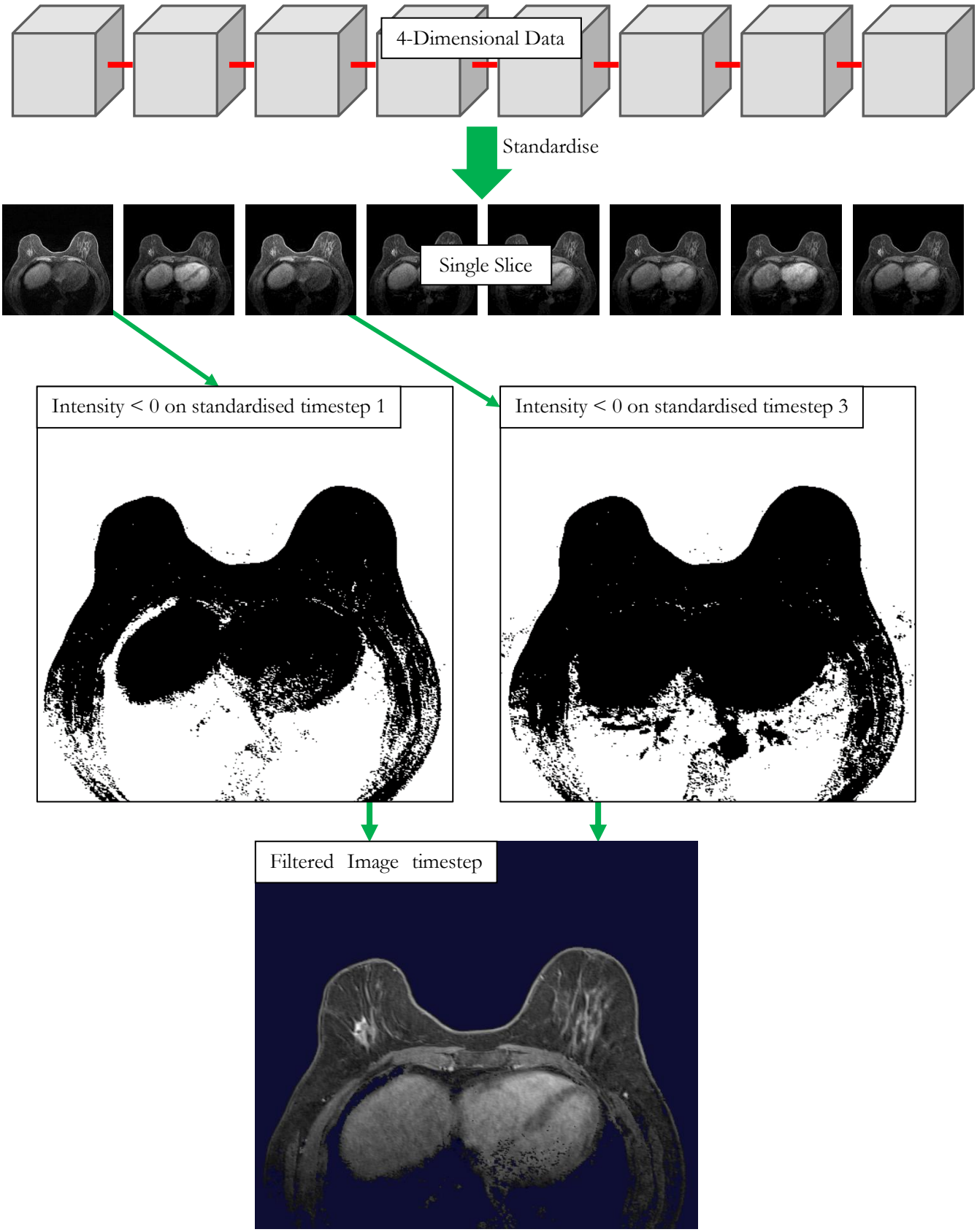


Figure 4.1: thresholding algorithm.

4.3 Related Work

Methods for detecting breast lesions from DCE-MRI scans often rely on exhaustive search techniques and manually crafted features ([51 – 54]). Vignati et al. ([53]) introduced a method that involves thresholding an intensity-normalized DCE-MRI to identify voxel candidates, which are then combined to form lesion candidates. The classification process utilizes hand-designed region and kinetic features. However, the method exhibits low accuracy, attributed to its reliance on assumptions about DCE-MRI intensity while neglecting texture, shape, location, and size features, achieving a True Positive Rate of 0.89 on a largely uniform dataset. Renz et al. ([54]) expanded on Vignati et al.'s work by incorporating additional hand-designed morphological and dynamical features, demonstrating improved results, with TPR of 0.96. Gubern-Merida et al. ([51]) achieved further enhancements by introducing hand-designed shape and appearance features, with a TPR of 0.89 on a significantly harder dataset, as it featured extreme variability in image quality, and lesion sizes and shape characteristics.

McClymont et al. ([52]) improved upon these methods by introducing unsupervised voxel clustering for initial lesion candidate detection, followed by a structured output learning approach that simultaneously detects and segments lesions. While this approach significantly enhances detection accuracy, it comes at a notable increase in computational cost, but achieved TPR of 1.00 in a small and varied dataset. The multi-scale deep learning cascade approach addressed runtime complexity concerns, enabling the extraction of optimal and efficient features while maintaining competitive detection accuracy, with a TPR of 0.80 on a small dataset (117 scans).

The recent research has been focused on reduction of inference time and false positive rate (FPR), such as [56], in which a deep reinforcement learning algorithm was trained to optimize the resources and restrict the potential locations of lesions to subsets of the images, achieving a TPR of 0.80 on a dataset closer to real-world when compared to the ones from previous approaches.

A critical aspect of previously proposed approaches involves two key issues: the absence of a standardized dataset for evaluating diverse methodologies and the lack of a consistent lesion detection criterion. In [53, 54], detections were visually inspected by a radiologist, while [51, 52] considered a lesion detected if a single voxel in the ground truth was identified. This study followed the methodology in [53, 54], with the addition of not considering false positives as part of the performance indicators, as presented in Chapter 3.

The absence of a standardized dataset also means that the biases that might be present in each dataset are grossly overlooked, hence any comparison would lack of scientific objectivity.

Moreover, there is no example in literature in which an index of suspiciousness of an area has been proposed, which would greatly reduce the impact of false positives in the diagnostic process, as well as reducing the likelihood of false negatives.

4.4 Methodology

4.4.1 Problem Definition

Given the challenges in data coherence and labelling outlined in Chapter 3, the chosen methodology for this study was an unsupervised one. Suspect regions in a DCE-MRI scan are identifiable as outliers or anomalies. This approach, which to the author's knowledge does not appear in literature, allows for annotation-agnostic highlighting of suspect areas.

The signal obtained by observing the intensity of a single pixel over time is inputted in a neural network encoder, which distils the information within 8 timesteps into n dimensions, with n being a small number (maximum 3). During training, the data is then decoded back to its original dimension. The reconstruction error of this process is defined as the MSE between the output of the decoder and the input and is the error function used to train the model. The distilled dimensions are processed via algorithmic implementations of heuristics derived from observation of the resulting data into an index which correlates with the likelihood of a pixel being part of a cancerous region. It is of paramount importance to note that a distance-based algorithm (i.e. k-nearest neighbours) would not work in this specific dataset, as all data features suspect lesions, hence a scan from a healthy patient would likely result in numerous false positives.

A summary of the methodology is available in Figure 4.2 and Figure 4.3. A schematic of the training process for a single datapoint is shown in Figure 4.4.

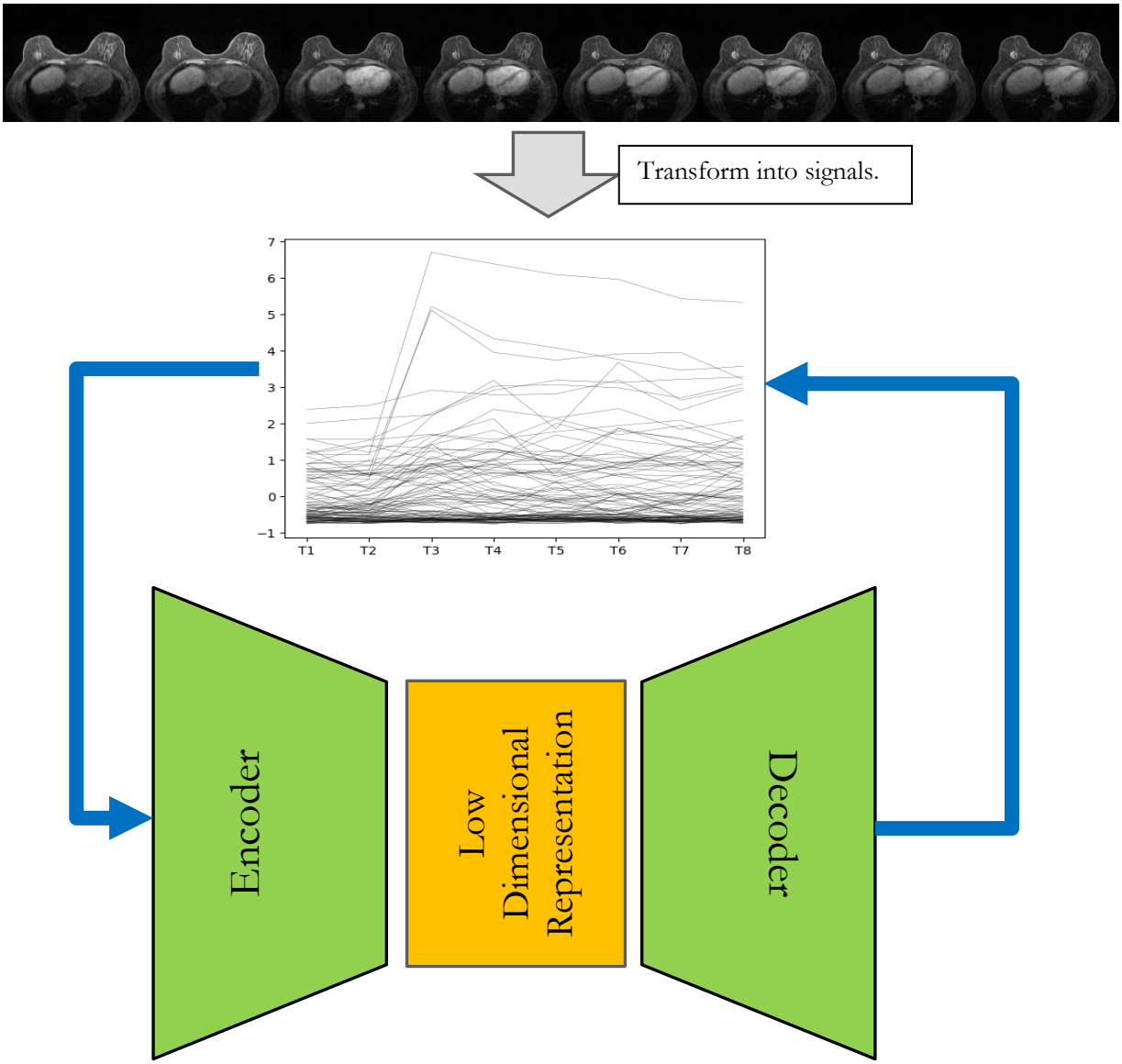


Figure 4.2: summary of methodology

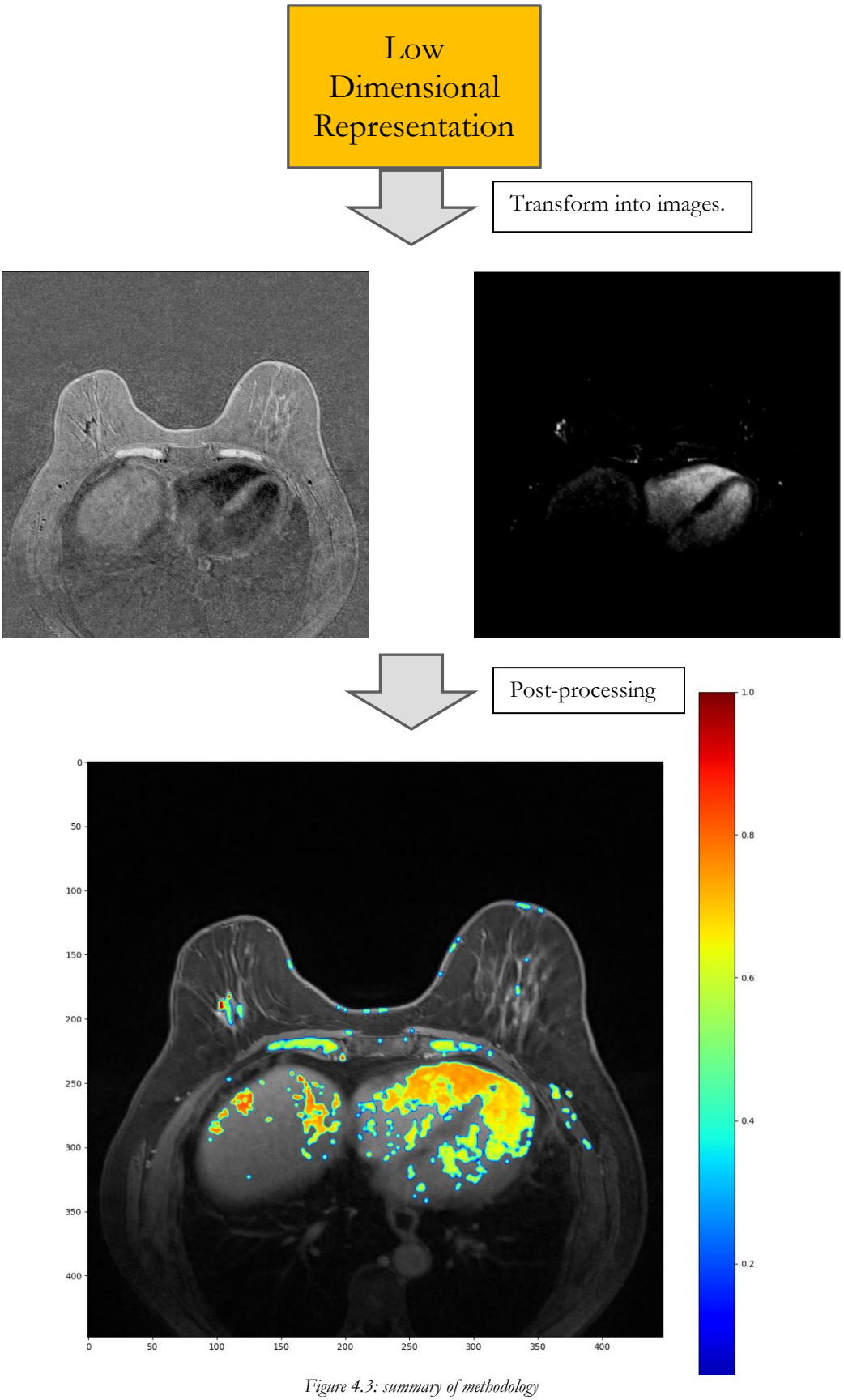


Figure 4.3: summary of methodology

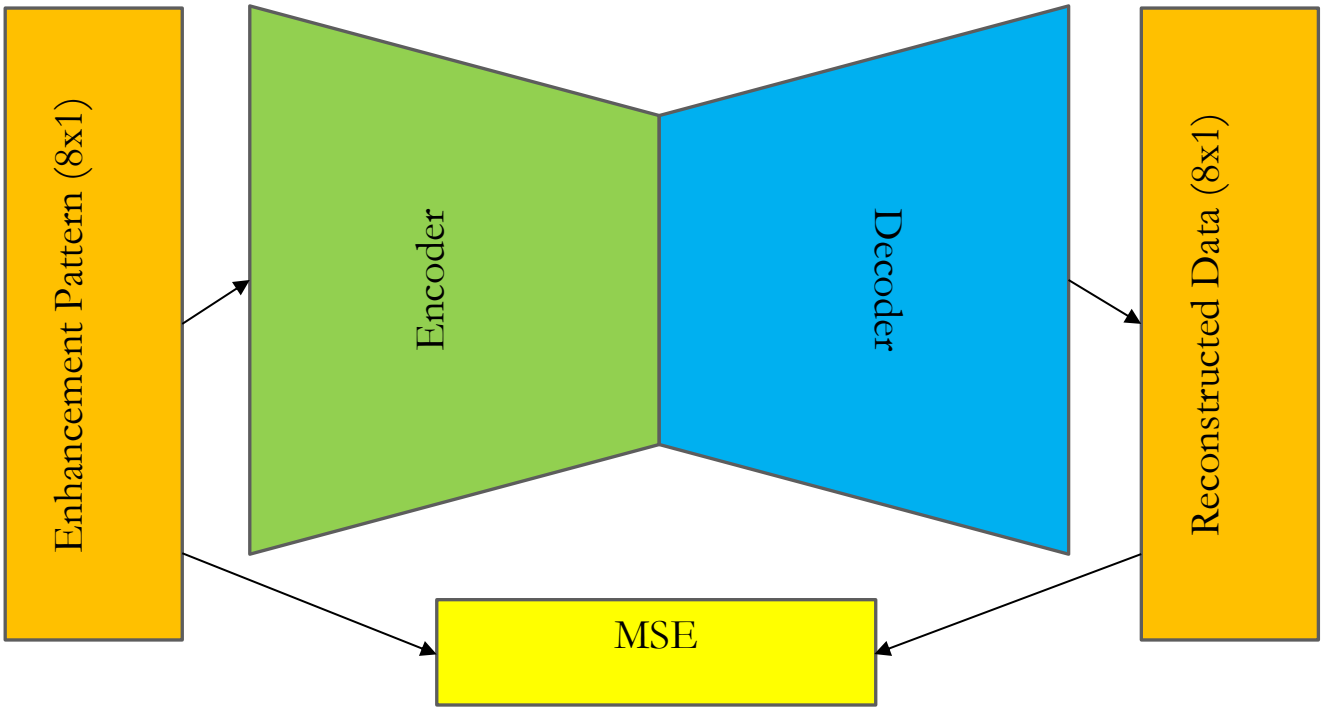


Figure 4.4: training process for the proposed Encoder-Decoder structure.

The data is inputted in the Encoder-Decoder architecture, which outputs an array of same dimensions. The arrays are confronted with MSE. The training aims to minimise this error, also known as reconstruction error. As the data from the labelling provides areas in which there is certainty of the presence of a suspect lesion, but there is no assurance on whether a single pixel is part of a lesion or not, the metric to determine the success of the methodology was chosen on overall accuracy of detection, with disregard of whether other areas highlighted were false positives or true positives. Part of further work on this data would be to improve upon the metric chosen in this project, as the only viable evaluation methodology for this study was a manual one.

This phase of the work aimed to develop an algorithm that is inherently incapable of modelling the entirety of the possible enhancement patterns. This results in activations corresponding to the anomalous enhancement patterns to be extreme in the low-dimensional layers, allowing for easier and granular estimation of the abnormality of any given datapoint. The ability to map the enhancement space in lower dimension is highly desired, hence a lower reconstruction MSE was found to correlate well with lesion detection.

4.4.2 Neural Network Architectures

The study focused on determining the best deep learning architecture in detecting the confirmed lesions in the dataset. The key factor in the quality of the outcome is the encoder architecture, hence, the decoder architecture was set to be mirroring the one of the encoders. Given the low dimensionality of the input data, a fully connected, feed forward architecture was chosen for the neural networks.

An encoder architecture is a function of depth and width of the network, as well as the dimensionality of the low-dimensional space. Therefore, the investigation focused on the effects of scaling these parameters on the prediction ability of the methodology. The chosen architectures, then, were combinations of different widths, depths, and bottleneck connections. Given the large number of combinations of these variables, a discretization of the search space was performed, illustrated in the following section.

4.4.3 Experimental Setup

The encoder search space for width of the layers was discretised by dividing it in two categories of size, expressed in the number P, which could assume the value of 2 or 3.

Depth, width, and low dimensionality were ablated to obtain an optimal size and shape of the network. To this end, two depth levels were chosen, as a function parameter numbers. The shallow depth variant would have $P + 1$ layers, while the deep variation would have $2P + 1$ layers.

This process was similarly conducted for width, which features a slim variant, with maximum width of 2^{2P+1}, and a large variation with maximum width of 2^{3P+1}. The maximum width was always assigned to the first layer, and any subsequent hidden was assigned a width equal to half of the one for the previous layer. As an example, if layer 1 had a width of 64, layer 2 would have a width of 32, layer 3 would have a width of 16, and so on. The minimum width was fixed at 4, hence a layer with width of 4 would have to feed into the low-dimensional layer. This only applies to the encoder/decoder structure, and not to the low-dimensional layer.

The low-dimensional layer was assigned depths of 3 and 2 for every encoder/decoder architecture. It is important to note that these dimensions were later combined into one via hand-crafted algorithms to the best of the researcher's ability, hence the results of study should not be considered as reliable in terms of dimensionality of low-dimensional representation size.

All models were trained to convergence five times, and the results reported represent the best performing model. The architectures covered in this study are presented in Table 4.1 below.

Table 4.1: deep learning architectures overview.

Name	Maximum Width	Depth	Low Dimensionality
P2_Low_Large_3	128	3	3
P2_Low_Large_2	128	3	2
P2_Low_Slim_3	32	3	3
P2_Low_Slim_2	32	3	2
P2_High_Large_3	128	6	3
P2_High_Large_2	128	6	2
P2_High_Slim_3	Not tested due to insufficient dimensions – would be equal to Low_Slim_3		
P2_High_Slim_2	Not tested due to insufficient dimensions – would be equal to Low_Slim_2		
P3_Low_Large_3	1024	4	3
P3_Low_Large_2	1024	4	2
P3_Low_Slim_3	128	4	3
P3_Low_Slim_2	128	4	2
P3_High_Large_3	1024	8	3
P3_High_Large_2	1024	8	2
P3_High_Slim_3	Not tested due to insufficient dimensions – would be equal to P2_High_Large_3		
P3_High_Slim_2	Not tested due to insufficient dimensions – would be equal to P2_High_Large_2		

4.4.4 Visual Evaluation

The performance of the models was evaluated through a visual inspection of results on all slices in which a lesion was confirmed to be present by an expert radiologist. The encoded data (2 or 3-dimensonal) was transformed in 1-dimensional data and reconstructed in a heatmap to be superimposed on the original image. Various methodologies for reducing the encoded data into a single dimension were tested, however it was observed that selecting the maximum absolute value among the activations yielded the best results in suspect area detection. Afterwards, a thresholding

algorithm was applied to remove low activation values. Finally, the values were normalised, and an expansion-erosion algorithm was applied to reduce the noise in the images.

4.5 Experiments

The unsupervised deep learning methodology for automatic suspect area identification in breast DCE-MRI scans was implemented and evaluated using a dataset consisting of 19,200 sequences and over 153,000 images. The methodology compressed the temporal information of eight timesteps into a limited number of dimensions for enhanced human interpretability, facilitating faster lesion identification. The results were assessed based on the ability of the models to reconstruct the input signals from the encoded space. Therefore, the chosen performance metric was MSE between the input and outputs of the network. The performance was later evaluated visually on the ability to detect cancerous lesions, and the TPR metric is reported.

The best results for each model presented in Table 4.1 above are shown in Table 4.2 below.

Table 4.2: reconstruction Error (MSE) and True positive rate (TPR) overview for all examined architectures.

Name	Reconstruction Error (MSE)	True positive rate (TPR)
P2_Low_Large_3	0.0945	0.7881
P2_Low_Large_2	0.1466	0.6415
P2_Low_Slim_3	0.1198	0.7089
P2_Low_Slim_2	0.1478	0.6386
P2_High_Large_3	0.0746	0.8673
P2_High_Large_2	0.0830	0.8313
P3_Low_Large_3	**0.0534**	**0.9788**
P3_Low_Large_2	0.0717	0.8804
P3_Low_Slim_3	0.0845	0.8255
P3_Low_Slim_2	0.1566	0.6193
P3_High_Large_3	0.0559	0.9635
P3_High_Large_2	0.0766	0.8583

The results of the model evaluation, as presented in Table 4.2, provide valuable insights into the performance of various neural network architectures in the context of breast DCE-MRI suspect area identification. The reconstruction error, quantified by the Mean Squared Error (MSE), serves as an essential metric reflecting the fidelity of the models in reconstructing the original data. Lower MSE values indicate better reconstruction performance. Notably, architectures such as P3_Low_Large_3 and P2_High_Large_3 exhibit the lowest reconstruction errors, with MSE values of 0.0534 and 0.0746, respectively. These models effectively compress the temporal information into a low-dimensional space while maintaining a high degree of accuracy in reproducing the original data. Simultaneously, it is possible to observe how the True Positive Rate (TPR) is highly correlated to the ability of a network to reconstruct the input data, hence the best performing models are P3_Low_Large_3 and P2_High_Large_3, with TPR values of 0.9788 and 0.9635, respectively.

It is worth noting how the performance scales with width, depth, and bottleneck dimensionality of the networks. Analysing the performance of networks with equal width but different depth, such as P2_High_Large_3 and P2_Low_Large_3, yields an average improvement on MSE of 23.11%, which translates to an average TPR increase of 7.47%. A similar analysis results in a 41.07% decrease in MSE corresponding to a 15.08% increase in TPR for bottleneck dimensionality, and a 51.06% decrease in MSE with a 16.22% increase in TPR for width scaling. These improvements are phenomena with clear diminishing returns, and at higher widths and depths, detrimental, as it is possible to observe a decrease in MSE of 6.40% when depth is scaled beyond the optimal level, likely due to overfitting.

4.5.1 Post-processing and Visual Results

The lower-dimensional space was processed to allow it to be human readable. First, the absolute value of the output of the bottleneck layer was used to reconstruct images by using the known pixel position of the input signal. An example of this process is shown in Figure 4.5 below, which showcases the best performing P3_Low_Large_3 model on a known lesion.

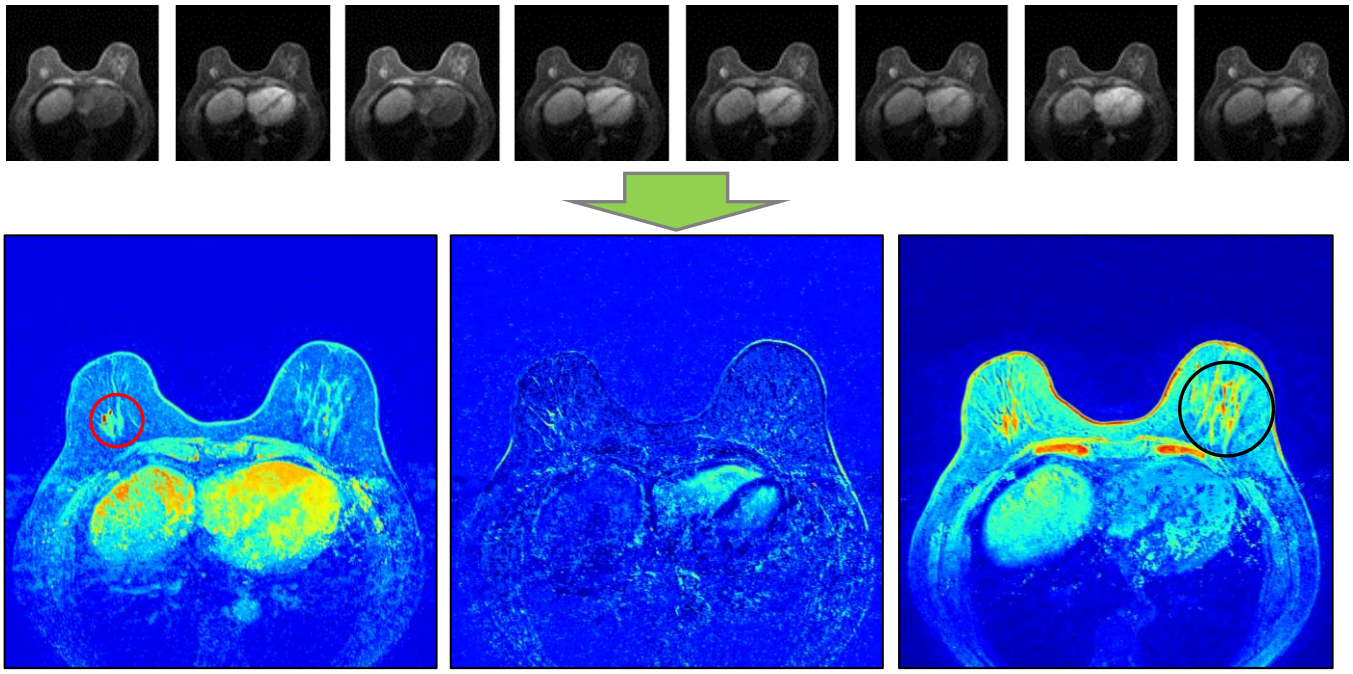


Figure 4.5: output of the low-dimensional layer for the best performing model (P3_Low_Large_3). The expert-confirmed lesion (ground truth) is circled in red on the left image. It is possible to note, however, high activations in the area circled in black on the rightmost image. These were later confirmed by the radiologist to be non-mass enhancement in early formation, further proving how reliance on ground truth diagnosis could lead to flawed algorithms.

The lower-dimensional images were then merged into one by taking the maximum value at each position. The resulting image was then filtered using a user-set threshold (mimicking the clinical use and offering customisability to the user) and superimposed onto one of the original images. All evaluations reported in Table 4.2 above were made by setting the threshold to $\frac{V_{max}}{3}$, where V_{max} was the maximum value in a whole scan. Finally, the image was inputted in an erosion-dilation algorithm to remove noise, as shown in Figure 4.6 below. The resulting image is normalised for the whole scan and presented to the user, which can conveniently refer to a score between 0 and 1 to gauge at the likelihood of an anomaly in any given area.

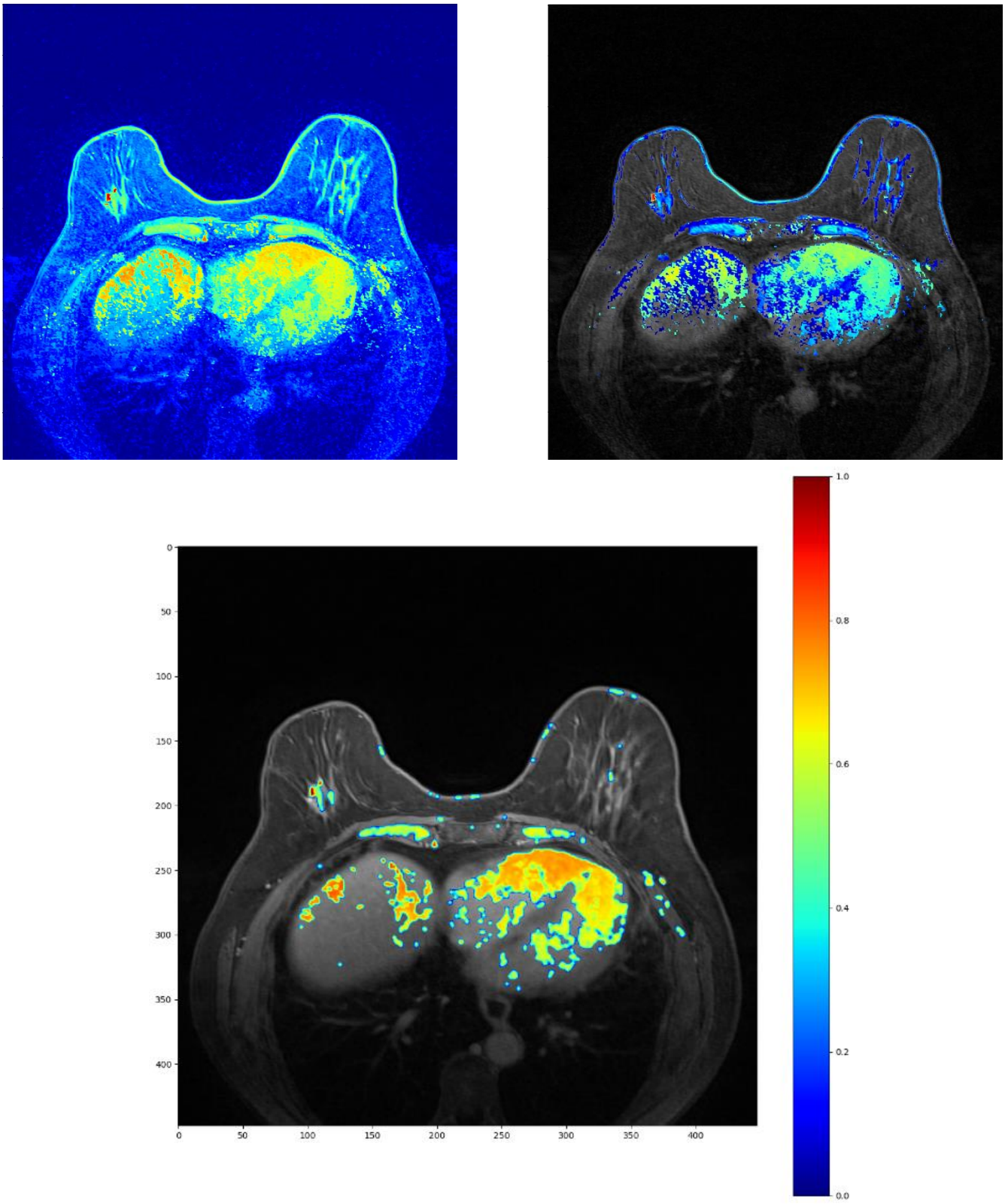


Figure 4.6: outputs of the post-processing step for the low-dimensional layer for the best performing model (P3_Low_Large_3).

4.6 Discussion

The proposed unsupervised methodology demonstrated considerable advantages over existing approaches that heavily rely on supervised techniques. First and foremost, this approach is the first in considering real-world clinical applications and usage of the methodology, offering the user a clear and linear way to interact with the results of the algorithms and adjust them to maximise the utility of them. Secondly, the proposed approach maximises interpretability by operating on enhancement curves, the same data as the practitioner during diagnosis, allowing the user to immediately and intuitively verify whether a suspect area is, in fact, worth considering. At the same time, operating on pure enhancement curves allows for fundamental transferability of this methodology to different imaging protocols and data from different MRI manufacturers. Moreover, the unsupervised nature of the solution allows for better scalability by reducing the need for labelled data, which is often biased, whilst maintaining similar or better performance levels compared to the state of the art. Finally, the methodology introduces an index of suspiciousness, offering a novel perspective on lesion detection in breast DCE-MRI scans and increases trustworthiness.

The best performing model out of the ones examined was P3_Low_Large_3, however there are still open questions regarding performance scalability of width, along with the behaviour on larger datasets.

Acknowledging the study's promising results, certain limitations were identified. The manual evaluation methodology used in this study requires refinement, and future work will explore alternative metrics for better assessing the performance of the models. The algorithm tends to identify the region inside the chest wall as suspicious, due to the behaviour of the heart (as shown in chapter 3). This limitation, while not detrimental for the accuracy of lesion detection, significantly impacts the clinical useability of the methodology by introducing noise in the algorithmic results.

4.7 Conclusion

The study successfully developed an unsupervised deep learning methodology for automatic suspect area identification in breast DCE-MRI scans. The detailed evaluation of various neural network architectures demonstrated the potential for improved diagnostic efficiency. The proposed approach showcased competitive results compared to existing techniques, highlighting its capability for annotation-agnostic detection of suspect lesions.

In conclusion, the developed methodology holds promise in enhancing diagnostic capabilities and reducing dependence on ground truth labelling. The findings contribute to advancements in breast cancer diagnosis through

automated imaging analysis, paving the way for future research to refine evaluation metrics. Part of the limitations stemming from the noise generated in the heart region are addressed in Chapter 5.

Chapter 5. Automatic thoracic cavity segmentation in DCE breast MRI using deep convolutional neural networks

5.1 Introduction

The segmentation of the thoracic cavity in dynamic contrast-enhanced (DCE) breast MRI is a fundamental task in medical imaging that plays a crucial role in the development of computer-aided diagnostic (CAD) systems. Accurate segmentation is essential for isolating the region of interest, enabling the detection and analysis of pathological structures while minimizing computational overhead associated with irrelevant anatomical regions. Despite the advancements in medical image processing, achieving high-precision, automated segmentation remains a challenging task due to variability in anatomical structures, imaging artifacts, and limited availability of annotated datasets.

Deep learning-based methodologies have emerged as powerful tools for automating segmentation tasks by leveraging neural networks trained on manually annotated data. These approaches have demonstrated superior performance compared to traditional image processing techniques, achieving state-of-the-art results in various medical imaging applications. However, challenges remain in ensuring interpretability and generalizability of these methods, as well as addressing their dependency on large, well-annotated datasets and significant computational resources. These requirements present a barrier to widespread adoption, particularly in clinical environments with limited resources. for automating segmentation tasks by leveraging neural networks trained on manually annotated data. These approaches have demonstrated superior performance compared to traditional image processing techniques, achieving state-of-the-art results in various medical imaging applications. However, the success of deep learning algorithms often depends on large, well-annotated datasets and significant computational resources. These requirements present a barrier to widespread adoption, particularly in clinical environments with limited resources.

In this chapter, a novel approach to thoracic cavity segmentation in DCE breast MRI using deep convolutional neural networks is presented, specifically focusing on addressing the challenges of data efficiency and computational feasibility. Our methodology employs a Dynamic UNet architecture with a pre-trained ResNet encoder to leverage transfer learning and mitigate the impact of limited training data. Additionally, architectural enhancements are incorporated, including self-attention mechanisms, blurring layers to reduce checkerboard artifacts, and bottleneck connections to preserve spatial information.

This work aims to bridge the gap between segmentation accuracy and practical implementation by optimizing network configurations to operate effectively with limited data and computational resources. By evaluating multiple architectural variations, the configurations that deliver high segmentation accuracy while maintaining computational efficiency suitable for clinical environments were selected. The proposed approach demonstrates robust performance, achieving results comparable to or exceeding existing state-of-the-art methods, and offers a scalable solution for CAD applications in breast MRI.

The remainder of this chapter outlines the methodologies employed, including dataset preparation, network architecture, and training strategies. An extensive evaluation of model performance based on segmentation accuracy, computational efficiency, and robustness to data augmentation strategies is also presented. Finally, a discussion on the implications of our findings and potential avenues for future research in medical image segmentation is presented.

5.2 Related work in thoracic cavity segmentation

The current state of the art addresses the thoracic cavity segmentation challenge by training a 3-dimensional cluster of 2D UNets ([57]). It is argued that implementing more recent deep learning techniques can lead to better results and generalisation performance.

Given the importance of the task and its complexity, there have been numerous approaches proposed in the published literature. Marrone et al. DL-based approaches ([58]) are also included in the following overview. While the DL category has limited representation right now, it is currently considered as the state of the art. Moreover, it is expected that the popularity of this subfield would increase due to the vast representation of DL-based techniques in parallel fields, such as hand and brain segmentation ([59]). For these reasons, the section dedicated to DL approaches is being given higher importance.

5.2.1 Pixel-based approaches

Approaches in this category rely on classifying pixels or voxels individually, or with simple computations on the surrounding pixels ([60, 61]). Results are not always fully automatic and tend to produce suboptimal results, especially the boundary between the sternum and the internal organ, which is often wrongly segmented. On the other hand, they require minimal computational costs. For example, in the approach by Vignati et al. ([62]), the images are processed

using Otsu's thresholding and a sequence of dilations and erosions. Results show good breast parenchyma segmentation performance. However, the limitations become apparent as the examples provided show imprecise chest wall segmentation, as specified by the authors, and require the aid of fat-saturated images, or an atlas-based segmentation. The study demonstrated that segmentation of the outer boundary of the breast is achievable by computationally efficient methodologies, while chest cavity segmentation would require more complex solutions. Given the importance of minimising the presence of internal organs (see Figure 1.6), specifically the heart, is fundamental in the development of lesion detection systems for DCE-MRI. Future studies should then focus on the detection of the chest wall.

5.2.2 Atlas-based approaches

The solutions in this category are generated by comparing anatomical atlas generated by manually segmented data ([15, 63-66]). The approaches usually require a high number of instances in the atlas to guarantee generalisability to different anatomical features and acquisition protocols. The size of the atlas, however, is directly linked to an increased computational cost. To counteract such limitations, the solutions are often restricted to a specific anatomical part. An example of such an application was provided by Fooladivanda et al. ([67]), in which the authors use an atlas-based approach to segment the pectoral muscle, relying on a simpler pixel-based approach to segment the chest wall edge.

5.2.3 Geometrical-based approaches

The proposed solutions within the category revolve around constraining the segmentation results to predetermined anatomical and physiological characteristics. Their most common usage is as a refinement to pixel-based approaches ([68, 69]). The main criticisms of the techniques are the extreme computational cost and extremely poor generalisation performance. Notable examples come from Wu et al. ([23]) in which the authors propose a methodology to extract the chest wall line from sagittal breast MRI. Results are based on a refinement of edge detection via enforcing geometrical constraints and are extremely effective. The fundamental assumptions, however, highlight the limitations in generalisability, as the methodology does not account for the edge detection algorithm failing. This limits the potential application in a contrast-enhanced MRI environment, as post-enhancement images usually feature high-intensity chest wall and heart regions, as shown in the following figure.

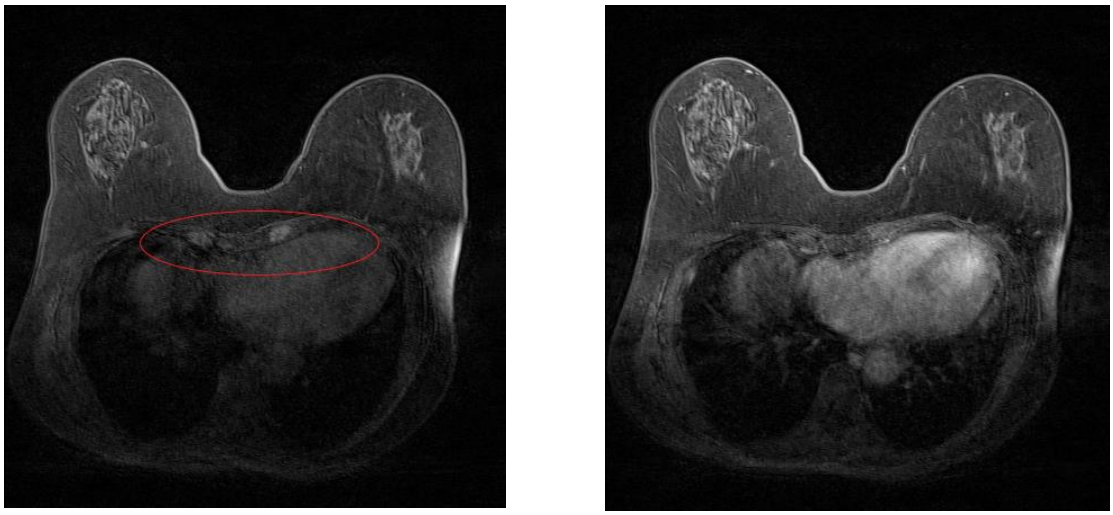


Figure 5.1: pre-enhancement and post-enhancement (6 minutes after CA injection) images of a central slice of the breast. The area of the sternum highlighted in the picture on the left is not clearly defined through an edge detection algorithm in the picture on the right.

5.2.4 Deep learning-based approaches

Deep learning-based solutions allow for the creation of fully automated segmentation methodologies by exploiting previous examples in the form of manually segmented data. The limitations of the approaches consist in the training process, for which large quantities of data and computational resources are needed. The techniques have been applied with great success to an increasing amount of research problems in recent years ([70]).

The pioneering solution is attributable to Dalmis et al. ([71]), in which the authors propose an automated breast and fibro-glandular tissue (FGT) segmentation using a UNet approach ([72]), obtaining considerable improvements in accuracy over the state of the art at that time in terms of DSC Similarity Coefficient (DSC), achieving a DSC of 0.944 against 0.863 from the previous reported state of the art. The current state of the art is attributable to the previously mentioned work by Piantadosi et al., who have implemented a multi-planar UNet approach, obtaining some improvements over Dalmis et al. ([27]). The methodology uses three discrete UNets on the transverse, sagittal and coronal planes instead of a three-dimensional (3D) UNet to increase computational efficiency for the solution. The 3D aspect of the network, however, effectively triples the computational cost. Moreover, the current state of the art was trained on a dataset comprised of fully labelled data from 117 patients, well above the data usually available for training to most developers. As the labelling process is the most time-consuming aspect of the development, solutions

should aim at maximising performance with a minimum amount of data. To this end, novel techniques need to be employed, from both an architectural and a data pre-processing point of view.

Given the limitations of the approaches presented and their advantages, the convolutional neural network (CNN) methodologies are the most applicable for problems in computer-aided diagnostics (CAD). However, all DL approaches available in the literature fail to address the most limiting aspect of the issue, which is data availability. The proposed solution aims to improve the current state of the art by focussing on data efficiency, by employing architectures that are less prone to overfitting and behave better in transfer learning scenarios. Moreover, recent advancements in DL research, such as self-attention ([33]), have the potential to further improve upon the state of the art. In addition, computational efforts need to be compliant with hardware that is potentially available in a hospital; hence the algorithms should process patient data in a 2-dimensional fashion, as 3D data can be generated if needed.

The work performed in these years has contributed to the improving the current state of the art by employing several techniques that have been successful in parallel fields. To this end, a total of 18 configurations were evaluated, corresponding to the combination of 3 different architectures and 3 different techniques to be applied.

5.3 Datasets and labelling strategy

A dataset comprising breast DCE MRI data from 44 patients served as the foundation for both training and evaluating the proposed segmentation model. The data collection took place on a 1.5T scanner (MAGNETOM Avanto, Siemens Healthcare GmbH, Erlangen, Germany), with patients lying face down. Each scan utilized a TR/TE/flip angle of 4.33s/1.32s/10°, employing a slice thickness of 1.1 mm with no inter-slice gaps. The resolution per slice stood at 448 x 448 pixels. The DCE-MRI protocol encompassed a pre-contrast T1-weighted sequence and seven post-contrast T1-weighted sequences, separated by intervals of 1:01 minutes.

For manual segmentation, a subset of slices was judiciously chosen and validated by a breast radiologist. Before manual segmentation, slices underwent size cropping to remove the posterior thorax half. In this development phase, the lowermost row of 131 pixels was discarded, as it yielded the optimal cropping outcome for our dataset. To account for anatomical variations across the upper body, eleven evenly distributed slices within the central 70 slices of the pre-contrast sequence were meticulously selected for manual segmentation. To accommodate variations caused by contrast agent injection, the six central slices of the pre-contrast sequence were also manually segmented. The process was

replicated for these slices across all seven post-contrast sequences. Consequently, a total of 2552 slices underwent manual segmentation. From this pool, slices from 37 patients (n = 2146 slices) were allocated for model training, while the remaining seven patients' segmented slices (n = 406 slices) were dedicated for model testing (in summary: 81% of data was used for training, the remaining 19% for test).

5.4 Methodology

The automatic chest cavity segmentation algorithm proposed in this work is composed of two main parts. In the first part, the volume is sliced in its transverse plane images, and it is inputted in a Dynamic UNet (fast.ai, n.d.), inspired by the work of Iglovikov and Shvets ([73]). A schematic of the proposed solution is shown in Figure 5.2 below.

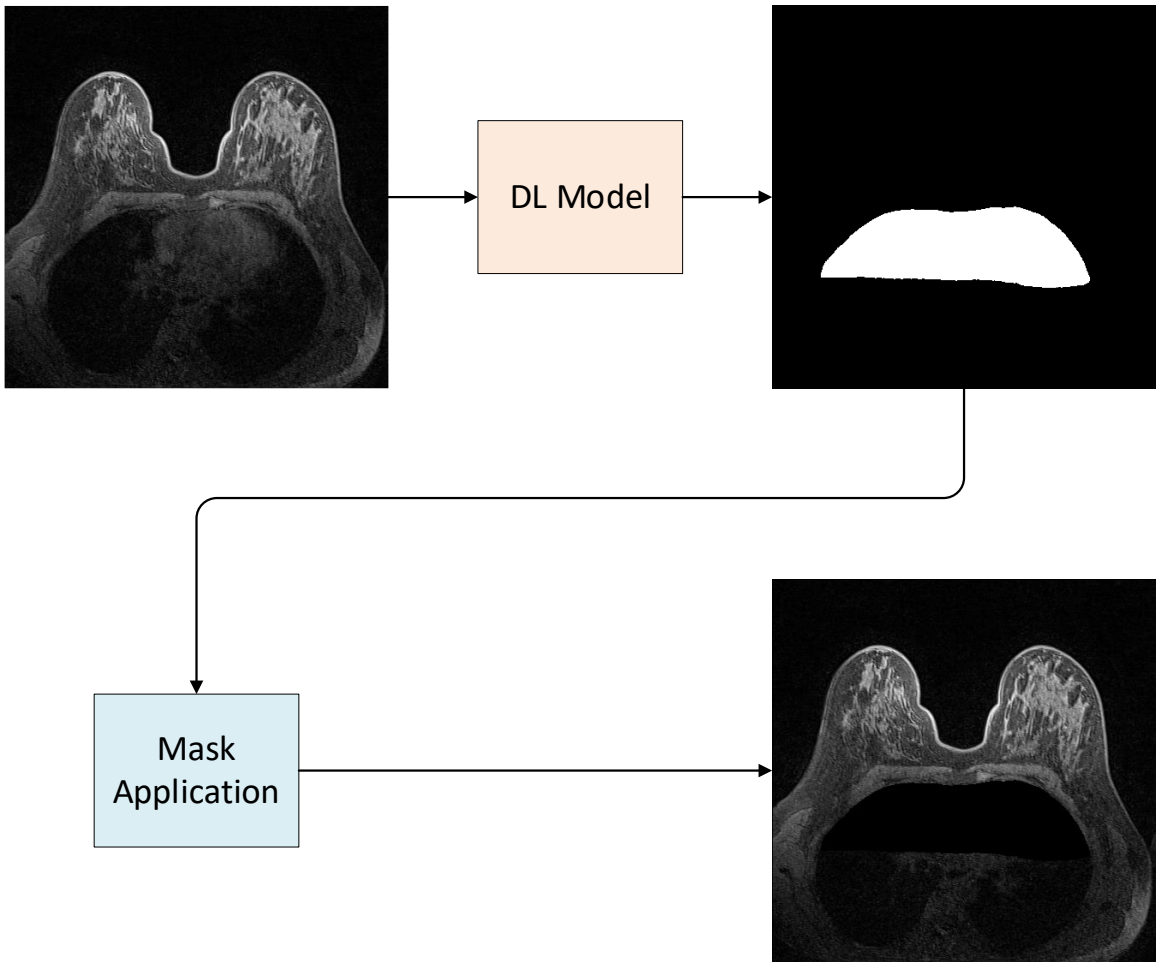


Figure 5.2: schematic of the proposed solution. The input image is inputted in a deep learning model, which outputs a generated mask. The segmentation mask is then applied to the input image.

This work proposes an automatic algorithm for chest cavity segmentation, designed to accurately delineate the chest cavity in medical imaging. The model processes 3D medical volumes by first slicing them into 2D transverse plane images. These images are then fed into a Dynamic UNet model, which performs the segmentation by generating a mask that highlights the chest cavity region.

5.4.1 Deep learning model

The segmentation model is based on a Dynamic UNet with a pre-trained ResNet encoder. A transfer learning approach was utilised with a pre-trained ResNet encoder to address the challenge of insufficient training data. Transfer learning ([74]) is a widely used technique in computer vision tasks outside of medical image segmentation ([75, 76]). To leverage transfer learning, the technique employed in the proposed solution is the Dynamic UNet (fast.ai, n.d.), based on the original Unet proposed by Ronneberger et al. ([26]). A UNet architecture was chosen as it is the best methodology in terms of overall performance in medical imaging applications ([24, 77]) The flexibility derived from using custom encoders provided by Dynamic UNet allows for a much greater degree of exploitation of transfer learning.

The original UNet architecture can be divided into an encoding part, or "down sampling", and a decoding part, or "up sampling". The encoding side performs a similar task to a conventional CNN, with regular down sampling steps, performed through the maxpool operation. At the same time, the decoding path follows a symmetrical structure, with the up-sampling steps performed through a fractionally-strided convolution layer. The symmetry allows for the activations of the down sampling layers to be concatenated to the activations of the up-sampling layers, thus better retaining spatial information throughout the up-sampling path.

The Dynamic UNet architecture was originally presented in the previously mentioned work by Iglovikov and Shvets. It follows the same basic architecture as the original UNet but adds a pre-trained model as the encoder. The approach not only achieves considerable improvements over the traditional UNet, but it also allows to experiment with a variety of pre-trained encoders. In this study, the ResNet encoders were used, as it has been demonstrated that they reduce the reliance on regularization techniques ([78]).

ResNets were first introduced in 2015 ([79]) to counteract the vanishing/exploding gradient problem in deeper networks, which had been a mainstay challenge in the deep learning research field since the inception of ConvNets ([80]). The fundamental component of a ResNet is the residual block, which incorporates a "skip-connection". The input to the block is passed through an identity mapping and is summed to the activation of a series of convolutional

layers. ResNets can reach theoretically infinite depths, due to the self-regularization method provided by the identity mapping ([43]). The flexibility of ResNets allows training for longer (more epochs), thus reducing the likelihood of overfitting ([43]).

5.4.2 Model configurations

In addition to the novel application of the Dynamic Unet in this field, a set of architectural configurations were made to the model. These model configurations further improve upon the current state of the art. The configurations introduced to the model were:

- ResNet encoders.
- Self-attention layer as part of the up-sampling path of the model.
- Blurring algorithm to avoid checkerboard artefacts.
- Bottleneck connection from input to output.

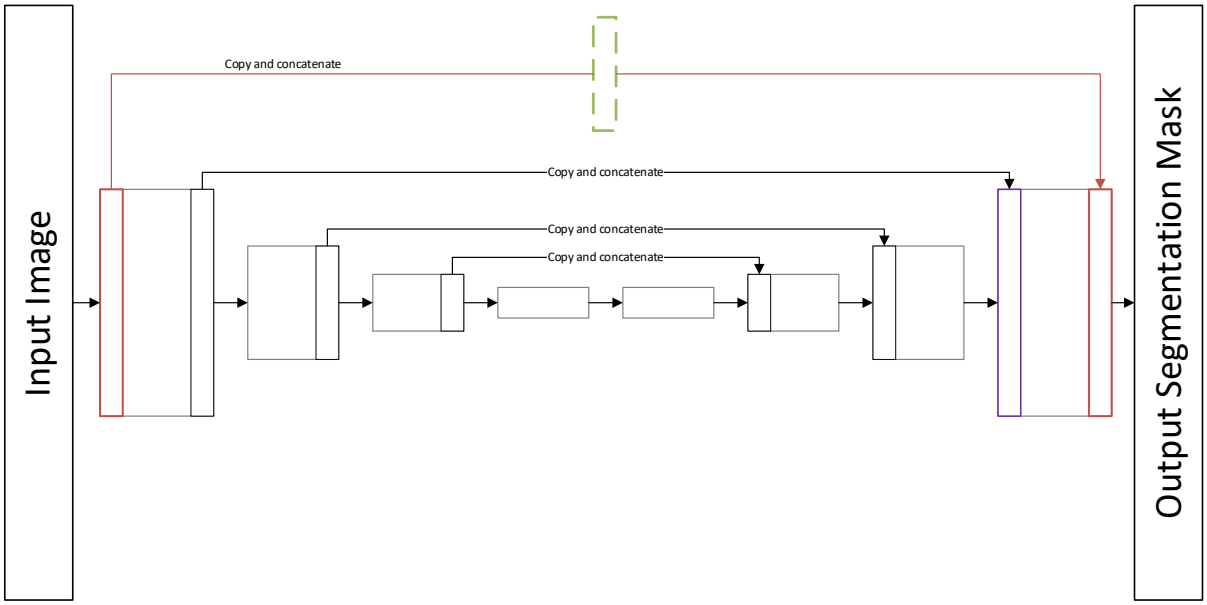


Figure 5.3: experimental layout. Purple represents the location of the self-attention layer (here represented in a simplified view of a ResNet18). Red represents the skip connection from the first to last layer. Green is the bottlenecked connection.

5.4.2.1 ResNet encoders

A ResNet encoder refers to the encoder portion of a neural network architecture known as ResNet, introduced in paragraph 2.1.9. ResNet is a deep learning architecture that was introduced by researchers at Microsoft Research ([81]) in 2015.

The ResNet architecture is commonly used for tasks like image classification and object detection, where deep networks are required. The encoder part of a ResNet typically refers to the initial layers that process the input data and extract features from it. The encoder processes the input data through a series of convolutional layers and pooling operations, gradually reducing the spatial dimensions while increasing the number of channels (i.e., features) extracted. This encoder part is usually followed by a fully connected layer or a global pooling operation that produces the final feature representation used for the task at hand.

The chosen encoder architectures were ResNet18, ResNet34 and ResNet50 ([39]). It was important to convey the effect of architectural scaling on performance, and ResNets have been demonstrated to be an excellent choice for medical imaging work ([82, 83]). While the number of training parameters is considerably higher than most other solutions, the increased transferability of ResNets allows for improved results with lower amounts of data, validating the effort that was put into increasing data efficiency. The choice of not experimenting with bigger architectures, such as ResNet101, was made due to insufficient data for such architectures. This assumption was confirmed correct by the results.

5.4.2.2 Self-attention layer

Attention is a mechanism that was introduced by Bahdanau et al. ([84]) to improve neural machine translation tasks and attention-based models have been shown to excel in all contexts in which capturing global dependencies are necessary. Attention layers have since then been an integral part of transformer-based models ([29]). Attention-based models have demonstrated to excel in all contexts in which capturing global dependencies are necessary, including hybrid text-image tasks ([85, 86]), and computer vision tasks ([87]). Self-attention consists of a block of layers that outputs the attention feature maps of an input sequence with each element of the same sequence. Self-attention has been featured in promising research within CAD ([88]), with the usage of self-attention as a region-of-interest detection tool, thus allowing to obtain a cropped local image in which to perform a second classification.

The experiments were run with the implementation described by Zhang et al. ([89]), who introduced a ResNet-like approach to self-attention by including a skip-connection within the final output of a layer y_i:

$$y_i = \gamma o_i + x_i \tag{5.1}$$

Where γ is a learnable parameter that scales the output of a self-attention layer and x_i is the input of the layer. The approach results in greater spatial awareness by the model, a trait that is highly desirable in medical imaging, as images often present themselves with a pseudo-symmetrical and repetitive structure.

5.4.2.3 Blurring

To counteract the natural occurrence of checkerboard artefacts in CNNs ([90]), a blurring mechanism was introduced. An average pooling layer with 2x2 dimensions and the unit stride was added after each activation in the up-sampling path of the Dynamic UNet.

5.4.2.4 Bottleneck connection

The original UNet architecture did not feature any direct connection between input and the final layer, opting instead for a concatenation of the activation of the third layer and the activation of the last deconvolution layer ([28]). Adding an even less processed pass-through could lead to better spatial awareness and improved performance. In this work, a bottlenecked connection is included, aiming at forcing the model to synthetise the input information by using a bottleneck within the residual block, which halves the number of features in the convolutional path of the block.

5.4.2.5 Data augmentation

In addition to transfer learning, we further address the issue of insufficient training data by employing an aggressive data augmentation strategy compared to published literature. Data augmentation is an ensemble of techniques that are employed to artificially increase the amount of available data to reduce overfitting and improve generalisation. Specifically, in computer vision images are slightly altered through affine or lighting transformations. These can include rotations, cropping, contrast, and colour correction. By heavily employing data augmentation, the chances for overfitting are drastically lowered, allowing for improved performance with data being equal (by employing a bigger architecture), or by achieving similar results with lower amounts of data, as presented by Wong et al. ([92]).

Table 5.1: data augmentation overview. The probability column refers to the likelihood of any transformation to happen.

Transformation	**Parameters**	**Probability**
Horizontal flip	N/A	0.5
Rotation	±10°	0.75
Cropping	1.1 magnification	0.75

Contrast adjustment	±20%	0.75
Brightness adjustment	±10%	0.75
Perspective warp	±20% position of the observation plane	0.75

Several augmentation strategies were employed (summarised in Table 5.1), and augmented images were used for all model configurations. All transformations were performed with a reflection padding mode, mirroring the pixel values along the image border to fill the shape. Examples of augmented data can be seen in Figure . The combined probability of a transformation to a specific feature is:

$$p = 1 - \prod(1 - p_i) \tag{5.2}$$

Where p_i is the probability of the transformations that affect the specific feature to occur. The dataset was then comprised of ~99.95% data augmented images. The perfectly horizontal edge of the manually segmented mask was impacted by the perspective warp and the rotation transformation. This led to an overwhelming imbalance in the dataset, with ~93.75% of the images featuring an inclined mask as ground truth.

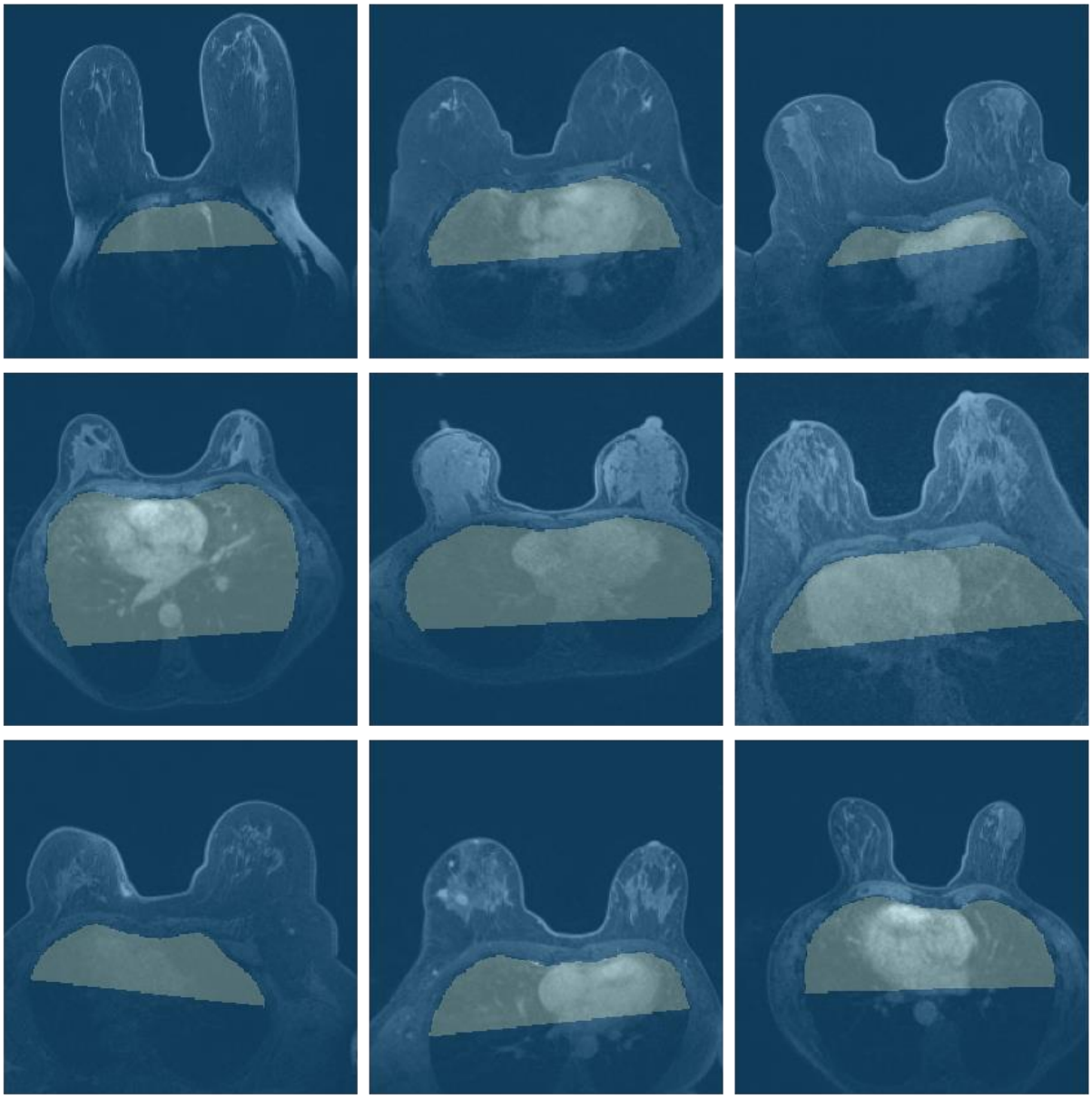


Figure 5.4: example of data augmentation on a small batch of data. The reflection padding can be seen in the top left corner. All images have their manually labelled ground truth featuring a straight line at the bottom of the mask.

5.4.2.6 Hyperparameters

To identify the best segmentation model for the segmentation task, a total of 18 different segmentation models were trained, each with a different combination of the configurations described above (see Table 5.2). The same training datasets (2146 images from 37 patients) were used for the training of all model configurations. Every model was trained with three distinct random seeds to ensure minimal stochastic noise in the results.

The optimiser for the training was Adam ([93]), with β_1 of 0.9 and β_2 of 0.99. The optimal learning rate for each architecture was found as shown by Smith ([94]). The training phase featured learning rate annealing, as described by the 1-cycle policy ([95]).

The models were trained using a NVIDIA P40 with 24 GB of VRAM. Batch sizes were chosen empirically, with ResNet18 models having 64 (31,208,178 parameters), ResNet34 having 32 (41,316,338 parameters), and ResNet50 having 8 (341,254,226 parameters).

5.5 Experiments

The performance of each of the 18 different models was tested with images from seven patients not used for training. The computational demand of each model was evaluated by calculating inference times for segmentation. Inference times were calculated for each image and for all images of the DCE-MRI sequence of each patient (n = 1260).

The segmentation result of each model was compared to 406 images (58 images per patient) that were manually segmented. Comparisons of agreement between the model segmentation and the manual segmentation were made using the DSC Similarity Coefficient (DSC) and the Jaccard Similarity Coefficient (JSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{5.3}$$

$$JSC = \frac{|X \cap Y|}{|X \cup Y|} \tag{5.4}$$

Where X is the number of pixels inside the thoracic cavity that are part of the manual segmentation and Y is the number of pixels inside the thoracic cavity that are part of the model segmentation. The DSC represents the mean overlap and the JSC represents the union overlap of pixels that are common to both the manual and automatic segmentation. For each model, DSC and JSC were determined for each image and the mean ± standard deviation (SD) of each coefficient was calculated for all 406 images.

5.5.1 Results

DSC and JSC for each model are summarised in Table 5.2. Inference times are summarised in Table 5.3. The best agreement between model segmentation and manual segmentation was found for the ResNet34 with self-attention, bottlenecked connection, and blurring mechanism model configuration with a DSC of 0.9359, JSC of 0.8874 and inference times of 33.56 ms/image and 42.3 s for all 1260 images of one DCE sequence. The worst agreement between

model segmentation and manual segmentation was found for the ResNet50 with self-attention model configuration with a DSC of 0.9210, JSC of 0.8670, and inference times of 194.8 ms/image and 245.2 s for all 1260 images of one DCE sequence.

Table 5.2: similarity coefficients for all models. The similarity coefficients are represented as mean ± standard deviation. The best performing model was the ResNet34 model with self-attention layer, bottleneck connection and blurring mechanism, which is highlighted in bold. Abbreviations: BC – bottleneck connection, BL – blurring mechanism, DSC – DSC Similarity Coefficient, JSC – Jaccard Similarity Coefficient, SA – self-attention layer.

Model configuration	**Mean DSC (n = 406)**	**Mean JSC (n = 406)**
Resnet18	0.9253 ± 0.1034	0.8760 ± 0.0758
ResNet18 + SA	0.9296 ± 0.1028	0.8757 ± 0.0765
ResNet18 + BL	0.9273 ± 0.1033	0.8795 ± 0.0764
ResNet18 + BC	0.9283 ± 0.1045	0.8742 ± 0.0763
ResNet18 + SA + BL	0.9348 ± 0.1045	0.8846 ± 0.0760
ResNet18 + SA +BL + BC	0.9293 ± 0.1036	0.8755 ± 0.0765
Resnet34	0.9244 ± 0.1017	0.8721 ± 0.0756
ResNet34 + SA	0.9230 ± 0.1012	0.8652 ± 0.0752
ResNet34 + BL	0.9227 ± 0.1015	0.8714 ± 0.0749
ResNet34 + BC	0.9292 ± 0.1009	0.8754 ± 0.0755
ResNet34 + SA + BL	0.9337 ± 0.1008	0.8780 ± 0.0750
ResNet34 + SA + BL + BC	**0.9359 ± 0.1004**	**0.8874 ± 0.0748**
Resnet50	0.9240 ± 0.1055	0.8717 ± 0.0781
ResNet50 + SA	*0.9210 ± 0.1069*	*0.8670 ± 0.0790*
ResNet50 + BL	0.9233 ± 0.1063	0.8708 ± 0.0777
ResNet50 + BC	0.9257 ± 0.1059	0.8730 ± 0.0775
ResNet50 + SA + BL	0.9278 ± 0.1061	0.8727 ± 0.0766
ResNet50 + SA +BL + BC	0.9289 ± 0.1053	0.8740 ± 0.0770

Table 5.3: inference times for all models. Times are represented in ms/image for single image processing, and in seconds for batch processing. The best performing model was the ResNet18 with no additions, which is highlighted in bold. Abbreviations: BC – bottleneck connection, BL – blurring mechanism, DSC – DSC Similarity Coefficient, JSC – Jaccard Similarity Coefficient, SA – self-attention layer.

Model configuration	Inference time (ms/image)	Batch inference time (1260 images) (s)
Resnet18	**26.83**	**30.0**
ResNet18 + SA	33.56	42.3
ResNet18 + BL	31.88	40.2
ResNet18 + BC	27.68	34.9
ResNet18 + SA + BL	33.56	42.3
ResNet18 + SA +BL + BC	33.56	42.3
Resnet34	31.88	40.2
ResNet34 + SA	33.56	42.3
ResNet34 + BL	33.56	42.3
ResNet34 + BC	30.2	38.1
ResNet34 + SA + BL	33.56	42.3
ResNet34 + SA + BL + BC	33.56	42.3
Resnet50	179.5	226.2
ResNet50 + SA	194.6	245.2
ResNet50 + BL	196.3	247.3
ResNet50 + BC	194.6	245.2
ResNet50 + SA + BL	196.3	247.3
ResNet50 + SA +BL + BC	*198*	*249.5*

Upon visual inspection of the model segmentation results, it was found in all models that the lower edge of the segmented area was not straight, which influenced the results of the similarity coefficients. This phenomenon is an

artefact caused by the data augmentation strategy, which allowed for reduced overfitting, and better training outcomes. The training data rarely was not augmented before being inputted into the models, resulting in the horizontal line at the bottom of the manually segmented mask to be interpreted as always inclined. A simple adjustment step was then taken to ensure the real-world usability metrics of the methodology. The refinement algorithm consists of a simple loop over the columns of the output of the model, forcing any pixel between the highest true value in a column and the 131st from the bottom to be true. The updated schematic of the solution can be seen below in Figure 5.4.

The sole purpose of the refinement algorithm is to remove the bias inserted into the training process by the aggressive data augmentation strategy. Such an algorithm would not be used in a workflow for lesion detection, or in clinical practice, as it aims at improving performance in areas that result in no additional value in CAD or diagnosis. Performance metrics of the refined masks should then be interpreted as directly correlated with the actual capability of models to segment the chest cavity.

Application of the refinement algorithm led to considerable improvements in the similarity coefficients, but with a slight increase in inference times (see Table 5.5 and Table 5.5). The best agreement between model segmentation and manual segmentation was again found for the ResNet34 with self-attention, bottlenecked connection, and blurring mechanism model configuration with a DSC of 0.9612, JSC of 0.9789 with inference times of 98.99 ms/image. The worst agreement between model segmentation and manual segmentation was found for the ResNet50 with self-attention model configuration with a DSC of 0.9465, JSC of 0.9708, with inference times of 261.7 ms/image.

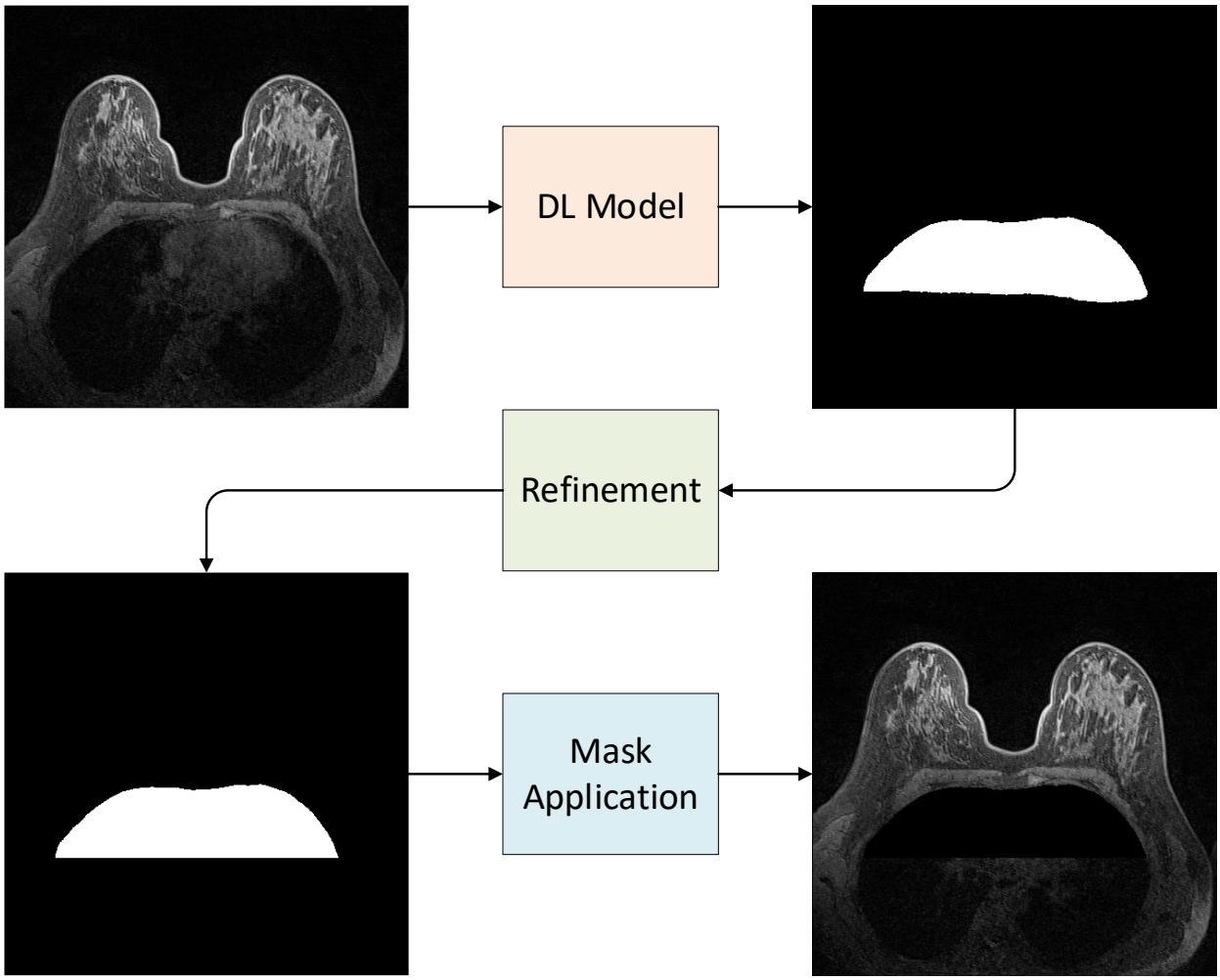


Figure 5.5: schematic of the solution. The input image is inputted in a deep learning model, which outputs a generated mask. A refinement algorithm then removes the bottom part of the mask. The segmentation mask is then applied to the input image.

Table 5.4: similarity coefficients for all models after refinement algorithm. The similarity coefficients are represented as mean ± standard deviation. The best performing model was the ResNet34 model with self-attention layer, bottleneck connection and blurring mechanism, DSC – DSC Similarity Coefficient, JSC – Jaccard Similarity Coefficient, SA – self-attention layer.

Model configuration	**DSC (n = 406)**	**JSC (n = 406)**
Resnet18	0.9750 ± 0.0451	0.9552 ± 0.0667
ResNet18 + SA	0.9731 ± 0.0449	0.9552 ± 0.0669
ResNet18 + BL	0.9739 ± 0.0447	0.9536 ± 0.0701
ResNet18 + BC	0.9717 ± 0.0450	0.9497 ± 0.0670

Model configuration	DSC (n = 406)	JSC (n = 406)
ResNet18 + SA + BL	0.9751 ± 0.0448	0.9556 ± 0.0665
ResNet18 + SA +BL + BC	0.9737 ± 0.0447	0.9533 ± 0.0674
Resnet34	0.9744 ± 0.0420	0.9541 ± 0.0682
ResNet34 + SA	0.9734 ± 0.0419	0.9527 ± 0.0628
ResNet34 + BL	0.9698 ± 0.0417	0.9512 ± 0.0642
ResNet34 + BC	0.9766 ± 0.0422	0.9577 ± 0.0677
ResNet34 + SA + BL	0.9775 ± 0.0425	0.9584 ± 0.0633
ResNet34 + SA + BL + BC	**0.9789 ± 0.0411**	**0.9612 ± 0.0621**
Resnet50	0.9718 ± 0.0493	0.9541 ± 0.0627
ResNet50 + SA	*0.9665 ± 0.0502*	*0.9538 ± 0.0721*
ResNet50 + BL	0.9708 ± 0.0488	0.9465 ± 0.0704
ResNet50 + BC	0.9732 ± 0.0499	0.9526 ± 0.0706
ResNet50 + SA + BL	0.9712 ± 0.0501	0.9532 ± 0.0710
ResNet50 + SA +BL + BC	0.9766 ± 0.0497	0.9577 ± 0.0708

Table 5.5: inference times for all models with subsequent refinement algorithm. Times are represented in ms/image for single image processing, and in seconds for batch processing. The best performing model was the ResNet18 with no additions, which is highlighted in bold. Abbreviations: BC – bottleneck connection, BL – blurring mechanism, DSC – DSC Similarity Coefficient, JSC – Jaccard Similarity Coefficient, SA – self-attention layer.

Model configuration	Inference time (ms/image)	Batch inference time (1260 images) (s)
Resnet18	**92.28**	**116.3**
ResNet18 + SA	93.96	118.4
ResNet18 + BL	92.28	116.3
ResNet18 + BC	92.28	116.3
ResNet18 + SA + BL	95.64	120.5

Model configuration	Inference time (ms/image)	Batch inference time (1260 images) (s)
ResNet18 + SA +BL + BC	95.64	120.5
Resnet34	90.6	114.2
ResNet34 + SA	95.64	120.5
ResNet34 + BL	95.64	120.5
ResNet34 + BC	92.28	116.3
ResNet34 + SA + BL	97.32	122.6
ResNet34 + SA + BL + BC	98.99	124.7
Resnet50	258.4	325.6
ResNet50 + SA	263.4	331.9
ResNet50 + BL	261.7	329.7
ResNet50 + BC	260.1	327.7
ResNet50 + SA + BL	261.7	329.7
ResNet50 + SA +BL + BC	*266.8*	*336.2*

5.6 Discussion

The main aim of this work was to develop a fully automatic methodology for the segmentation of the chest cavity to use for the development of automatic lesion detection systems. All 18 models showed excellent agreement with the manual segmentation with DSCs of over 0.92 and JSCs of over 0.86. After the introduction of a refinement algorithm to compensate for artefacts due to the aggressive data augmentation strategy, the DSCs and JSCs improved to over 0.95 and 0.96, respectively. The best performing model was found to be the ResNet34 model with self-attention layer, blurring mechanism and bottleneck connection resulting in very high agreement with manual segmentation with a DSC of 0.9612 and a JSC of 0.9789.

Overall, the techniques proposed show a significant improvement over the current state of the art (Piantadosi et al., 2020), which features a DSC of 0.9660 on the full breast segmentation. While the tasks are not directly comparable, the body-air segmentation has been shown to achieve more than optimal results in the past ([25]), hence highlighting

the need for novelty solely on the chest cavity segmentation. Moreover, DSC is expected to be lower with smaller regions to segment. Hence, it is reasonable to assume that the cost to DSC and JSC of generating an incorrect chest wall segmentation is lower in Piantadosi et al. ([11]).

The best configuration also improves on inference timings, matching the previous state of the art after post-processing on comparable hardware, but dramatically improving on it without. As previously mentioned, the post-processing algorithm is solely used to highlight the validity of the proposed solution and is not meant to be included in any CAD workflow.

The agreement between the model segmentation and the manual segmentation is very good to excellent for all model configurations. The addition of each configuration to the model provided marginal improvements for all architectures, however, the models with all configurations added consistently resulted in higher performance. This improvement was particularly striking for the ResNet 34 encoder, which showed an increase in the DSC from 0.9541 (ResNet34 only) to 0.9612 (ResNet34 with self-attention layer, blurring and bottleneck). However, adding to the complexity of the model comes with an increased computational cost, especially when adding the self-attention layer. For example, inference time for the entire DCE image sequence is 114.2 s for ResNet34 only, but this increases by 6.3 s when adding the self-attention layer while adding the bottleneck connection causes an increase of only 2.1 s (Table). The best performing model before post-processing is the ResNet34 featuring all proposed additions. The considerable improvement in segmentation accuracy over a standard ResNet encoder may justify the additional computational cost. ResNet50 configurations performed worse than the smaller ResNet34 architectures. This is likely due to under fitting and having access to more data would be extremely beneficial, and most likely would lead to better performance than ResNet34 on larger training datasets.

In cases of limited computational resources, both at training time and at inference time, the ResNet18 configurations would be the recommendation.

As the refinement algorithm is applied, improvements can be uniformly observed across all model configurations. By removing the bottom boundaries of the generated masks, the mean JSC is increased by 0.098, and the average DSC is increased by 0.026. The low variance of both similarity coefficients corroborates the observation that aggressive data augmentation is one of the main causes of irregularities in the generated masks (Table).

In summary, this chapter reported about a series of experiments that were run on Dynamic UNet architectures to determine their performance in chest cavity segmentation, achieving the best performance with a ResNet34 down sampling path, a self-attention layer, a blurring mechanism, and a bottlenecked connection between the activation of the first block and the last block of the model. The average model performance without any post-processing achieved a DSC of 0.9359 and a JSC of 0.8874. Applying a simple algorithm aimed at correcting artefacts that were generated by aggressive data augmentation yielded an average DSC of 0.9612 and JSC 0.9789, thus setting a new state of the art. While the results are a significant step in the direction of clinical usability, future research should aim to expand the overall training datasets, as well as test the solution with different acquisition protocols and/or tasks.

Chapter 6. Lesion characterisation

6.1 Introduction

This chapter focuses on the development and evaluation of advanced machine learning models for the classification of breast cancer lesions based on their morphological characteristics, specifically lesion margins. It highlights the importance of selecting the right classifiers and methodologies for improving diagnostic accuracy in medical imaging. The chapter presents a comprehensive overview of various architectures, data augmentation techniques, and training methodologies aimed at enhancing model performance in a challenging dataset of breast cancer lesions.

The dataset includes a diverse set of images, annotated by radiology specialists, with detailed descriptions of irregular and smooth lesions. The chapter also includes a detailed comparison of model performance, analysing the effects of different training techniques and architectures, offering insights into their applicability to medical image classification tasks. This work sets a new benchmark in the field, contributing to the ongoing efforts to enhance machine learning's role in supporting medical professionals with accurate diagnostic tools. Classification is the final step for providing a comprehensive support to professionals in breast cancer diagnosis.

6.2 Datasets and Labelling Strategy

The margin of the lesion is one of most important morphological descriptors ([96]) for breast cancer diagnosis, and therefore margin identification is the key outcome of the classifier. The training data was chosen by manually selecting suspicious cases, according to involved radiology specialist, and properly labelling them; in addition, the considered cases have been object of further biopsy analysis for review and confirmation. The training data was comprised of 2405 images of irregular lesions (referring to 96 cases) and 561 images of smooth lesions (referring to 20 cases), from the total of 8 time steps, plus the high-resolution scan of lesions that were validated by an expert radiologist. The validation set was comprised of 476 images of irregular lesions (19 cases) and 63 of smooth lesions (2 cases).

The images were obtained by cropping the images to an area of 224 by 224 pixels symmetrically around the lesion, in a way to enforce centrality of the suspect tissues. In cases where insufficient pixels were available to enforce centrality (see Figure 6.2 left), the image was cropped asymmetrically.

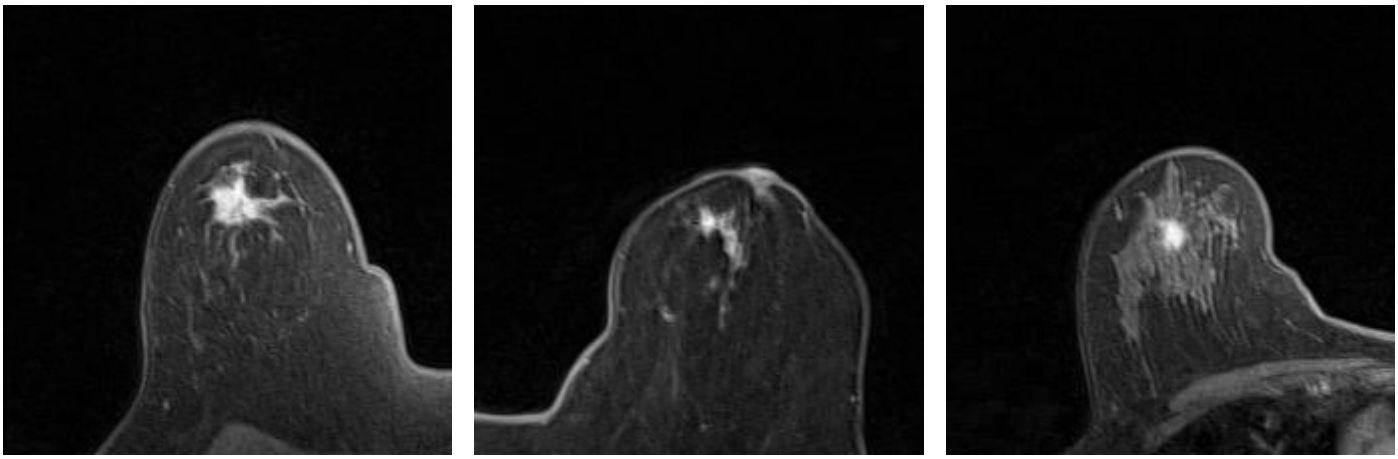


Figure 6.1: examples of irregular lesions.

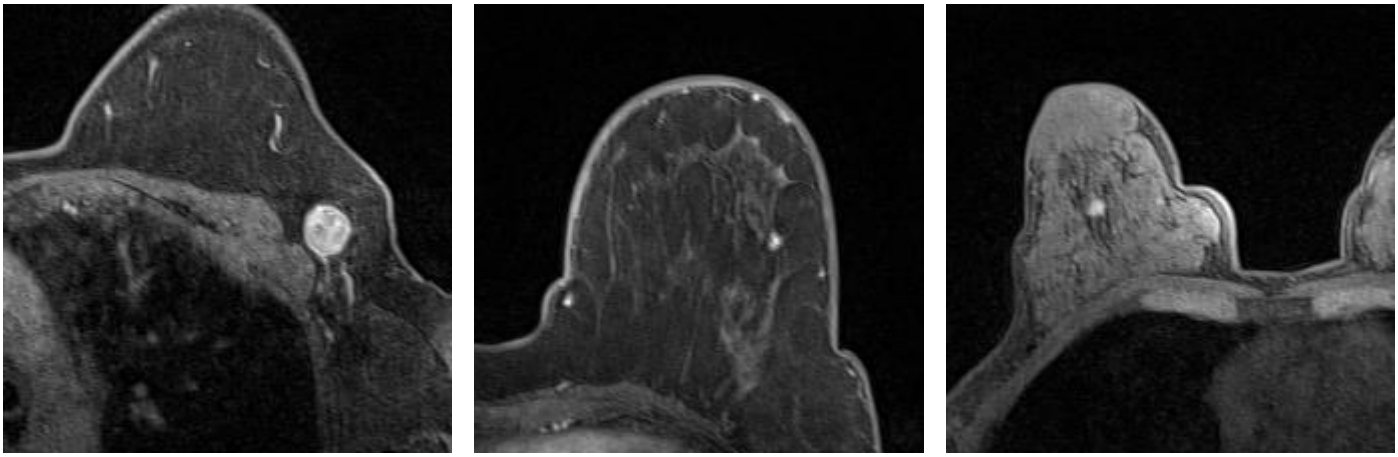


Figure 6.2: examples of smooth lesions.

6.3 Methodology

To create a state-of-the-art margin classifier to support doctors in their diagnostic process, a selection was made between several pre-trained architectures for image analysis available on an open-source repository ([97]). As a side product, it was also possible to verify the transfer learning capabilities of these architecture to a small and difficult dataset such as this one.

In the end two transformer architectures (the type of architectures that have proven to be the best performers in natural language processing) were chosen and tested in combination with three fine-tuning techniques, and the performances were duly compared on the dataset. Transformer-based architectures have shown to strongly outperform traditional convolutional approaches in this specific task. The potential explanation is the greater spatial awareness that transformer-based architecture benefit from ([98]). To further provide substance to the claim, it is worth noting that

both reported architectures (best performing out of all experiments) were developed to enhance spatial awareness in ImageNet classification.

6.3.1 XCiT: Cross Co-Variance Image Transformers

Cross-Covariance Image Transformers, or XCiT ([99]), builds upon the transformer architecture and introduces cross-covariance attention, which allows the model to capture relationships between patches in an image.

The XCiT model replaces the self-attention mechanism found in traditional transformers with cross-covariance attention. In the original transformer, self-attention calculates the attention weights between all pairs of tokens (in the case of vision transformers, tokens correspond to image patches). However, this self-attention mechanism has a quadratic complexity concerning the number of tokens, making it computationally expensive for large images.

XCiT addresses this problem by introducing cross-covariance attention, which computes attention weights based on the cross-covariance between token pairs. Cross-covariance measures the relationship between two variables, indicating how changes in one variable are associated with changes in another variable. In the case of XCiT, the cross-covariance is computed between the feature representations of two tokens.

The XCiT model consists of an encoder that processes the input image by dividing it into patches and linearly projecting them into a higher-dimensional feature space. These projected patches are then passed through multiple layers of the XCiT encoder. Each encoder layer performs cross-covariance attention, followed by feed-forward neural networks (FFNs) and layer normalization. The FFNs introduce non-linear transformations to capture complex relationships between tokens.

During cross-covariance attention, the model calculates the cross-covariance matrix between pairs of tokens and applies SoftMax normalization along rows and columns to obtain attention weights. These attention weights are then used to compute weighted sums of token representations, resulting in context-aware representations for each token.

By using cross-covariance attention, XCiT can capture long-range dependencies and relationships between tokens efficiently, enabling the model to understand the global context of the image. This helps the model perform well on various vision tasks such as image classification, object detection, and segmentation.

6.3.2 BEiT: BERT Pre-Training of Image Transformers

BEIT ([100]) (Bidirectional Encoder representation from Image Transformers), or BERT ([101]) with vision, borrows concepts from natural language processing to computer vision tasks.

The main idea behind BEiT is to adapt the transformer architecture, originally designed for sequential data like text, to handle images. It aims to leverage the success of pre-training transformers on large text corpora and apply similar techniques to pre-train vision transformers on large-scale image datasets.

BEiT introduces several key modifications to adapt BERT for vision tasks:

- Image Patch Embeddings: Like vision transformers, BEiT divides an input image into smaller patches. Each patch is then linearly projected to a high-dimensional feature space to obtain patch embeddings. These embeddings serve as the input tokens for the transformer.
- Positional Embeddings: To capture the spatial information of the image, BEiT incorporates positional embeddings that encode the relative positions of the image patches. These embeddings help the model understand the location of each patch within the image.
- Patch Tokenization: BEiT introduces a patch tokenization strategy to encode the information about patch locations. Each patch is assigned a unique token, including special tokens like [CLS] and [SEP] used in BERT. These tokens help the model differentiate between patches and provide additional positional information.
- Pre-training: BEiT pre-trains the vision transformer in a self-supervised manner. It uses a large-scale dataset containing unlabelled images and applies various pretext tasks, such as patch order prediction and masked patch prediction, to learn meaningful representations. The model learns to predict the correct order of randomly shuffled patches and to reconstruct the original image from masked patches.

By adapting the BERT architecture to images, BEiT demonstrates strong performance on various vision benchmarks, often surpassing or achieving competitive results compared to other vision transformer models. It benefits from the large-scale pre-training on unlabelled data, which helps the model learn general visual representations that can be fine-tuned for specific tasks.

In specific, BEiTv2 ([102]) was used during this study, which greatly enhances BEiT's performance by employing a Vector-Quantized Knowledge Distillation (VQ-KD) algorithm to discretise a semantic space, rather than a pixel-space,

to employ as an input and target for the optimisation of BEiTv2. This has led to significant improvements in the transfer learning capabilities of the original BEiT.

6.3.3 Progressive resizing

The intuition of using differently sized images during training was first introduced in 2019 by Hoffer at al. ([103]). It leverages the properties of modern deep learning architectures ([104]) to allow for scale-agnostic learning of relevant features by training networks with gradually increasing input dimensions. This has been empirically proven to yield results akin to having additional data, with diminishing returns by myself in this study and other practitioners ([105, 106]). The additional benefits of using lower resolution images during training is to force the training process to involve low-frequency patterns first (high-level), and later adding finer details.

In this work, training images with resolution of 224 by 224 pixels were used to generate datasets representing the same images, although with resolutions of 112 by 112 pixels and 168 by 168 pixels. The training process was then carried out by transferring from a pretrained architecture, fine tuning on the lowest resolution dataset, followed by subsequent transfers and trainings to higher resolution datasets.

Progressive resizing is not always possible, as in the case of BEiT, which has invariable input dimensions. The aim of this work also includes the comparison between state-of-the-art methodologies that allow for progressive resizing as part of the fine tuning, and ones that do not.

6.3.4 Mix-up augmentation

Mix-up augmentation is a data augmentation technique that aims at improving the generalization and robustness of models by creating new training examples through mixing pairs of existing examples ([107]).

The key idea behind Mix-up augmentation is to combine two or more training images and their corresponding labels, to generate a new training example. This mixing process occurs at the input level, meaning that the input data is linearly interpolated, and the labels are also mixed proportionally.

In essence, Mix-up is carried out as follows:

- Select two training examples randomly from the dataset.
- Randomly generate a mixing coefficient, typically drawn from a beta distribution. The mixing coefficient determines how much influence each example will have on the mixed example.

- Combine the input data of the selected examples by taking a weighted average. This is done by blending the pixel values based on the mixing coefficient.
- similarly combine the corresponding labels. For classification tasks, such as the one employed in this work, the labels are one-hot encoded vectors, and the mixing process involves taking a weighted average of these vectors.
- Repeat the process for a specified number of times or until a desired augmented dataset size is achieved.

The main intuition behind Mix-up augmentation is that by creating new examples that lie between two existing examples, the model is encouraged to learn more generalizable representations and decision boundaries. It introduces a form of regularization that encourages the model to be more robust to small perturbations and variations in the input data.

Mix-up augmentation is effective in reducing overfitting, improving model accuracy, and enhancing generalization in various computer vision tasks, such as image classification, object detection, and segmentation.

6.3.5 Label smoothing

Label smoothing is a regularization technique involving smoothing the target labels during training by redistributing a portion of the probability mass from the true label to other classes ([108]).

The target labels are represented as one-hot encoded vectors, where the true class is assigned a probability of 1 and all other classes have a probability of 0. However, this can lead to models that are excessively confident and prone to overfitting. Label smoothing addresses this issue by introducing a small amount of uncertainty into the training process. Instead of assigning a probability of 1 to the true class, label smoothing redistributes a portion of that probability mass to other classes. This is typically done by subtracting a small value, ε, from the true class probability and adding $\varepsilon/(K-1)$ to the other classes, where K is the total number of classes. The label smoothing operation results in a smoothed target distribution, where the true class has a probability of $1-\varepsilon$, and the other classes have a probability of $\varepsilon/(K-1)$.

The main idea behind label smoothing is to encourage the model to be more calibrated and to learn more robust decision boundaries. By introducing some uncertainty into the training labels, the model becomes less certain and more tolerant of small errors or noise in the input data. This regularization can help prevent overfitting, improve generalization, and make the model more resilient to adversarial examples.

Label smoothing is particularly useful in scenarios where the training dataset may contain noisy or incorrect labels, such as this one. It helps the model to be less reliant on individual training samples and reduces the risk of over-optimization on the training set. The hyperparameter ε could be a source of underperformance, as too high values could lead to uncertainty and instability in the prediction at inference time. Multiple values of ε were tested (ranging from 0 to 0.3 in steps of 0.03), and only the best performing one is presented in this work (ε = 0.1).

6.3.6 Training Hyperparameters

To identify the best lesion characterization model for the classification task, a total of 11 different classification models were trained, using the methodologies described above. Each model was trained 5 times each with varying random seeds with a weight decay of 0.01; 5 times with no weight decay; 5 times with a weight decay of 0.01. All best results feature weight decay of 0.01. Results are reported for the best performing model on the validation set.

The optimizer for the training was Adam, with β_1 of 0.9 and β_2 of 0.99. The optimal learning rate for each architecture was found as shown by Smith ([86]). The training phase featured learning rate annealing, as described by the 1-cycle policy ([87]).

The models were trained using a NVIDIA RTX 3090 with 24 GB of VRAM. Batch sizes were chosen empirically to be 8 for all models, however the number is inconsequential as gradient accumulation was employed.

A benchmark ResNet34 model was also used to gauge the gains that these new architectures bring to the field of medical images classification. The training methodology employed was identical, with differences in batch size, which was chosen to be 64.

6.4 Experiments

The models were evaluated on the relevant performance metrics, namely Accuracy, Recall, Precision and F1 Score, summarised in tables 6.1, 6.2, 6.3. Each model was evaluated with the proposed additions during training, with the exceptions of progressive resizing for BEiTv2, as the fixed input nature of the architecture does not allow it.

The best performing models were the BEiTv2 architectures, with the configuration featuring Mix-up augmentation and label smoothing, scoring highest in all metrics, with an accuracy of 0.9314, a recall if 0.8923, a precision of 0.9748 and an F1 score of 0.9317. Inference times were comparable, as the model sizes are, with BEiTV2 being slightly faster

than alternatives. It is important to note how larger these times are to the benchmark ResNet34. This is due to inefficient operations (self-attention), as well as smaller model size in the case of BEiTv2.

The losses agree with the actual performances of the models, as expected given the similarities within the training set. However, it is important to note how a validation loss of 0.2174 (such as the one from the best performing BEiTv2 model) and a loss of 0.5141 (such as the one of the worst performing ResNet34 model) will not feature dramatically different metrics (i.e. F1 scores of 0.961498 and 0.872776 respectively).

Table 6.1: results for the architecture xcit_tiny_12_p8_224_dist

	Model Size	**Train loss**	**Valid loss**	**Inference time (images/s)**	**Accuracy**	**Recall**	**Precision**	**F1 score**
Normal	7M	0.6199	0.3273	216.5	0.8479	0.8807	0.9622	0.9197
+ Progressive resizing	7M	0.6019	0.3111	216.5	0.8553	0.8885	0.9706	0.9277
+ Mix-up	7M	0.5718	0.3098	216.5	0.8590	0.8923	0.9745	0.9317
+ Label smoothing	7M	0.5422	0.3628	216.5	0.8738	0.9036	0.9453	0.9240

Table 6.2: results for the architecture xcit_small_24_p16_224

	Model Size	**Train loss**	**Valid loss**	**Inference time (images/s)**	**Accuracy**	**Recall**	**Precision**	**F1 score**
Normal	48M	0.6259	0.3263	174.4	0.8798	0.8898	0.9552	0.9214
+ Progressive resizing	48M	0.6254	0.3231	174.4	0.8851	0.8952	0.9610	0.9269
+ Mix-up	48M	0.6072	0.3215	174.4	0.9017	0.9119	0.978992	0.9443
+ Label smoothing	48M	0.6229	0.3945	174.4	0.8850	0.9036	0.945378	0.9240

Table 6.3: results for the architecture beitv2_base_patch16_224_in22k

	Model Size	**Train loss**	**Valid loss**	**Inference time (images/s)**	**Accuracy**	**Recall**	**Precision**	**F1 score**
Normal	87M	0.4126	0.2655	193.2	0.8544	0.9267	0.9875	0.9561

+ Progressive resizing	87M	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
+ Mix-up	87M	0.3986	0.2174	193.2	0.9314	0.9526	0.9706	0.9615
+ Label smoothing	87M	0.5980	0.4284	193.2	0.8664	0.9529	0.8929	0.9219

Table 6.4: results for the benchmark architecture ResNet34

	Model Size	**Train loss**	**Valid loss**	**Inference time (images/s)**	**Accuracy**	**Recall**	**Precision**	**F1 score**
Normal	21.8M	0.93952	0.514113	299.4	0.795058	0.928364	0.823468	0.872776
+ Progressive resizing	21.8M	0.897209	0.496763	299.4	0.80743	0.94541	0.83157	0.884844
+ Mix-up	21.8M	0.86314	0.49291	299.4	0.807524	0.949459	0.832978	0.887413
+ Label smoothing	21.8M	0.849426	0.4748	299.4	0.820037	0.952915	0.836134	0.890713

6.5 Discussion

The main aim of this work was to develop a lesion morphology classifier. All 11 models showed excellent agreement with the manually annotated data and decisive improvements over the benchmark models. The best performing model was a BEiTv2 with Mix-up augmentation which resulted in values for accuracy, recall, precision and F1 score of 0.93, 0.95, 0.97, 0.96 respectively on the validation set.

Overall, the novel application of these state-of-the-art techniques to medical images classification has shown remarkable improvements over the previous state of the art, represented by the benchmark ResNet34 model, with an improvement of 0.07 for the F1 score (7.9%) and an improvement of 0.11 for accuracy (13.5%). The significancy of these results is further exemplified by the wide gap in validation loss between the best performing BEiTV2 model and the reference ResNet34, with the former being 2.18 times lower than the latter.

While a comparison to previous literature is impossible, as to the author's knowledge there are no comparable proposed solutions to the lesion morphology classification, it is observable how the application of state-of-the-art techniques to the problem establishes a strong benchmark as state of the art for future work to aim at improving upon.

Progressive resizing and Mix-up augmentation provided added performance to all configurations, proving the effectiveness of these proposed techniques. The performance gain, however, differed from architecture to architecture. XCiT featured an average validation loss improvement of 3.00% with the application of progressive resizing, however the improvement is consistent only on smaller variants (XCiT Tiny, 7M parameters, -4.8%). The behaviour with Mix-up augmentation is similar for XCiT, with average reductions of 3.47% in validation losses, and with minute differences in the bigger variants (XCiT Small, 48M parameters, -1.53%). It is possible to conclude, then, that these data efficiency techniques, while improving on all fronts, are much more suited to smaller architectures, where selective learning of size-invariant features is of higher importance. Hence, Mix-up augmentation and label smoothing are strong recommendations for tasks in which data is limited and smaller models need to be employed.

BEiTV2 featured strong improvements with Mix-up augmentation, and in general the models would have performed worse than XCiT Small variants without it. The improvement is so massive (-19.5% in validation losses) that it should be mandatory for any solution in the medical imaging space looking to transfer from this pretrained architecture.

Overall, the recommended solution for datasets similar in size and nature to the one in this work is the BEiTV2 with Mix-up augmentation, as it represents a marked improvement over the benchmark ResNet34 in terms of performance, with minimal cost in terms of inference timings. To further elaborate, a transferability evaluation of these architectures was performed by obtaining top-1 accuracy metrics on ImageNet ([109]) and comparing to accuracies obtained in this work. The most transferrable architecture was defined to be the one with highest improvement over the pretrained model on ImageNet in terms of accuracy. Once again, the recommended architecture was the only one to improve upon this metric (0.9314 vs 0.8974), proving its extreme capacity for transfer learning to medical images.

Chapter 7. Conclusions and Recommendation for Further work

7.1 Key Findings

Several advancements in the field of computer-aided diagnosis for breast cancer in DCE-MRI are presented in this work. All the objectives for this project were met. In chapter 4, a novel unsupervised deep learning methodology was presented to allow less reliance on costly manual data annotation, while maintaining comparable true positive rates against the state of the art. Moreover, the technique is based on the radiologists' diagnostic protocol, allowing for greater interpretability compared to other examples in literature. The limitations of this approach, namely the tendence of identifying the heart region as suspicious, was addressed in chapter 5, in which a state-of-the-art semantic segmentation algorithm was presented, which decisively improved upon the previous state of the art, while being significantly more computationally efficient. Finally, a novel deep learning algorithm for lesion morphology classification was shown in chapter 6, laying the groundwork for further improvement in this area of study.

Overall, the results of the project are a significant step forward in integrating machine learning methodologies in clinical practice, as the proposed algorithms address the key weaknesses of the previously available methodologies by greatly reducing clinicians' involvement in the data labelling, increasing the interpretability of results, and by performing groundbreaking progress in morphology characterisation.

7.2 Limitations

The proposed methodologies, while significant, are limited to a relatively small dataset of images compared to the variability in human anatomy. Validation on more data, especially for ethnic groups that are underrepresented in the UK women population, is still needed to ensure the safety and usability of the solutions. Regarding this matter, the validation strategy proposed in chapter 4 of this work was chosen to ignore false positives, as there was no certainty on the nature of the highlighted area without a biopsy. A more comprehensive study would incorporate evaluation in clinical practice. Similarly, the evaluation of the accuracy in chapter 6 is based on ground truth that was obtained by a biopsy, but with labels based on visual perception.

While this study furthers the usability of machine learning tools in breast DCE-MRI diagnosis by having algorithms that mimic the diagnostic process, a clinician would still need to reconstruct the algorithmic process by inspecting the data, without the possibility of accessing a logical reasoning. On this topic, the algorithms proposed in this work would

greatly speed up and reduce errors in the identification and analysis of suspect areas only if paired with today's diagnostic process, which consists of analysing the volume by looking at every slice. Currently, there is no functionality for aggregating the results across the z-axis of the volume, allowing doctors to gain information of lesions that span multiple slices.

7.3 Implications

The results presented in this work propose several points of interest for the research community. First and foremost, by demonstrating that annotated data can be avoided, the ever-present challenge of obtaining labelled data could be considered voided, opening to research on wider datasets. Secondly, the empirical demonstration that neural networks can predict not only the location, but also morphology of lesions could lead to diagnostic support tools that aid radiologists in more ways than previously thought possible. Lastly, the quantitative observations made on the data and its annotations raise serious questions on the current research methodologies in the field, with researchers assuming correctness of labels attached to data. While unverified, it is reasonable to assume that similar inductive biases in the data are present for parallel research fields.

7.4 Recommendations for Further Work

The primary aim of research in the field of automated tools for diagnostic processes should be to reduce the need for labelled data, as requiring medical professionals to devote their time to data validation and annotation defeats the purpose of the research itself. Another main goal of research should be to allow for doctors to work alongside the algorithms, allowing them to understand the intricacies of an algorithmic decision. To this end, further research on this topic is recommended to follow the intuitions and ideas presented in chapter 4.

In detail, the following avenues for research are, to the extent of the author's interpretation of the challenges ahead, what would maximise real-world impact.

7.4.1 Logical Reasoning Functionality

As mentioned in the limitations section, the proposed methodologies are restricted to a workflow in which the radiologist examines the scan volume with the aid of deep learning algorithms to ensure faster and safer decisions. At the same time, if the doctor and the algorithm have diverging results, the clinician is forced into trying to understand why the outcome of the automated process is as such. Following future developments in large language modelling, it

would be possible to provide a brief explanation in natural language on the decision, allowing to add another layer of interpretability to the process.

7.4.2 Validation Methodologies

The challenges highlighted in chapter 3 and 4 regarding the truthfulness of labelled data and subsequent decisions to ignore false positive data stem from the limitations of current evaluation methodologies, which are far too reliant on ground truth data. The desired alignment, however, resides in increased accuracy and speed in clinical settings of users. To this end, standardised validation methodologies should be explored. In absence of these solutions, a more achievable target could be to have a publicly accessible validation dataset, in which the effectiveness of various methodologies could be objectively measured. Such dataset would have several categories and workflows to allow for benchmarking at various steps of the diagnostic process (i.e. one could start from the state-of-the-art lesion detection tool output to evaluate the capability of a lesion morphology classifier).

7.4.3 Improved methodologies

As highlighted previously, the proposed solution for lesion localisation relies on data on a singular enhancement curve and can highlight suspect lesions with state-of-the-art TPR. The methodology could be further improved by including information of the surrounding areas, both in the same plane, as well as on the x-axis. Taking this a step further, while losing interpretability, a whole slice or even a whole scan could be used as input for an unsupervised methodology.
As the field of machine learning keeps evolving at a fast pace, it is foreseeable that the methodologies presented in this work will eventually be superseded by equivalents employing more advanced neural networks architectures that will be developed for parallel fields.

7.4.4 Parallel Fields

The techniques presented in this work are, in principle, appliable to all other branches of radiology in which a contrast-enhanced MRI is employed for diagnosis. Attempts at adapting the contents of this work to different fields could provide valuable insight into potential improvements and limitations, as well as allowing for a broader impact of this work.

References

[1] NHS England and NHS Improvement, "Diagnostic imaging dataset annual statistical release 2019/2020.", London: NHE England 2020.

[2] The Royal College of Radiologists. "Clinical radiology UK workforce census 2020 report.", 2020.

[3] The Royal College of Radiologists. "The breast imaging and diagnostic workforce in the United Kingdom.", 2016.

[4] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, 2018. CA: A Cancer Journal for Clinicians 68, 394–424.

[5] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, 2021. CA: A Cancer Journal for Clinicians 71, 209–249.

[6] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, 2015, International Journal of Cancer 136, E359–E386.

[7] S. Vinnicombe, "How I report breast magnetic resonance imaging studies for breast cancer staging and screening.", Cancer Imaging 2016 Vol. 16, https://doi.org/10.1186/s40644-016-0078-0.

[8] N. Cho, S.A. Im, I.A. Park, K.H. Lee, M. Li, W. Han, D.Y. Noh, W.K. Moon, "Breast cancer: Early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging." Radiology 2014 Vol. 272, 385–396, https://doi.org/10.1148/radiol.14131332.

[9] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M.J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K.J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C.L. Zitnick, M.P. Recht, D.K. Sodickson, Y.W. Lui, "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI", 2018, https://doi.org/10.48550/arXiv.1811.08839.

[10] Y.C. Chang, Y.H. Huang, C.S. Huang, P.K. Chang, J.H. Chen, R.F. Chang, "Magnetic Resonance Spectroscopy and Imaging Guidance in Molecular Medicine: Targeting and Monitoring of Choline and Glucose Metabolism in Cancer", Magnetic Resonance Imaging 2012 Vol. 30, 312–322, https://doi.org/10.1002/nbm.1751.

[11] S.R. Dubey, S.K. Singh, B.B. Chaudhuri, "Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark", Neurocomputing, vol. 503, 2022, pp 92-108, https://doi.org/10.1016/j.neucom.2022.06.111.

[12] S. Narayan, "The generalized sigmoid activation function: Competitive supervised learning", Information Sciences, vol. 99, Issues 1–2, 1997, pp 69-82, ISSN 0020-0255, https://doi.org/10.1016/S0020-0255(96)00200-9.

[13] A.F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)", 2018, [1803.08375] Deep Learning using Rectified Linear Units (ReLU) (arxiv.org).

[14] A. Mollahosseini, D. Chan, M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1-10, https://doi.org/10.1109/WACV.2016.7477450.

[15] D. Hendrycks, K. Gimpel, "Gaussian Error Linear Units (GELUs)", arXiv:1606.08415, https://doi.org/10.48550/arXiv.1606.08415.

[16] B. Xu, N. Wang, T. Chen, M. Li, Mu, "Empirical Evaluation of Rectified Activations in Convolutional Network", 2015, arXiv:1505.00853, https://doi.org/10.48550/arXiv.1505.00853.

[17] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, "Self-normalizing neural networks" Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, 972--981 [1706.02515] Self-Normalizing Neural Networks (arxiv.org).

[18] S.R. Dubey, S.K. Singh, B.B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark", Neurocomputing, Volume 503, 2022, pp 92-108, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2022.06.111.

[19] A.Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance", ICML '04: Proceedings of the twenty-first international conference on Machine learning, 2004, https://doi.org/10.1145/1015330.1015435.

[20] K. Anders, J.A. Hertz, A simple weight decay can improve generalization, NIPS'91: Proceedings of the 4th International Conference on Neural Information Processing Systems, 1991, pp 950–957, A Simple Weight Decay Can Improve Generalization (nips.cc).

[21] I. Loshchilov, F. Hutter, "Decoupled Weight Decay Regularization", 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 2019, https://arxiv.org/abs/1711.05101.

[22] Y. Lecun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, V. Vapnik, Comparison of learning algorithms for handwritten digit recognition, International Conference on Artificial Neural Networks, 1995.

[23] S. Ioffe, C.Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 2015, pp 448–456, [1502.03167] Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift (arxiv.org).

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", Journal of Machine Learning Research, Vol. 15, 2014, pp 1929-1958, Dropout: A Simple Way to Prevent Neural Networks from Overfitting (jmlr.org).

[25] D. Rumelhart, G. Hinton, R. Williams, "Learning representations by back-propagating errors", *Nature* **323**, 533–536 (1986). https://doi.org/10.1038/323533a0.

[26] H. Robbins, S. Monro, "A Stochastic Approximation Method", The Annals of Mathematical Statistics vol. 22 N. 3, Institute of Mathematical Statistics, pp. 400-407, 1951, https://doi.org/10.1214/aoms/1177729586.

[27] J. Duchi, E. Hazan, Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization", Journal of Machine Learning Research, Vol. 12, 2011, pp. 2121-2159, https://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf.

[28] D. Kingma, J Ba, "Adam: A Method for Stochastic Optimization", International Conference on Learning Representations, 2014, https://arxiv.org/abs/1412.6980.

[29] Y. Lecun, Y. Bengio, "Convolutional Networks for Images, Speech, and Time-Series", The Handbook of Brain Theory and Neural Networks, 1998, pp. 255–258, https://dl.acm.org/doi/10.5555/303568.303704.

[30] M.D. Zeiler, R. Fergus, „Visualizing and Understanding Convolutional Networks", In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689, Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53.

[31] Z. Li, Y. Zhang, S. Arora, Sanjeev. (2020). "Why Are Convolutional Nets More Sample-Efficient than Fully Connected Nets?", 9th International Conference on Learning Representations, ICLR 2021, Why Are Convolutional Nets More Sample-Efficient than Fully Connected Nets? | OpenReview.

[32] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", arXiv:1512.03385, https://doi.org/10.48550/arXiv.1512.03385.

[33] H Li, Z. Xu, G. Taylor, C. Studer, T. Goldstein, "Visualizing the Loss Landscape of Neural Nets", NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018, pp. 6391-6401, https://doi.org/10.3929/ethz-b-000461393.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, "Attention is all you need", Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017, Attention is All you Need (neurips.cc).

[35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", International Conference on Learning Representations, 2021, https://doi.org/10.48550/arXiv.2010.11929.

[36] P.A. Rink, "Magnetic Resonance in Medicine: A Critical Introduction", 2019, Books on Demand GmbH Editor

[37] P. Lauterbut, "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance". Nature 242, 190–191 (1973), https://doi.org/10.1038/242190a0.

[38] S.K. Kang, "Measuring the value of MRI: Comparative effectiveness & outcomes research." J Magn Reson Imaging. 2019 Jun;49(7): e78-e84. Doi: 10.1002/jmri.26647. Epub 2019 Jan 10. PMID: 30632255.

[39] T. S. Sindhu, N. Kumaratharan, P. Anandan, "A Review of Magnetic Resonance Imaging and its Clinical Applications," 6th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 2022, pp. 38-42, Doi: 10.1109/ICDCS54290.2022.9780834.

[40] R. Mann, N. Cho, L. Moy, "Breast MRI: State of the Art", Radiology, Vol. 292, No. 3, 2019, https://doi.org/10.1148/radiol.2019182947.

[41] NEMA, DICOM PS3.1 2023d - Introduction and Overview, https://dicom.nema.org/medical/dicom/current/output/chtml/part01/chapter_1.html#sect_1.1

[42] M. Dietzel, P.A.T. Baltzer, "How to use the Kaiser score as a clinical decision rule for diagnosis in multiparametric breast MRI: a pictorial essay", *Insights Imaging* **9**, 325–335 (2018). https://doi.org/10.1007/s13244-018-0611-8.

[43] C. Grippo, P. Jagmohan, T. H. Helbich, P. Kapetas, P. Clauser, P. A.T. Baltzer, "Correct determination of the enhancement curve is critical to ensure accurate diagnosis using the Kaiser score as a clinical decision rule for breast MRI", European Journal of Radiology, Volume 138, 2021, 109630, ISSN 0720-048X, https://doi.org/10.1016/j.ejrad.2021.109630.

[44] C. Kuhl, "The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice", Radiology, 2007 Aug, 244(2):356-78, doi: 10.1148/radiol.2442051620. PMID: 17641361.

[45] C. Kuhl, "Current status of breast MR imaging. Part 2. Clinical applications", Radiology, 2007 Sep;244(3):672-91. doi: 10.1148/radiol.2443051661. PMID: 17709824.

[46] Z. Xu, Y. Ding, K. Zhao, C. Han, Z. Shi, Y. Cui, C. Liu, N. Lin, X. Pan, P. Li, M. Chen, H. Wang, X. Deng, C. Liang, Y. Xie, Z. Liu, "MRI characteristics of breast edema for assessing axillary lymph node burden in early-stage breast cancer: a retrospective bicentric study", Eur Radiol, 2022 Dec;32(12):8213-8225. doi: 10.1007/s00330-022-08896-z. Epub 2022 Jun 15. Erratum in: Eur Radiol. 2022 Sep 8; PMID: 35704112.

[47] C. Grippo, P. Jagmohan, T. H. Helbich, P. Kapetas, P. Clauser, P.A.T. Baltzer, "Correct determination of the enhancement curve is critical to ensure accurate diagnosis using the Kaiser score as a clinical decision rule for breast MRI", European Journal of Radiology, Volume 138, 2021, 109630, ISSN 0720-048X, https://doi.org/10.1016/j.ejrad.2021.109630.

[48] https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads#MRI

[49] https://www.meduniwien.ac.at/kaiser-score/#:~:text=Kaiser%20Score%20ranges%20from%201,in%20a%20number%20of%20studies

[50] W.A. Kaiser, E. Zeitler, "MR imaging of the breast: fast imaging sequences with and without Gd-DTPA. Preliminary observations", Radiology, Vol. 170 N. 3, 1989, https://doi.org/10.1148/radiology.170.3.2916021.

[51] A. Gubern-Mérida, R. Martí, J. Melendez, JL Hauth, RM Mann, N. Karssemeijer, B. Platel, "Automated localization of breast cancer in DCE-MRI", Med Image Anal. 2015 Feb;20(1):265-74, doi: 10.1016/j.media.2014.12.001, Epub 2014 Dec 8, PMID: 25532510.

[52] D. McClymont, A. Mehnert, A. Trakic, D. Kennedy, S. Crozier, "Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph", J Magn Reson Imaging, 2014 Apr; 39(4):795-804, doi: 10.1002/jmri.24229, PMID: 24783238.

[53] A. Vignati, V. Giannini, M. De Luca, L. Morra, D. Persano, L.A. Carbonaro, I. Bertotto, L. Martincich, D. Regge, A. Bert, F. Sardanelli, "Performance of a fully automatic lesion detection system for breast DCE-MRI", J Magn Reson Imaging, 2011 Dec; 34(6):1341-51, doi: 10.1002/jmri.22680, Epub 2011 Sep 30, PMID: 21965159.

[54] D.M. Renz, J. Böttcher, F. Diekmann, A. Poellinger, M.H. Maurer, A. Pfeil, F. Streitparth, F. Collettini, U. Bick, B. Hamm, E.M. Fallenberg, "Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI", J Magn Reson Imaging, 2012 May; 35(5):1077-88, doi: 10.1002/jmri.23516, Epub 2012 Jan 13, PMID: 22247104.

[55] G. Maicas, G. Carneiro and A. P. Bradley, "Globally optimal breast mass segmentation from DCE-MRI using deep semantic segmentation as shape prior," *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, Australia, 2017, pp. 305-309, doi: 10.1109/ISBI.2017.7950525.

[56] G. Maicas, G. Carneiro, A.P. Bradley, J.C. Nascimento, I. Reid, "Deep Reinforcement Learning for Active Breast Lesion Detection from DCE-MRI", In: "Medical Image Computing and Computer Assisted Intervention" − MICCAI 2017, Lecture Notes in Computer Science (), vol 10435. Springer, Cham. https://doi.org/10.1007/978-3-319-66179-7_76.

[57] G. Piantadosi, M. Sansone, R. Fusco, C. Sansone, "Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks", Artificial Intelligence in Medicine 2020 Vol. 103, 101781, https://doi.org/10.1016/j.artmed.2019.101781.

[58] S. Marrone, G. Piantadosi, R. Fusco, A. Petrillo, M. Sansone, C. Sansone, "Breast segmentation using Fuzzy C-Means and anatomical priors in DCE-MRI", 23rd IEEE International Conference on Pattern Recognition (ICPR), 2016, pp. 1472–1477.

[59] B. Kayalibay, G. Jensen, P. van der Smagt, "CNN-based Segmentation of Medical Imaging Data", 2017, https://doi.org/10.48550/arXiv.1701.03056.

[60] H.S. Alshanbari, S. Amin, J. Shuttleworth, K.A. Slman, S. Muslam, "Automatic Segmentation in Breast Cancer Using Wa-tershed Algorithm", International Journal of Biomedical Engineering 2015 Vol. 2 N. 2.

[61] L. Wang, B. Platel, T. Ivanovskaya, M. Harz, H.K. Hahn, "Fully automatic breast segmentation in 3D breast MRI", 9th IEEE International Symposium on Biomedical Imaging (ISBI), 2012.

[62] A. Vignati, V. Giannini, M. de Luca, L. Morra, D. Persano, L.A. Carbonaro, I. Bertotto, L. Martincich, D. Regge, A. Bert, F. Sardanelli, "Performance of a Fully Automatic Lesion Detection System for Breast DCE-MRI", Journal of Magnetic Resonance Imaging, 2011, Vol. 34, 1341–1351, https://doi.org/10.1002/jmri.22680.

[63] C. Gallego Ortiz, A.L. Martel, "Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation", Medical physics, 2012, Vol. 39 (10), 5835-5848.

[64] A. Gubern-Mérida, M. Kallenberg, R.M. Mann, R. Martí, N. Karssemeijer, "Breast segmentation and density estimation in breast MRI: a fully automatic framework", IEEE Journal of Biomedical and Health Informatics, 2015, Vol. 19, 349-357.

[65] F. Khalvati, C. Gallego-Ortiz, S. Balasingham, A.L. Martel, "Automated Segmentation of Breast in 3D MR Images Using a Robust Atlas", IEEE Transactions on Medical Imaging, 2015, Vol. 34.

[66] V.K. Reed, W.A. Woodward, L. Zhang, E.A. Strom, G.H. Perkins, W. Tereffe, J.L. Oh, T.K. Yu, I. Bedrosian, G.J. Whitman, T.A. Buchholz, L. Dong, "Automatic segmentation of whole breast using atlas approach and deformable image registration", International Journal of Radiation Oncology Biology Physics, 2009, Vol. 73, 1493–1500.

[67] A. Fooladivanda, S.B. Shokouhi, M.R. Mosavi, N. Ahmadinejad, "Atlas-based automatic breast MRI segmentation using pectoral muscle and chest region model", 21st Iranian Conference on Biomedical Engineering (ICBME), IEEE, 2014, pp. 258–262.

[68] M. Mustra, J. Bozek, J., "Breast border extraction and pectoral muscle detection using wavelet decomposition", IEEE EUROCON 2009.

[69] S. Wu, S.P. Weinstein, E.F. Conant, M.D. Schnall, D. Kontos, "Automated chest wall line detection for whole-breast segmentation in sagittal breast MR images", Medical physics, 2013, Vol. 40(4), 1–12.

[70] L. Cai, J. Gao, D. Zhao, "A review of the application of deep learning in medical image classification and segmentation, Annals of Translational Medicine", 2020, Vol. 8, 713–713, doi: 10.21037/atm.2020.02.44.

[71] M.U. Dalmış, G. Litjens, K. Holland, A. Setio, R. Mann, N. Karssemeijer, A. Gubern-Mérida, "Using deep learning to segment breast and fibroglandular tissue in MRI volumes", Medical Physics, 2017, Vol. 44, 533–546, https://doi.org/10.1002/mp.12079

[72] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation.", Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015, pp. 234–241.

[73] V.I. Iglovikov, A. Shvets, "TernausNet". In: Bernal, J., Histace, A. (eds) Computer-Aided Analysis of Gastrointestinal Videos. Springer, Cham. https://doi.org/10.1007/978-3-030-64340-9_15.

[74] D. sen Maitra, U. Bhattacharya, S.K. Parui, "CNN based common approach to handwritten character recognition of multiple scripts", 13th IEEE International Conference on Document Analysis and Recognition (ICDAR), 2015, pp. 1021–1025.

[75] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging", Advances in Neural Information Processing Systems, Curran Associates, 2019.

[76] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

[77] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, „ A Survey on Deep Transfer Learning", Artificial Neural Networks and Machine Learning – ICANN 2018, 270–279.

[78] S. Serte, A. Serener, F. Al-Turjman, "Deep learning in medical imaging: A brief review", Transactions on Emerging Telecom-munications Technologies, 2020, https://doi.org/10.1002/ett.4080.

[79] H. Zhang, Y.N. Dauphin, T. Ma, "Fixup Initialization: Residual Learning Without Normalization", International Conference on Learning Representations, 2019, https://doi.org/10.48550/arXiv.1901.09321.

[80] D. Bahdanau, K. Cho, Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", 2nd International Conference on Learning Representations, 2014.

[81] K. He, X. Zhang, S. Ren, J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification", International Conference on Computer Vision ICCV, 2015.

[82] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks", Communications of the ACM, 2017, Vol. 60, 84–90, https://doi.org/10.1145/3065386.

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

[84] M.I. Razzak, S. Naz, A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future", Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics, 2017, Vol. 26, https://doi.org/10.1007/978-3-319-65981-7_12.

[85] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, X. He, "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[86] Z. Yang, X. He, J. Gao, L. Deng, A. Smola, "Stacked attention networks for image question answering", CoRR abs/1511.02274, arXiv preprint arXiv:1511.02274, 2015.

[87] K. Gregor, I. Danihelka, A. Graves, D.J. Rezende, D. Wierstra, "Draw: A recurrent neural network for image generation", arXiv, 2015, arXiv preprint arXiv:1502.04623.

[88] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, Y. Yang, "Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification", arXiv, arXiv preprint arXiv:1801.09927, 2018, https://doi.org/10.48550/arXiv.1801.09927.

[89] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, "Self-Attention Generative Adversarial Networks", Proceedings of the 36th International Conference on Machine Learning, 2019, pp. 7354-7363.

[90] A. Odena, V. Dumoulin, C. Olah, "Deconvolution and Checkerboard Artifacts", Distill, 2016, 1.10: e3.

[91] A. Mikolajczyk, M. Grochowski, "Data augmentation for improving deep learning in image classification problem", 2018 IEEE International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117–122.

[92] S.C. Wong, A. Gatt, V. Stamatescu, M.D. McDonnell, "Understanding data augmentation for classification: when to warp?", IEEE International conference on digital image computing: techniques and applications (DICTA), 2016, pp. 1-6.

[93] D.P. Kingma, J. Ba, "A method for stochastic optimization", arXiv, 2014, arXiv preprint arXiv:1412.6980.

[94] L.N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay", arXiv, 2018, arXiv preprint arXiv:1803.09820.

[95] L.N. Smith, "No more pesky learning rate guessing games", arXiv, 2015, arXiv preprint arXiv 1506.01186, 1506.01186v1.pdf (arxiv.org).

[96] G. Agrawal, M.Y. Su, O. Nalcioglu, S.A. Feig, J.H. Chen, "Significance of breast lesion descriptors in the ACR BI-RADS MRI lexicon", 2019, Cancer: Interdisciplinary International Journal of the American Cancer Society, 115(7), 1363-1380, https://doi.org/10.1002/cncr.24156

[97] GitHub - huggingface/pytorch-image-models: PyTorch image models, scripts, pretrained weights -- ResNet, ResNeXT, EfficientNet, EfficientNetV2, NFNet, Vision Transformer, MixNet, MobileNet-V3/V2, RegNet, DPN, CSPNet, and more

[98] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?", Advances in Neural Information Processing Systems, 34, 12116-12128, 2021, https://doi.org/10.48550/arXiv.2108.08810.

[99] A. Ali, H. Touvron, M. Caron, P. Bojanowski, M. Douze, A. Joulin, H. Jégou, "Xcit: Cross-covariance image transformers", Advances in neural information processing systems, 34, 20014-20027, 2021, https://doi.org/10.48550/arXiv.2106.09681.

[100] H. Bao, L. Dong, F. Wei, "Beit: Bert pre-training of image transformers", arXiv 2021. arXiv preprint arXiv:2106.08254, https://doi.org/10.48550/arXiv.2106.08254.

[101] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805, https://doi.org/10.48550/arXiv.1810.04805.

[102] Z. Peng, L. Dong, H. Bao, Q. Ye, F. Wei, "Beit v2: Masked image modelling with vector-quantized visual tokenizers", arXiv preprint arXiv:2208.06366, 2022, https://doi.org/10.48550/arXiv.2208.06366.

[103] E. Hoffer, B. Weinstein, I. Hubara, T. Ben-Nun, T. Hoefler, D. Soudry, "Mix & match: training convnets with mixed image sizes for improved accuracy, speed and scale resiliency", arXiv preprint arXiv:1908.08986, 2019, https://doi.org/10.48550/arXiv.1908.08986.

[104] M.L. Richter, W. Byttner, U. Krumnack, A. Wiedenroth, L. Schallner, J. Shenk, "Size matters", Artificial Neural Networks and Machine Learning–ICANN: 30th International Conference on Artificial Neural

Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30 (pp. 133-144). Springer International Publishing, https://doi.org/10.1007/978-3-030-86340-1_11.

[105] D. Jacob, O. Nankar, S. Gite, S. Patil, K. Kotecha, "Lyme Disease Detection Using Progressive Resizing and Self-Supervised Learning Algorithms", pre-print article Available at SSRN 4059738, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4059738

[106] F. Colangelo, F. Battisti, A. Neri, "Progressive Training Of Convolutional Neural Networks For Acoustic Events Classification," 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021, pp. 26-30, https://doi.org/10.23919/Eusipco47968.2020.9287362.

[107] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, "mixup: Beyond empirical risk minimization", arXiv preprint arXiv:1710.09412, 2017, https://doi.org/10.48550/arXiv.1710.09412.

[108] R. Müller, S. Kornblith, G.E. Hinton, "When does label smoothing help?", Advances in Neural Information Processing Systems (NeurIPS 2019), 2019, https://arxiv.org/abs/1906.02629.

[109] https://github.com/kentaroy47/timm_speed_benchmark.

[110] Carriero, A.; Groenhoff, L.; Vologina, E.; Basile, P.; Albera, M. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics* **2024**, *14*, 848. https://doi.org/10.3390/diagnostics14080848

[111] Saha, A., Harowicz, M.R., Grimm, L.J., Kim, C.E., Ghate, S.V., Walsh, R. and Mazurowski, M.A., 2018. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. British journal of cancer, 119(4), pp.508-516.

[112] Zhao, X.; Liao, Y.; Xie, J.; He, X.; Zhang, S.; Wang, G.; Fang, J.; Lu, H.; Yu, J. BreastDM: A DCE-MRI Dataset for Breast Tumor Image Segmentation and Classification. Comput. Biol. Med. 2023, 164, 107255

[113] Adam, R.; Dell'Aquila, K.; Hodges, L.; Maldjian, T.; Duong, T.Q. Deep Learning Applications to Breast Cancer Detection by Magnetic Resonance Imaging: A Literature Review. Breast Cancer Res. 2023, 25, 87.

[114] Janse, M.H.A.; Janssen, L.M.; Van Der Velden, B.H.M.; Moman, M.R.; Wolters-van Der Ben, E.J.M.; Kock, M.C.J.M.; Viergever, M.A.; Van Diest, P.J.; Gilhuijs, K.G.A. Deep Learning-Based Segmentation of Locally Advanced Breast Cancer on MRI in Relation to Residual Cancer Burden: A Multi-Institutional Cohort Study. J. Magn. Reson. Imaging 2023, 58, 1739–1749.

[115] Chung, M.; Calabrese, E.; Mongan, J.; Ray, K.M.; Hayward, J.H.; Kelil, T.; Sieberg, R.; Hylton, N.; Joe, B.N.; Lee, A.Y. Deep Learning to Simulate Contrast-Enhanced Breast MRI of Invasive Breast Cancer. Radiology 2023, 306, e213199.

[116] Osuala, R.; Joshi, S.; Tsirikoglou, A.; Garrucho, L.; Pinaya, W.H.L.; Diaz, O.; Lekadir, K. Pre- to Post-Contrast Breast MRI Synthesis for Enhanced Tumour Segmentation. arXiv 2023

[117] Li, Y.; Fan, Y.; Xu, D.; Li, Y.; Zhong, Z.; Pan, H.; Huang, B.; Xie, X.; Yang, Y.; Liu, B. Deep Learning Radiomic Analysis of DCE-MRI Combined with Clinical Characteristics Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Front. Oncol. 2023, 12, 1041142.

Table of tables